The copyright © of this thesis belongs to its rightful author and/or other copyright owner. Copies can be accessed and downloaded for non-commercial or learning purposes without any charge and permission. The thesis cannot be reproduced or quoted as a whole without the permission from its rightful owner. No alteration or changes in format is allowed without permission from its rightful owner.

DATA VIRTUALIZATION DESIGN MODEL FOR NEAR REAL TIME DECISION MAKING IN BUSINESS INTELLIGENCE ENVIRONMENT

AYAD HAMEED MOUSA ALBADRI

Universiti Utara Malaysia

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2017

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

AYAD HAMEED MOUSA

calon untuk Ijazah PhD
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"DATA VIRTUALIZATION DESIGN MODEL FOR NEAR REAL TIME DECISION MAKING IN BUSINESS INTELLIGENCE ENVIRONMENT"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***23 Februari 2017.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:
February 23, 2017.

Pengerusi Viva: *(Chairman for VIVA)*	**Assoc. Prof. Dr. Wan Rozaini Sheik Osman**	Tandatangan *(Signature)* Rozain
Pemeriksa Luar: *(External Examiner)*	**Prof. Dr. Mohd Nordin Abdul Rahman**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Dr. Mazni Omar**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Prof. Dr. Norshuhada Shiratuddin**	Tandatangan *(Signature)* Shuhada
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Muhamad Shahbani Abu Bakar**	Tandatangan *(Signature)*

Tarikh:
(Date) *February 23, 2017*

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Universiti Utara Malaysia

Declaration

I declare that this thesis is my own work and has not previously been submitted in any form for another degree or diploma at any other university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Ayad Hameed Mousa Albadri
10 June 2017

UNIVERSITI UTARA MALAYSIA ILMU BUDI BAKTI UUM Universiti Utara Malaysia

Abstrak

Tujuan utama Kepintaran Perniagaan (BI) ialah untuk memberi tumpuan kepada menyokong pembuatan keputusan strategik, operasi, dan taktikal sesebuah organisasi dengan menyediakan salinan data yang menyeluruh, tepat, dan jelas kepada pembuat keputusan. Sebuah Gudang Data (DW) dianggap sebagai input dalam aktiviti-aktiviti sistem pembuatan keputusan, dan dibentuk melalui proses Ekstrak (Extract), (Ubah) Transform, dan Beban (Load) (ETL). ETL beroperasi dalam tempoh yang tertentu dan memerlukan data untuk memproses dan menghantar data. Walaubagaimanapun, menyediakan maklumat menghampiri masa nyata bagi membantu pengintergrasian data dalam menyokong pembuatan keputusan adalah satu isu yang telah diketahui. Ketidakbolehcapaian maklumat menghampiri masa nyata boleh diatasi menggunakan pemayaan data (DV) kerana ia menyediakan maklumat yang utuh, abstrak, menghampiri masa nyata, dan terlindung bagi memenuhi permintaan pengguna. Tidak ketinggalan, ketika ini, terdapat kajian yang terhad berkaitan model BI bagi membangun dan mengurus data dalam persekitaran maya yang mampu memenuhi keperluan organisasi. Oleh itu, matlamat utama kajian ini adalah bagi mengusulkan sebuah model DV untuk pembuatan keputusan menghampiri masa nyata dalam persekitaran BI. Kaedah penyelidikan sains reka bentuk diadaptasi bagi mencapai objektif kajian. Sebagai hasil kajian, sebuah model yang dinamakan Model Pembangunan Pemayaan Data (DVDeM) diusulkan, yang mengemukakan fasa-fasa dan komponen-komponen yang mempengaruhi persekitaran BI. Bagi mengesahkan model, semakan pakar dan perbincangan kumpulan fokus telah dijalankan. Sebuah prototaip berdasarkan model yang diusulkan telah dibangunkan, dan dilaksanakan dalam dua kajian kes. Selain itu, satu alat pengukuran telah dibangunkan bagi menilai kebolehgunaan dan keupayaan model dalam menyediakan data menghampiri masa nyata. Sejumlah 60 subjek kajian telah terlibat, dan dapatan menunjukkan 93% daripada subjek kajian bersetuju bahawa prototaip yang menerapkan DVDeM berupaya menyediakan data menghampiri masa nyata dalam menyokong proses pembuatan keputusan. Daripada jakian tersebut, dapatan juga menunjukkan bahawa majoriti responden (melebihi 90%) dalam sektor pendidikan dan perniagaan, telah mengakui kefungsian DVDeM dan kebolehgunaan prototaip, khasnya keupayaan menghantar data pembuatan keputusan menghampiri masa nyata. Dapatan juga menunjukkan sumbangan teoririkal dan praktikal bagi pembangun untuk membangun aplikasi BI yang cekap menggunakan teknik DV. Juga, min bagi setiap item ukuran adalah lebih besar dari 4 yang menunjukkan responden setuju dengan setiap pernyataan bagi setiap item pengukuran. Sementara itu, skor min bagi atribut kebolehgunaan model reka bentuk DVDeM secara keseluruhan adalah “tinggi” atau “agak tinggi”. Oleh itu, keputusan menyediakan petunjuk yang cukup bahawa pembangunan sistem yang menerap DVDeM membuahkan sistem yang dilihat oleh majoriti responden sebagai berkebergunaan tinggi dan berupaya menyokong data pembuatan keputusan yang menghampiri masa nyata.

Keywords: Kepintaran Perniagaan, Sistem Sokongan Keputusan, Pemayaan Data, Gudang Data, Pembuatan Keputusan.

Universiti Utara Malaysia

Abstract

The main purpose of Business Intelligence (BI) is to focus on supporting an organization's strategic, operational and tactical decisions by providing comprehensive, accurate and vivid data to the decision makers. A data warehouse (DW), which is considered as the input for decision making system activities is created through a complex process known as Extract, Transform and Load (ETL). ETL operates at pre-defined times and requires time to process and transfer data. However, providing near real time information to facilitate the data integration in supporting decision making process is a known issue. Inaccessibility to near real-time information could be overcome with Data Virtualization (DV) as it provides unified, abstracted, near real time, and encapsulated view of information for querying. Nevertheless, currently, there are lack of studies on the BI model for developing and managing data in virtual manner that can fulfil the organization needs. Therefore, the main aim of this study is to propose a DV model for near-real time decision making in BI environment. Design science research methodology was adopted to accomplish the research objectives. As a result of this study, a model called Data Virtualization Development Model (DVDeM) is proposed that addresses the phases and components which affect the BI environment. To validate the model, expert reviews and focus group discussions were conducted. A prototype based on the proposed model was also developed, and then implemented in two case studies. Also, an instrument was developed to measure the usability of the prototype in providing near real time data. In total, 60 participants were involved and the findings indicated that 93% of the participants agreed that the DVDeM based prototype was able to provide near real-time data for supporting decision-making process. From the studies, the findings also showed that the majority of the participants (more than 90%) in both of education and business sectors, have affirmed the workability of the DVDeM and the usability of the prototype in particular able to deliver near real-time decision-making data. Findings also indicate theoretical and practical contributions for developers to develop efficient BI applications using DV technique. Also, the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item. Meanwhile, it was found that the mean scores for overall usability attributes of DVDeM design model fall under "High" or "Fairly High". Therefore, the results show sufficient indications that by adopting DVDeM model in developing a system, the usability of the produced system is perceived by the majority of respondents as high and is able to support near real time decision making data.

Keywords: Business Intelligence, Decision Support Systems, Data Virtualization, Data Warehouse, Decision-Making.

Acknowledgements

In the Name of Allah, the Most Gracious the Most Merciful

Praise to Allah, first and foremost all praise and thanks to Allah S.W.T. for giving me the opportunity and strength to go through this Ph.D journey until it complete. I am extremely grateful for all those who have provided me with their help, encouragement, and assistance in many different ways in order to complete my thesis.

My sincere gratitude is expressed to my 1st supervisor Prof. Dr. Norshuhada Shiratuddin and 2nd supervisor Dr. Muhamad Shahbani Abu Bakar for their help, motivation, enthusiasm, inspiration and endless support through my Ph.D study. Thank you very much for their patience, feedback, suggestion, immense knowledge in providing a guiding path through my Ph.D journey. Without their encouragement and inspiration, this thesis would not have emerged and been completed.

Not forgetting to express my grateful and thanks to the Ministry of Higher Education and Scientific Research (Iraq) and University of Karbala for supporting and funding my studies. Without their continuous support, this research study would not have materialized. In the same aspect, I would like to thank University Utara Malaysia for giving me the opportunity to undertake this research and providing me with the necessary resources and support.

Most importantly, my special acknowledgment goes to my wife. She has shown her endless love and understanding through my study. Thank you so much for your support and patience during this journey. Also to my piece of my heart Al-Fadhil, thank you so much for bringing the happiness and smile into my life.

Also, my deepest appreciation and gratitude to my family especially, my mother, sisters, and brothers and their families for believing in my ability and supporting me spiritually throughout my life, thank you so much for making my life so meaningful, bringing so much of joy in my life. I appreciate all of you very much. I pray to Allah that all of you will have a special place in the hereafter.

Also my special acknowledgment goes to my nephew Mohammed Al-Masuadi, thank you for bringing the happiness and smile into my life.

I would like to thank all my friends, colleges, students and fellow PhD candidates that I met during my time at UUM. During uncountable chats, discussions and coffee sessions, they all contributed in some way to my development and sometimes helped me to see things differently. I wish everybody the very best for their future.

Ayad Hameed Mousa Albadri

Universiti Utara Malaysia

10 June 2017

UUM Universiti Utara Malaysia

Table of Contents

Universiti Utara Malaysia

Universiti Utara Malaysia

Universiti Utara Malaysia

Universiti Utara Malaysia

Universiti Utara Malaysia

Universiti Utara Malaysia

List of Tables

Universiti Utara Malaysia

Universiti Utara Malaysia

List of Figures

Universiti Utara Malaysia

Universiti Utara Malaysia

Universiti Utara Malaysia

List of Appendices

Universiti Utara Malaysia

List of Abbreviations

BI	Business Intelligence
CRM	Customer Relationship Management
CG	Conceptual Graph
DW	Data Warehousing
DV	Data Virtualization
DSS	Decision Support System
IT	Information Technology
IS	Information System
LOB	line-of-Business
SOA	Service Oriented Architecture
UML	Unified Modeling Language
SEU	Subjective Expected Utility
ERP	Enterprise Resource Planning
VDIS	View based Data Integration Theory
WST	Work System Theory
OLTP	Online Transaction Processing
EII	Enterprise Information Integration
GD	Goal-Driven
VDIS	View based Data Integration Theory
WST	Work System Theory
RFID	Radio Frequency Identification
ODS	Operational Data Store
GDT	Goal Decomposition Tree
GUI	Graphical User Interface

Universiti Utara Malaysia

KPIs	Key Performance Indicators
OMT	Object-Method Table
RTBI	Real Time Business Intelligence
CA	Comparative Analysis

Universiti Utara Malaysia

List of Publication

The following are a few publications related to this that have been published in journals and proceedings:

JOURNALS

1. **Mousa, A. H**., Shiratuddin, N., & Bakar, M. S. A. (2015a). Process Oriented Data Virtualization Design Model for Business Processes Evaluation (PODVDM) Research in Progress. *Jurnal Teknologi*, 72(4).

2. **Mousa, A. H**., Shiratuddin, N., & Bakar, M. S. A. (2014c). Virtual Data Mart for Measuring Organizational Achievement Using Data Virtualization Technique (KPIVDM). *Jurnal Teknologi*, 68(3).

3. **Mousa, A. H**., Shiratuddin, N., & Bakar, M. S. A. (2014b). Generic Framework for Better Choosing Between Data Integration Types (GFCBDIT) During Build Business Intelligence Applications. International Journal of Digital Content Technology and its Applications, 8(5), 27.

4. **Mousa, A. H**., Shiratuddin, N., & Bakar, M. S. A. (2016). Evaluation Framework for Business Process Evaluation Approaches. *Journal of Computer Science & Computational Mathematics*, 6(3), 7.

PROCEEDINGS

1. **Mousa, A. H**., Shiratuddin, N., & Bakar, M. S. A. (2014). Data Warehouse for Business Process Evaluation Approach Opportunities and Challenges Paper presented at *the Knowledge Management International Conference (KMICe)*, Malaysia.

2. **Mousa, A. H**., Haider, N, Kh & Bakar, M. S. A (2013, August). Intelligent Iraqi Health System (IIHS) Using Online Analytical Process (OLAP) Model. Paper presented at *the 4th International Conference on Computing and Informatics, ICOCI*, Sarawak, Malaysia.

3. **Mousa, A. H**., & Shiratuddin, N. (2015). Data Warehouse and Data Virtualization Comparative Study. Paper presented at *the Developments of E-Systems Engineering (DeSE)*, 2015 International Conference on (pp. 369-372). IEEE.

4. **Mousa, A. H**., Shiratuddin, N., & Bakar, M. S. A. (2015b). RGMDV: An approach to requirements gathering and the management of data virtualization projects. Paper presented at *the Innovation and Analytics Conference and Exhibition (IACE 2015*): Proceedings of the 2nd Innovation and Analytics Conference & Exhibition.

CHAPTER ONE
INTRODUCTION

1.1 Overview

This chapter provides some background of the study that deliberates on issues that lead to the motivation aspects of the study, the specification of the problem, identification of research gap, and formulation of research questions and objectives. It also discusses the scope and limitations of the study, contribution of study, and research framework, operational definitions of terms used throughout the study, as well as chapter summary.

1.2 Background of Study

The background of this study can be classified into six main sections: business intelligence, decision support system, data sources, real-time business intelligence, data warehouse, and finally, data virtualization. The following are the list of details:

Universiti Utara Malaysia

1.2.1 Business Intelligence

Business Intelligence (BI) is the mechanism to provide insights for most of the operations and performance of organizations, in addition to identifying strategic business opportunities. Over the years, numerous definitions of BI have emerged; however, there is no comprehensive definition that is acceptable by all researchers in this area. BI from a technical point is a set of techniques, tools and methodologies that work together to transform the information and data belonging to the organizations into meaningful and actionable information and making this information available to decision makers in an organization (Cody, Kreulen, Krishna, & Spangler, 2002; Dayal, Castellanos, Simitsis, & Wilkinson, 2009; Kimball, Ross, Thorthwaite, Becker, & Mundy, 2008; Olszak & Ziemba, 2007; Reinschmidt & Francoise, 2000). The assistance provided by the BI tool is embodied in many points such as the identification and discovery of tactical and strategic orientation as well as to predict future events. This is achieved by using tools such as data mining and other knowledge discovery approaches such as predictive analytics approach (Hair Jr, 2007; Watson & Wixom, 2007; Wixom et al., 2011). BI tools are generally used to improve the performance of business intelligence in order to increase the competitive effectiveness of these organizations (Dayal, Castellanos, et al., 2009; Dayal, Wilkinson, Simitsis, & Castellanos, 2009; Mousa, Shiratuddin, & Bakar, 2014b, 2015a). These tools are also widely used to improve the efficiency of organizations, particularly in cost reduction, and help to attract and sustain customers as long as possible in order to improve sales and revenue (Ben Azvine, Cui, Nauck, & Majeed, 2006; B, Z, D, & B, 2006; Ramanigopal, Palaniappan, & Hemalatha, 2012; Van der Lans, 2012; Watson, Wixom, Hoffer, Anderson-Lehman, & Reynolds, 2006).

Universiti Utara Malaysia

1.2.2 Decision Support Systems

Decision Support Systems (DSS) is one of the types of software systems that have the ability to provide possible assistance to organizations to deal with the available data and transform into meaningful information. DSS also perform analysis onto these data and DSS represents one of the main areas of information systems (Hall, 2008). It can be concluded that BI is a DSS integrated approach that combines data collection and storage, as well as possessing the ability to analyze that data to reach meaningful and viable information for implementation. Therefore, BI can loosely be defined as “data-driven DSS” and relies the full adoption of the data (Negash &

Gray, 2008; Sauter, 2014). It can be concluded that the meaningful information can be considered as the backbone of decision support systems and the accuracy and success of DSS strongly associated with the quality of data that is obtained and fed into it.

In summary, the relationship between DSS and BI systems is DSS was an old system to assist decision maker to solve the business problem while BI was a novel as a successor of DSS. The writing also differentiates DSSs with BI systems based on their definition, architecture, and functionality.

1.2.3 Data Sources

Nowadays, organizations collect, generate and store meaningful information at an exponentially growing rate. This is driven by regulatory requirements such as proper accounting and tax records, and certifications as well as fulfilling business needs such as operations management, finance and others. The emergence of technologies such as bar codes, Radio-Frequency Identification (RFID) and the Internet has made it easier to capture data. The increasing volume of data does not necessarily have a positive impact on the decision-making process because it is difficult to obtain vivid and meaningful information (Barone, Yu, Won, Jiang, & Mylopoulos, 2010; Sargut & McGrath, 2011). To gain competitive advantage, organizations rely on sound information to make decision and promptly react to market changes. Previous BI systems rely entirely on internal data to be processed, analyzed and reported. Today, there are many new resources of data and information readily available for BI systems to analyze and report to give organizations a leading advantage (Ben Azvine et al., 2006; Cody et al., 2002; Ramanigopal et al., 2012; Van der Lans, 2012).

Universiti Utara Malaysia

1.2.4 Real-Time and Near Real Time Decision Making

Azvine (2006) explicitly expressed that there is no specific accurate definition or understanding of the term of "Real-Time" and he is introducing three different meanings. “Zero Latency” Processes, up to date information whenever needed by user, and Key performance Indicators (KPIs) relate to current situation (i.e. now) (Ben Azvine et al., 2006). However, there are two reasons been given by authors for the significance of real-time BI (RT-BI); firstly, the business environment, secondly, the advances in technology (Anderson-Lehman, Watson, Wixom, & Hoffer, 2008). The environment businesses operate in changes rapidly (e.g. share prices, sales pattern and others). The increasing needs for the live information that used to create reports. Today’s technology would look forward to the design of RT-BI systems. For example, the Internet is mentioned as a means of distributing data throughout an organization. However, according to literature, the current BI systems face two challenges in regards to providing RT-BI i.e. the converting from data to information and from information into action or knowledge (Marsden, 2008).

In line with the above situation, RT-BI refers to different timescales and may vary significantly. For example, if a RT-BI system is used to support a decision that is due once a month there should be sufficient time to run all ETL processes. For a RT-BI system that is used in a financial trading environment on the other hand, it is unlikely that an “every-night” schedule is sufficient for the purpose. In the context of this research, the term near real-time mean; the data processing systems are to slightly slower than real-time and based on organization requirements. Consequently, the next paragraphs discussed all types of data processing.

1.2.5 Data Warehouse

Most BI systems are connected to a central database called the Data Warehouse (DW). DW can be defined as a subject oriented, non-volatile, integrated and time variant collection of data in favor of decision making (Inmon & Hackathorn, 1994). There are many main business measures, usually extracted, transformed and loaded from various data sources and integrated in the DW to make it ready for use by online analytical processing or any other BI tools. Technically, DW data will be transferred and copied from one database to another. During this transfer-copy process, data cleansing, extraction and integration will be performed iteratively until an acceptable quality is attained before it is loaded into a new database. This process is named ETL (Inmon, 1996; Kimball et al., 2008; Mousa et al., 2014b). Most decision-making systems suffer from the fact that the input data is not in real time because the ETL process needs to be implemented earlier consumes time for processing, transport and it expends resources. On the other hand, this series of processes are long and complex, so any change to requirements makes it necessary for developers to redo the complex procedures which are time and resource consuming which potentially leads to a negative impact on the decision-making process. These constraints prompted organizations and developers to find a new technique called Data Virtualization (DV) technique.

1.2.6 Data Virtualization

DV presents numerous benefits to enterprises strategically and technically (Richter, McFarland, & Bredfeldt, 2012; Weng et al., 2004). Businesses rely heavily on data and DV presents a platform for business agility, action ability, information speed, and information quality. DV has been defined by several authors. In this study; the definition by Lans (2012) is adapted because this definition has in common that they stress the holistic character of DV that includes the technical, as well as non-technical aspects, hence, in this research this definition was adopted. Generally, data virtualization refers to a set of data stores which offer users to query access and manipulate data in a unified, abstracted and encapsulated manner, regardless of the data location. Data virtualization hides the fact that the data is being integrated to form that unified view (Ben Azvine et al., 2006; Bucher, Gericke, & Sigg, 2009; Marjanovic, 2007; Van der Lans, 2012). In a nutshell, using DV will lead to reduce the time for data integration, and quicker response time to meet business information requirement. In addition, DV provides opportunities to collect information and access new data sources which may have been added recently. On the other hand, DV can also deal with more complete data through a huge data size (Ferguson, 2011; Hopkins, 2011; Van der Lans, 2012). From the technical perspective, DV provides data of higher quality; this can be translated to business syntax and contexts, rather than delivery systems and data storage contexts. In addition, DV is based on information delivery that is quick, efficient and effective, since achieving data integration is made easier, in terms of information scope and timeliness (Eve & Davis, 2011; Van der Lans, 2012; Weng et al., 2004).

This research aims to propose a model that have a capability to deliver near real-time and relevant data by extracting it from multiple resources - whether these sources were located homogeneous and heterogeneous or from a single server or multiple servers by using DV technique. this data is then be integrated and used for BI environment; in order to facilitate near real-time decision making.

1.3 Motivation of Study

A few phenomena have motivated the acceleration of this study. Accordingly, this section summarizes those phenomena.

1.3.1 Importance of Data Management used as Inputs for Decision-Making Systems

According to INSIDE- ERP which is an industry leader in research and information for ERP professionals, today's business settings have companies constantly looking for ways to manage large amounts of data, analyze it, and present it in an organized way (IBM, 2016). This needs to be done in order to help paint a clear picture of the business life cycle to ensure efficiency.

DSS have been utilized by organizations to support faster and more reliable decision making. The Aberdeen Group study in March 2011 concluded that 43 percent of organizations or companies are facing difficulties in making timely decisions. There is less and less time available, and is "sometimes crucial" to make decisions, as shown in Figure 1.1. Top management of these organizations argue that the time required to make appropriate decisions is becoming critical and events happen before decisions can be made. Therefore, it is imperative that organizations are able to adapt and modify existing reports quickly towards effective decision making (Group, 2011).

Universiti Utara Malaysia

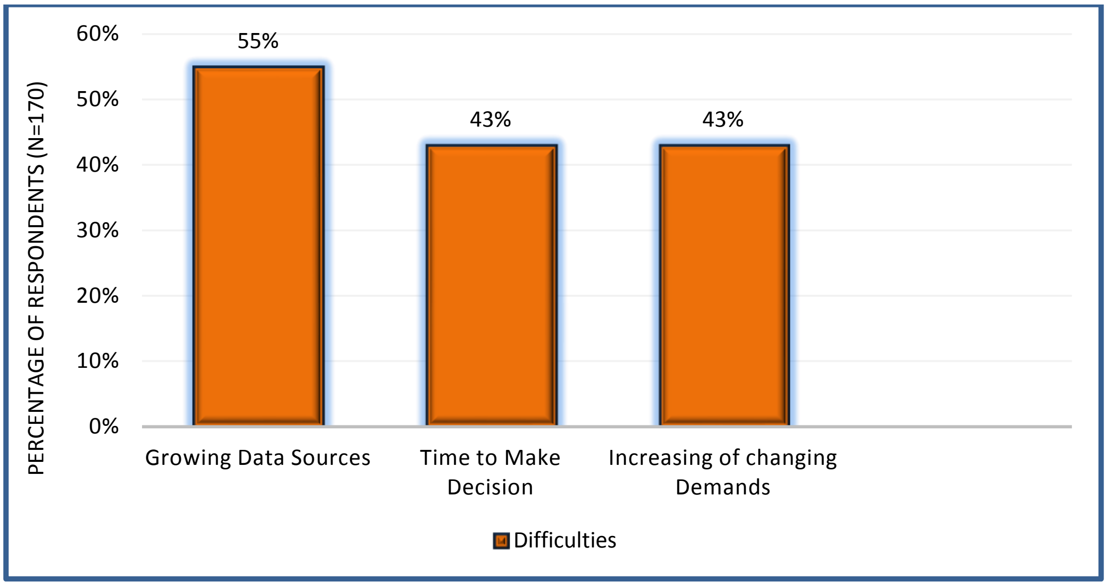


Figure 1.1. The Difficulties in Making Timely Decisions

1.3.2 The Emergence of Data Virtualization as a Data Processing Technology in Business Intelligence Environment

UUM Universiti Utara Malaysia

Virtualization is a concept that has been in existence for some time within the realm of Information Technology (IT). IBM applied virtualization by splitting mainframes into separate virtual machines. This enabled one machine to concurrently run multiple applications. Virtual memory or memory virtualization is also commonly known as paging. It was used to simulate more memory than was available in a machine. Currently, many objects such as processors, storage, network, data repositories and computer operating systems can be virtualized. (Ferguson, 2011; Hopkins, 2011; Nasir & Shahzad, 2007; Van der Lans, 2012; Weng et al., 2004). It can be concluded that DV is a potential concept that can be used for reporting and analytics. Lans (2012) argues that DV requires fewer databases and lesser numbers of transactions and transformation processes, thus leading to a shorter chain. It is also cost effective and less labor intensive as fewer databases are required. On the other hand, DV technology provides a unified, abstracted, and encapsulated view of information for querying. DV can deal and manipulate information that is kept, either in different or heterogeneous set of data stores. In other aspect, DV can reduce data movement which is required to integrate data sources on different platforms so as to produce research data sets (Richter et al., 2012).

1.3.3 Summary of Research Motivation

Organizations are constantly seeking a leading advantage in businesses and to make it more efficient have resorted to combining their own internal data with new data sources. This combination, which requires the presence of real-time data or near real-time have enriched the analytical and reporting capabilities resulting in organizations capable of making decisions and to react faster, it is, therefore, critical to have an agile architecture that is easy to change, and the best way to do that is to create an architecture that consists of fewer components such as databases and fewer transformation processes. The strategies of improving BI applications need to be considered for preparing relevant data as well as real-time data. One of the approaches that can be exploited to enhance BI applications is to improve data integration. Hence, the reasons for this should be further addressed. On the other hand, there is a lack of comprehensive models that can have used by BI developer as a guideline for data integration in a virtual way to provide a near real-time decision making in BI environment. Therefore, this study proposes a DV design model for near real-time decision making in BI environment.

1.4 Problem Statements

In organizational practices, all BI applications require extraction and interpretation of relevant information that have been gathered during the execution of processes (Ben Azvine et al., 2006; Barone et al., 2010; Reinschmidt & Francoise, 2000; Shahzad & Giannoulis, 2011; Van der Lans, 2012; Weske, 2012). For that reason, the accuracy, speed and live data for extracting and interpreting the information and

Universiti Utara Malaysia

providing them in a timely manner, would have a positive impact on the analysis of the performance of any organization. It will also provide support for decision-makers to make better decisions based on these analyses. Therefore, there is an urgent need for methods, models and guidelines to facilitate timely and accurate data integration as an input for BI. Without these, some important phases in the decision making systems may be overlooked (Zellner, 2011). On the other hand, the quality associated with making better decisions may suffer without satisfactory and sufficient approaches and techniques (Pourshahid et al., 2009; Shahzad, 2010).

Nevertheless, little work has been done to integrate data in the virtual manner to be used as an input for BI, which is vital in supporting decision-making process by provide real-time or near real-time data for decision-making systems. In particular, through an in-depth review of the current issues and approaches, this study does not find any published academic literature on the adequate methods, clearly defined steps or instructions that can guide the BI developers for designing and developing DV model. In the absence of such methods, guidelines or clearly defined steps, important steps may be ignored and credible steps cannot be taken.

The problem domain of this study is to provide relevant information at near real-time that can be used in BI environment in supporting decision making process to make better decisions. The main concern is to facilitate data integration process in virtual manner using data virtualization technique. Currently, the existing techniques for extracting, integrating, storing data and information pertaining to BI using DW has some limitations and constraints which can be classified into the following.

A DW is considered as an input for decision support systems activities and is created through the ETL process. ETL is a complex process and it operates at pre-defined

Universiti Utara Malaysia

times. The ETL process requires time to process and transfer data (Kemper & Baars, 2009; Meredith, O'Donnell, & Arnott, 2008; Negash & Gray, 2008; Trivedi, 2011). On the same aspect, in building the data warehouse, ETL, plays the role of data integration, injection and is the most time-consuming activity. Thus it is necessary to improve the performance of ETL or find new data integration technique (Guo, Yuan, Sun, & Yue, 2015). It can be concluded that there is a lack of information provided by ETL that is accessible and could be used by decision makers in near real-time. Furthermore, the conventional BI approaches do not completely support the emerging business needs of decision support.

It has been observed that most of the available BI models use DW. DW design model, however, is complex and not flexible causing difficulty for DW users to deal with large amounts of the organization's information, especially with changing business requirements and business needs (Jossen, Blunschi, Mori, Kossmann, & Stockinger, 2012). On top of that, due to this difficulty in improving the efficiency of the ETL process, designers are forced to re-design ETL when they intend to add new sources. Consequently, this resulted in changes or alterations in rules or sources of data. Therefore, it is noted that the process of redesigning and building ETL processes are challenging (Liutong Xu, 2011). On the other hand, data extraction using ETL can be more expensive and riskier than allowing an existing, stable data source to remain as is - as an isolated line-of-business (LOB) data sources (Davis & Eve, 2011).

Universiti Utara Malaysia

The amount of data is growing dramatically in today's world (Zikopoulos, deRoos, Bienko, Buglio, & Andrews, 2015). According to IBM (2015), 90 percent of the existing data in the world has been created in the last two years. The data exists in different formats such as data that generated by cell phones, GPS signals, social media sites and digital videos (Van der Meulen & Rivera, 2013). Singh et al. (2012) stated that, since there are more data to consider, there is a need of support on how to manage the data to make successful decisions.

Normally, the process of updating in the DW is often performed in an offline manner. During the update, all BI applications cannot get access to DW and be delivered in real time (Ricardo Jorge Santos, 2011). In line with the above situations, the DW is always (to varying degrees) out of date, a state that is not desirable in particular in RT-BI or near RT-BI systems.

Traditionally, data integration faces difficulties in supporting decision-making process in terms of near real-time data. These difficulties will generate a negative effects in the decision-making process particularly, in the business and education sectors as well as other sectors which need near real time data.

In education sector, the higher education institutions HEIs data integration process is becoming increasingly difficult to analyze due to its fast-growing trend. In this regard, getting previously data integration that could assist policy and decision making in the education sectors is becoming challenging (Akanmu & Jamaludin, 2016; Conrad et al., 2013; Gill, Borden, & Hallgren, 2014). It can be said that, the fast growth of data, and the consequent problem of data integration are also found in many sectors. Therefore, the data integration in education sector has been constrained by the limitations of the presently adopted technologies in the processing of the near real time data and the representation and presentation in supporting the decision-making process in making timely decisions (Botta-Genoulaz & Millet, 2006; Siemens et al., 2013). This, invariably, has been responsible for the unsatisfactory managerial decision and policy making process (Samoff, 1999; Siemens, Dawson, &

Lynch, 2013). This, therefore, forms the need for further studies that will attend to data management issues of HEIs data.

In business sector, data integration and acquiring knowledge are considered as an essential need for businesses in supporting decision making process to compete in the market (Sureephong, Chakpitak, Ouzrout, & Bouras, 2008); therefore, the importance of decision making and business intelligence have increased within businesses in such economies (Karim, 2011a). As a result, the adoption of decision-making tools based on near real-time data integration is of vital importance in the business sector. Besides, since businesses aim to enhance decision making to maintain and improve their business performance; taking the right decision of a right time.

In the context of this study, both education and business sectors are implemented as case studies. The details about case studies are described in the Chapter 5.

In other aspect, Wayne Eckerson of TechTarget has conducted a DV market research in April 2013. This research reveals that 35% of the respondents have invested in DV, while 27% and 18% of the respondents have partially or completely deployed the software. Furthermore, approximately one-third of the organizations have data virtualization under consideration (Lans, 2013). In line with the above situations, the need of a general and comprehensive model which can be used as a guideline to design and implement data virtualization is extremely significant. In addition, the non-existence of such model can be considered as one of shortcoming in developing BI environments. Therefore, the non-existence of a precise guideline, coupled with the absence of sufficient and efficient data virtualization design models as well as the lack of improved techniques to take advantage of the benefits of real time data for using by decision making in business intelligence environments, have made this study relevant and important. As such the Data Virtualization Design model for near Real-Time Decision Making in Business Intelligence Environment is proposed. In section 1.5, 1.6, and 1.7, the proposed solution, research questions and research aim and objectives are formed respectively.

1.5 Proposed Solution

In line with the above situations, there is a need for a comprehensive model for BI data that not only serves the decision-makers process in the delivery of near real-time, but this model to be a guide for developers in this area. Hence, the DVDeM model was proposed. Therefore, the efforts of this research are focused on providing a near real-time data to support decision-making process in BI environment.

1.6 Research Questions

From the above data integration issues discussed, it can be put forward that well-designed BI model and appropriately defined elements and components are necessary in BI environment. The main research question of this study is how to design and develop a DV design model for near real-time decision making in BI environment. Hence, this leads to the following research questions:

i. What are the main components of such design model?
ii. How to design a system for BI environments using data virtualization technique?
iii. How should such model be validated in BI environment?
iv. Based on the model, how can a prototype be developed for BI environment?
v. How should such prototype be tested in BI environment?

1.7 Research Aims and Objectives

As outlined above, the main aim of this study is to propose a DV Design Model for Near- Real Time Decision Making in BI Environment (DVDeM). To accomplish the main aim, the following specific objectives are outlined:

i. To identify the main components of the DVDeM model.
ii. To develop a DVDeM model using data virtualization technique for BI environments.
iii. To validate DVDeM in BI environment using expert review and focus group discussion.
iv. To develop a prototype based on the DVDeM model.
v. To test the DVDeM prototype in BI environment.

Accordingly, the mapping of research questions and research objectives diagram illustrate in Figure 1.2.

Universiti Utara Malaysia

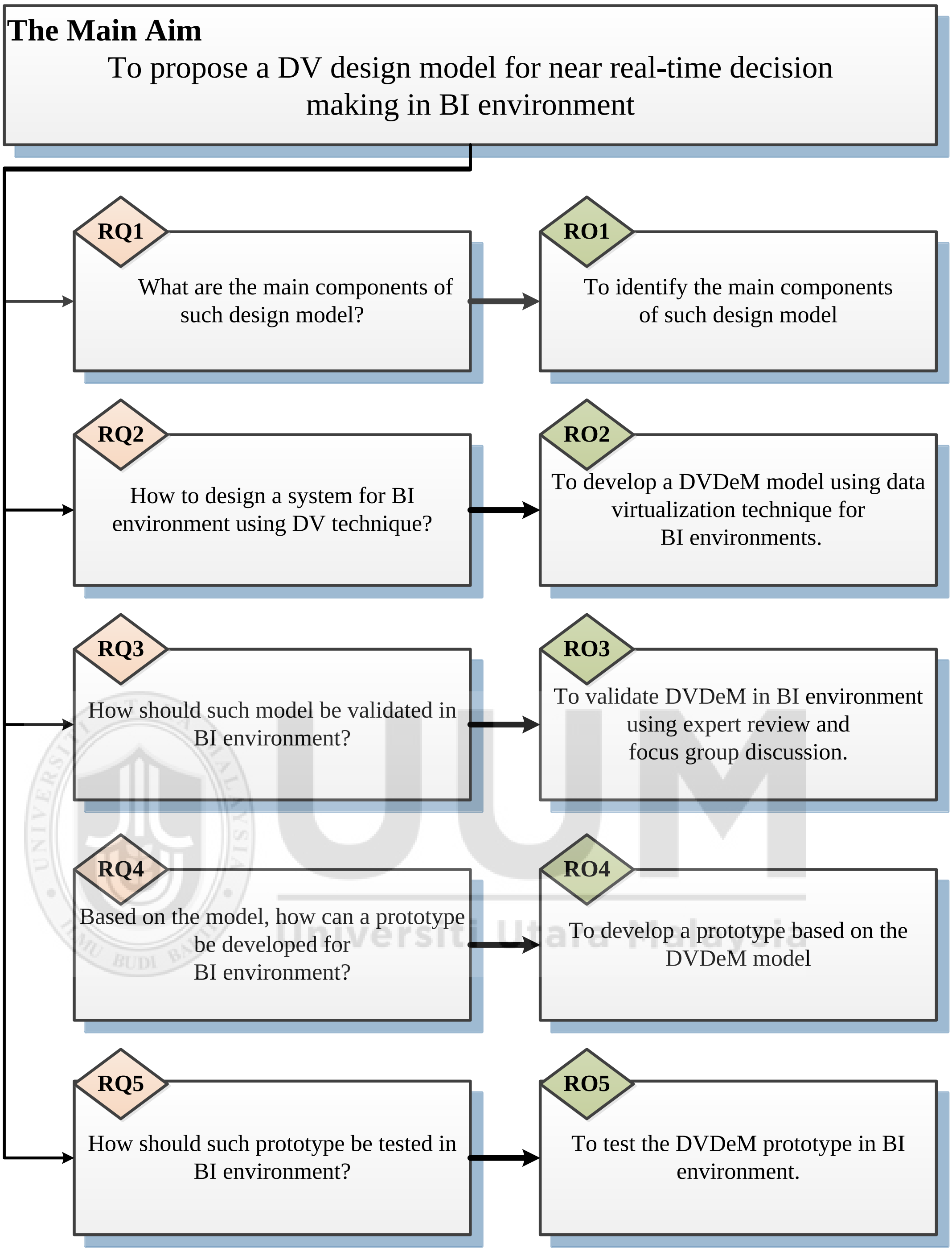


Figure 1.2. Research Questions and Research Objectives Mapping

1.8 Research Scope

This study focuses mainly on design and development of BI model using DV technology. Therefore, to address the scope of this study, two principal perspectives of BI need to be distinguished. The first one is the technical perspective while the other is the business perspective (Olszak & Ziemba, 2010). The technical perspective includes: an integrated set of techniques, tools, and software that is used to extracting heterogeneous data that resides in a variety of sources then integrate and analyze these data and make it available to decision makers. The business perspective (organizational perspective), on the other hand, includes the BI methodology which refers to the philosophy that deals with information and knowledge. The knowledge is then shared along with the analytical approach to business processes in organizations. The motivation that stands behind this research is to design a new or enhanced version of the traditional BI environment.

Decision making systems have to develop over time to reflect the changing environment they operate in (O'Leary, 2008). This change can occur in technical areas, like new and improved database systems, new communication technologies or organizational change like shifting user needs, new and more complex problems or problem understanding and similar. Besides, the motivation behind this research is to help to evolve BI by extending current technologies to better support businesses in their decision making. Therefore, in the context of this study, the focus is on technical perspective of BI especially data sourcing stage.

Universiti Utara Malaysia

1.9 Significances of the Study

The major purpose of this study is to propose data virtualization design model for near real-time decision making in BI environment. The proposed model includes phases, components, and activities for designing any BI application using DV as data integration (as illustrated in Chapter 4). Meanwhile, the proposed model has its individual components and characteristics as it provides specific guidelines on developing BI environment includes various theories and concepts such as data integration theory (physical and virtual), Decision Making Theory, View Based Data Integration Theory, and Work System Theory (as illustrated in Chapter 2). The proposed model with its relevant concepts could be significantly utilized for future research by academics, future BI developers, and future instructional development by instructional practitioners.

Moreover, this study increases the current literature by providing a research and theoretical framework that could be adopted to examine potential related theories, concepts, and issues for future studies.

1.10 Theoretical and Research Framework

This study is carried out based on theories and concepts related to BI environment which focus on data integration and design model using DV technology. Therefore, the research framework comprises three main steps, i.e., identification of problem, model development and evaluation. Figure 1.3 visualizes the research framework for this study, which is based on three phases are associated with sub-phases and detailed activities in achieving the objectives as will be discussed in Chapter 3.

In the first phase (Problem Identification) which contain two sub-phases (awareness of problem and suggestion). The research problem and scope are identified. A literature study and content analysis regarding concepts and theories were performed in identifying the components and phases of the proposed model. Additionally, comparative analyses on existing studies has been carried out to confirm the research gap as well as identifying all techniques used to integrate data used in BI environments.

As for the second stage (model development) which contain one sub-phase (construction) the theories and concepts of existing BI and DV techniques that have

Universiti Utara Malaysia

been reviewed was used as the basis to determine the components of the design model for BI environments. The outcomes from this phase are model development (DVDeM), which are validated iteratively using two validation methods (expert review and focus group discussion) to reach to final version of the (DVDeM).

In the third stage (model evaluation) which contain two sub-phases (evaluation and conclusion), the prototype based on the (DVDeM) was developed, prototype implementation in two case studies was conducted, and the usability test on DVDeM prototype was conducted. Besides, the analyzing the finding and report writing and research publication was conducted.

Universiti Utara Malaysia

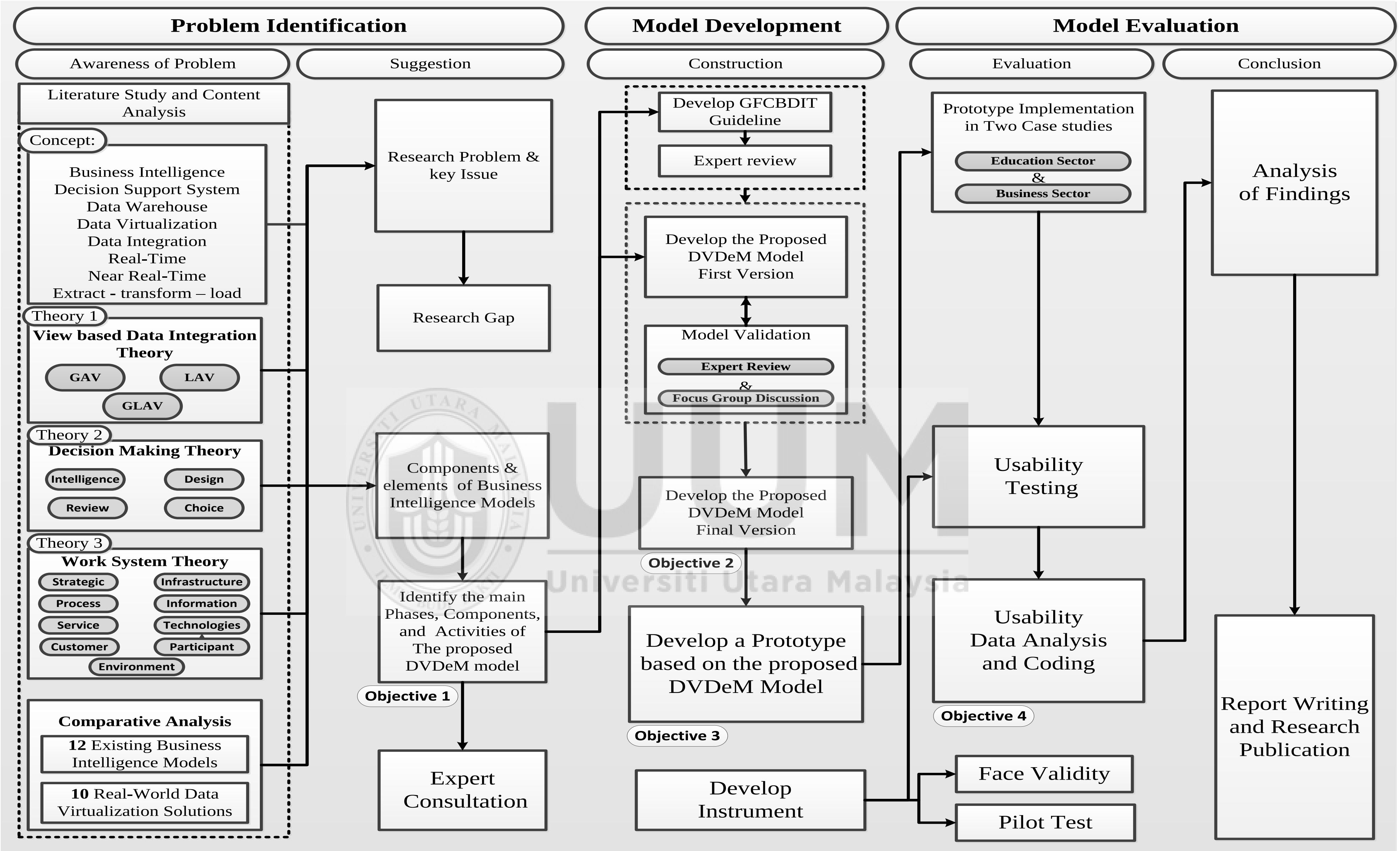


Figure 1.3. Theoretical and Research Framework

1.11 Contributions of study

This study contributes generally to the body of knowledge and industry within the specific area of BI environment. Generally, this study focuses on the design and use of data virtualization technique in facilitating and supporting the decision making process in BI environment. This study proposes the design and implementation of a comprehensive model that is practical and easy guide to developers of business intelligence. The proposed model also contains most of the important steps to be followed in the design of BI systems. In particular, the contributions of this study address the problems which are discussed in section 1.4 and will answer the questions aforementioned in section 1.6 in order to achieve the objectives as stated in section 1.7. Therefore, the detailed list of contributions is as follows:

1.11.1 Contribution of the Research to the Academics

Through a detailed literature review and empirical findings, this research contributes to knowledge with its proposed DVDeM model as a guideline for the developers and academics to be used for developing BI applications using DV as a data integration technique technology in BI environment, in addition to other contributions that associated with. The detailed list of the contributions to the academics is as follow:

1.11.1.1 Design and Development of DVDeM

This study proposes a DV design model of for near real-time decision making in BI environment, which is called DVDeM. The BI designers and developers would get benefits from this model, in which they can refer to develop BI applications that use DV as a data integration technique in BI environment. The model provides comprehensive guidelines which consist of 3 main phases which are DV gathering

Universiti Utara Malaysia

requirements (which consists of two sub-phases; organization and business requirement and data sources requirements), DV development (which consist of also two main sub-phases; data preparation and data connection and data manipulation and data management), and DV presentation, as well as the other components and flow between them. Besides, one of the most important components in the proposed model is to create a virtual table that contains the relevant and near real-time data that can be used by BI tools. Normally, a virtual table consists of the huge amount of data, in order to make a virtual table contain the relevant data, therefore, the Goal based approach (GODV) for virtual table has been proposed. This approach is discussed in greater detail in chapter 4, section 4.3.2.3. In line with above situation, the proposed DVDeM model equally avails the academics with the opportunity of guideline for developing BI applications in BI environment.

Universiti Utara Malaysia

1.11.1.2 Comparative Study of Existing Models and Approaches

In finding out the core phases and their components of DVDeM model and its prototype, two comparative studies of existing BI models and approaches as well as real-world DV solutions were conducted. These comparative studies compares the current studies which proposed by several researchers (whether these studies have focused on integrate the data via DW or DV) in identifying the important phases and components by explaining the focus of the studies and therefore. These comparative studies should provide significant analysis to other researchers and will further provide the research basis for future studies.

1.11.1.3 Evaluation Instrument of the Proposed DVDeM model

The evaluation instrument (Q-U) for the DVDeM model, has been adapted and validated through factor analysis and experts. Later, the instrument has been used to test the usability of DVDeM prototype. In the instrument, six evaluation dimensions were proposed to test the BI prototype usability namely; visibility, flexibility, learnability, application behavior, error control and help, and business intelligence. These dimensions were collected from previous literature which considers the criteria of good development and tested BI prototype. The instrument was found highly reliable in the pilot study with Cronbach's Alpha for each dimension is greater than 0.7. Hence, the validated instrument is allowed to be adapted by future researchers who study the same field.

1.11.2 Contribution of the Research to the Industries

The main objective of this study is to propose data virtualization design model for BI environment. This model consists of phases and components which cater for the decision makers by integrating data in a virtual manner in supporting decision-making process by deliver near real-time data. As demonstrated in the comparative analysis of real- world data virtualization solutions and existing BI models, as well as the development of the prototype based on DVDeM which has been tested with two case studies, therefore, In line with the above, this study has provided the business justification and technical roadmap required for effective adoption of data virtualization. The following are the list of significance:

Universiti Utara Malaysia

1.11.2.1 The Organizations are New to the Data Virtualization

If the organization is new to data virtualization, this study demonstrates two comparative studies based on real-world DV solutions as well as existing BI models and approaches. By using the proposed model as well as the prototype that have been developed based on DVDeM, therefore, this study is significant to help organizations move beyond traditional data integration techniques and use DV to improve their organization's business agility.

1.11.2.2 The Organizations are Already Adopting Data Virtualization

If the organizations are already adopting data virtualization, this study will help the organization to successfully accelerate and expand their adoption, compound their business agility gains and achieve additional business and IT benefits from data virtualization. The guideline derived from this study also facilitates in terms of maintenance because study propose a model and this model can be guideline in terms of the need of add new data sources or remove data sources as well as the organization can know the main components of the system and thus can avoid any mistakes.

1.11.2.3 Prototype Based on the Proposed DVDeM Model

The prototype of DVDeM has been developed for the validation process can be used in BI environment. This prototype could be used for the organizations that implement or intend to implement BI system.

1.12 Operational Definition and Terminologies

In the context of this study, some of the definitions of most used terminologies throughout this study that are related to the topic discussed were described, which may have been used in other study but conveying different interpretation. Besides, the other definitions are listed in Appendix E.

Data Warehouse (DW): a data warehouse is a subject-oriented, non-volatile, integrated, and time variant collection of data in favor of decision making.

Data Virtualization (DV): Is an approach to manage data by allowing other applications to retrieve and process data without the need to know any details about these data.

Business Intelligence: The application retrieves from structured, unstructured, internal and external data that is stored in a large variety of data sources, to gain the intelligence meaningful information.

Universiti Utara Malaysia

Decision Support System (DSS): is a computer-based information system that supports business or organizational decision-making activities.

Real-Time: A term used to describe computer systems that update information at the same rate as they receive data.

Near Real-Time: In the context of this research, the term near real-time mean; the data processing systems are to slightly slower than real-time and based on organization requirements and organization interpretation for near real time. Consequently, the next paragraphs discussed all types of data processing.

Wrapper: In information technology, a wrapper is data that precedes or frames the main data or a program that sets up another program so that it can run successfully.

Mapping: In computing and data management, data mapping is the process of creating data element mappings between two distinct data models. Data mapping is used as a first step for a wide variety of data integration tasks which include data transformation or data mediation between a data source and a destination.

Design Model: Design model is a set of propositions which expresses the relationship between components or concept. In this study, a design model is defined as a combination of components and processes that makes up a model to design a system.

1.13 Thesis Structure

This thesis comprises six chapters in total. The whole contents of each chapter are outlined as follows:

Chapter 1: Introduction – As an introductory part, this chapter provides some background of study that underlines detailed motivation of the study, the issues and problem that discuss on the scenario and justification which then lead to the selection of the research topic. In supporting the research problem, this chapter also discussed the formulation of research gap, research questions, and research objective. To avoid any misleading perception, the scope, contribution, operational definition and terminologies are also provided in this chapter, as well as a chapter summary.

Chapter 2: Literature Review – A systematic and in-depth review of concepts and theories that are needed before beginning with determining the appropriate proposed model components. It is important to ensure the proposed data virtualization model in BI environment is corresponding to the components needed and complies with the research objectives. Thus, reviews on concepts and theories underlying this study are expressed comprehensively in this chapter, as well as the two comparative analyses for existing's BI models and approaches. This chapter also deliberates on real world DV solutions, and how it is applicable to this study.

Chapter 3: Research Methodology – This chapter comprises of a step-by-step process from the beginning to the end on how the objectives are achieved. Overall, this study is accomplished through adopting the design science research joined with five sub-phases adapted from Design Science Research Methodology, to become the root for the whole of the study. Besides, the testing instruments provided for this study are discussed at length and expansively in this chapter. Overall, this chapter discusses the design and development of the proposed model.

Chapter 4: Model Development – The work involved in achieving objective one, two and three are discussed in detail in Chapter 4. It explains the process engaged in identifying the appropriate phases and components of the proposed model. It also discusses the steps involved in the development of the proposed model (first version). This chapter also presents how the study is validated through expert review and focus group discussion. Besides, based on validation finding, the proposed refine and modified and therefore, model (final version) has been developed, and later the prototype base on the proposed model (final version) has been developed.

Chapter 5: Prototype Implementation in Case Studies – In achieving the fourth objective of this study, the prototype which is developed based on the proposed model is implemented through two case studies. A usability test is conducted, the findings are discussed.

Chapter 6: Discussion and Conclusion – Finally Chapter six deliberates on the findings of this study by answering all the research questions and research objectives. It also concludes the whole work in this study. Besides, it discusses the contributions of the study to the body of knowledge, as well as highlighting recommendations for future directions of the study.

1.14 Summary

This chapter explains the overview of the research background. The problem statement, objectives, motivation and expected contribution of the research are highlighted. The expected outcome is explained to emphasize the benefits of the research. To summarize all research questions and objectives, the Object-Method Table (OMT) is used (Shiratuddin & Hassan, 2010), where all questions of research and its corresponding objectives are illustrated in Table 1.1. Additionally, methods that are used to achieve these objectives and expected outcomes for each objective are mentioned.

Table 1.1
Summary of Research Work

Main Objective	**Specific Objectives**	**Methods Used**	**Outcomes**
To propose a DV Design Model for Near-Real Time Decision Making in Business Intelligence Environment (DVDeM).	✓ To identify the components of DVDeM. ✓ To develop DVDeM using DV technique for BI environments. ✓ To develop prototype based on DVDeM. ✓ To test the DVDeM prototype in BI environment.	✓ Literature Review. ✓ Research based theory. ✓ Comparatives analysis. ✓ Relational database. ✓ DV technique. ✓ DW technique. ✓ Prototyping. ✓ Expert review. ✓ Focus group Discussions. ✓ Case studies.	✓ (DVDeM). ✓ Q-U Instrument. ✓ DVDeM Prototype. ✓ GODV Approach

Universiti Utara Malaysia

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

Decision Support Systems is an important aspect in the Information System (IS) Research. The DSS has experienced rapid growth and will continue to grow. The growth can be seen within the DSS environment as well as the number of organizations that embrace DSS design, techniques and implementation in their daily operations. It is therefore imperative that DSS constantly adapt to changes (Burstein & Holsapple, 2008; Reynolds et al., 2008). On the other hand, Edmunds and Morris (2000) argued that data that were used as an input of DSS is widely available in organizations, yet transforming and integrating these data into meaningful information remains a challenge. Barone (2010) argues that this problem still exists and that today's BI systems do not address business needs sufficiently.

Universiti Utara Malaysia

The motivation that stands behind this research is to design a new or enhanced version of the 'traditional' BI concept. To do so, this chapter describes the kinds of literature and theories related to this study, such as BI, DSS, DW, DV and the three related theories; First, this chapter reviews Decision Support Systems (DSS) and Business Intelligence. Then the work reviews data integration techniques which include Data Warehouse (DW) and Data Virtualization (DV). This thesis also reviews three research theories namely, View Based Data Integration Theory (VDIS), Decision Making Theory (DMT), and Work System Theory (WST). The focus of this chapter is to review the existing studies of data integration in BI environment in order to identify the gaps and understanding the implications of these studies for proposing data virtualization design model in BI environment.

The concepts of BI and DSS are linked to each other closely. In Section 2.2, the definition of each of these terms with some depth and detail as well as how they will have used in context of this research is given. Finally, through an in-depth review of the current issues and approaches, we present the research gap that forms the aim of this thesis.

Data integration techniques, be it in a physical manner like data warehouse (DW) or virtual manner like data virtualization (DV), is the backbone for BI systems. It is the motivation behind this research to leverage this data integration technique and its characteristics in BI systems to better align BI with business demands. Some of the most important features of this technique are shown in detail and the reasons that lie behind the fact that this technique is relevant in section 2.3. Section 2.3.6 presents an overview of DW. While section 2.3.7 presents an overview of DV.

In the context of this research, based on a selective literature review and some of the author's recent papers, a unifying theoretical approach of the most relevant BI specific concepts has been initiated. The review concludes that there is a need to improve on existing BI models and approaches, in terms of data integration as part of BI applications to be used by developers in developing BI environment, and also to consider the features of inclusive the core components in the proposed model. Figure 2.1 illustrates the overview of the literature review.

Universiti Utara Malaysia

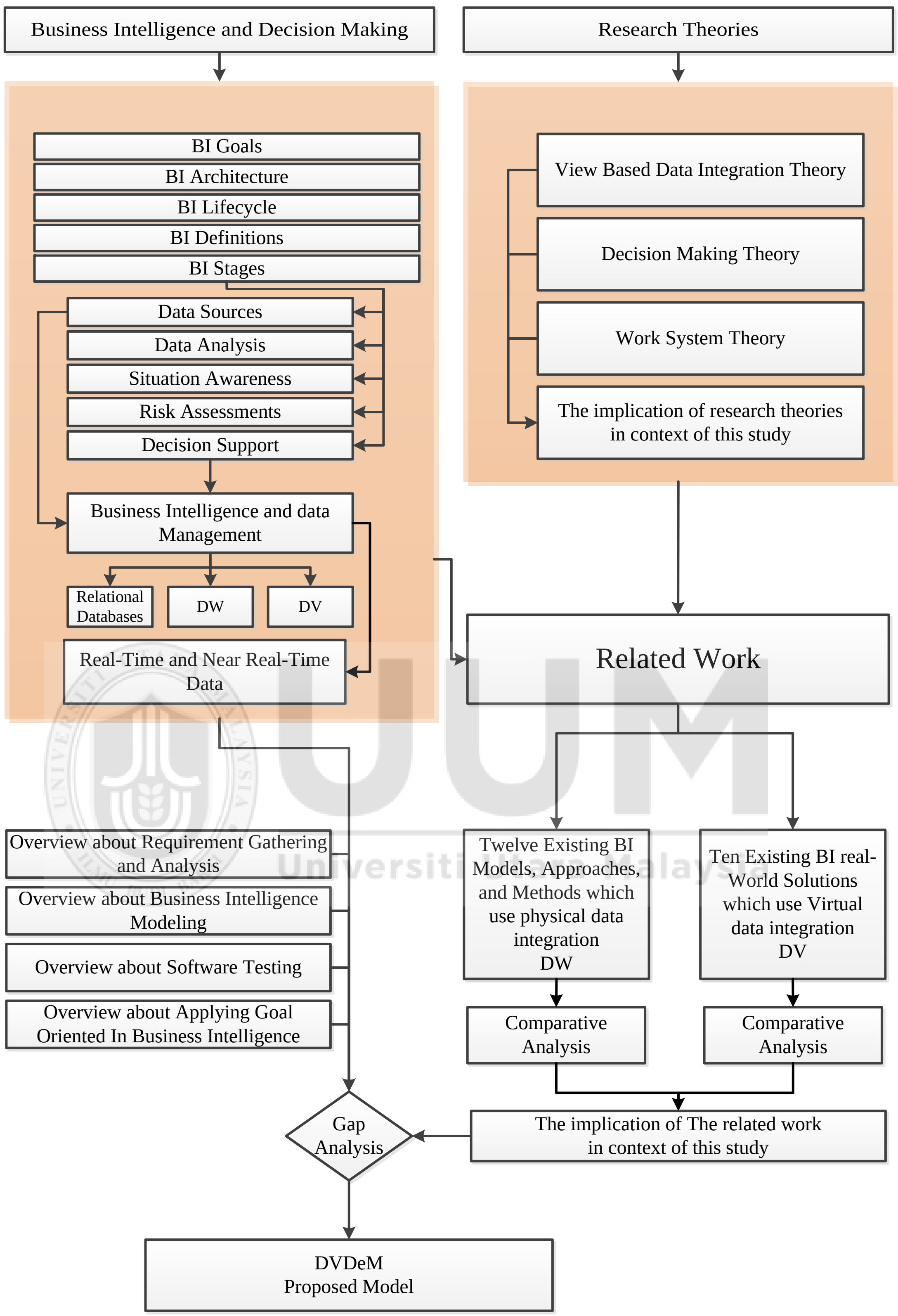


Figure 2.1. An Overview of Literature Review

2.2 Research Theory

A theory is a set of interrelated concepts, definitions, and propositions that explains or predicts events or situations by specifying relations among variables. A theory explains how a particular aspect of human behavior or performance is organized. Therefore, a theory enables us to deliver predictions about this behavior. Generally, a theory consists of two main ingredients which are concepts and principles (Moore, 1973). However, this study was conducted based on three theories namely, View-Based Integration theory, Work System theory, and Decision Making theory. In the next paragraphs, a brief description about each of them was given.

2.2.1 View based Data Integration Theory

Data Integration (or Information Integration) is the problem of obtaining and extracting data from different both heterogeneous and homogeneous sources. View based Data Integration (VDIS) is a theoretical framework that solves the data integration problem for structured data by integrating sources into a single unified view. However, the first appearance of VDIS was in the form of multi databases and federated systems in 1986 in the form of multi databases and federated systems (Landers & Rosenberg, 1986).

2.2.1.1 VDIS Architecture

In general, VDIS Architecture consists of four core stages namely; sources, wrappers, mediator, and applications as shown in figure 2.2. The sources are responsible for storing all types of data which normally exists in a heterogeneous manner, while the main function for wrappers is to solve the heterogeneity in the formats by transforming heterogeneous data formats into common formats in the data integration systems. Noteworthy, the wrappers refer to sources database (local schema). The mappings process is expressed in a mapping language to determine the relationship between wrappers and unified view which is exported by the mediator (global schema). Finally, the main task for applications is to retrieve data from the sources in an indirectly manner by querying the global schema. Noteworthy, the mediator's task is to conduct the mappings to determine which data to retrieve from the sources and how to consolidate these data in a proper way in order to make the answer the queries.

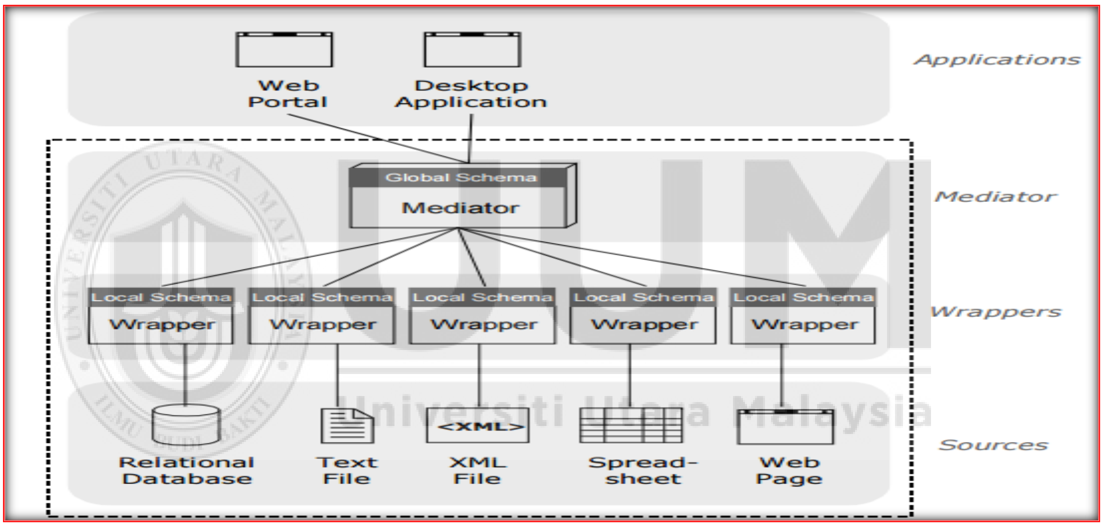


Figure 2.2. VDIS Architecture

Source: (Landers & Rosenberg, 1986).

2.2.1.2 VDIS Classification

VDIS can be classified based on three core axes:

i. Query language with common data model (data which exposed by wrappers to mediator and from mediator to application).

ii. Mapping Language: Global As View (GAV), Local As View (LAV) and Global and Local As View (GLAV). Due to being one of the most important components in a VDIS, The elaboration of each of them in the next paragraph.

iii. Data storage: it's responsible of determine the location where data are stored.

2.2.1.3 VDIS Mapping Categories

The VDIS mappings are expressed in a language, corresponding to some class of logic formulas. Languages proposed in the literature fall into three categories; GAV, LAV and GLAV (Katsis & Papakonstantinou, 2009). Accordingly, all of these languages categories will be further explained in the next section.

2.2.1.3.1 Global As View

In Global As View (GAV), and according to Katsis and Papakonstantinou (2009), the data integration system is constrained to the set of tuples mapped by the mediators while the set of tuples expressible over the sources may be much larger and richer. Figure 2.3 illustrates an example of GAV.

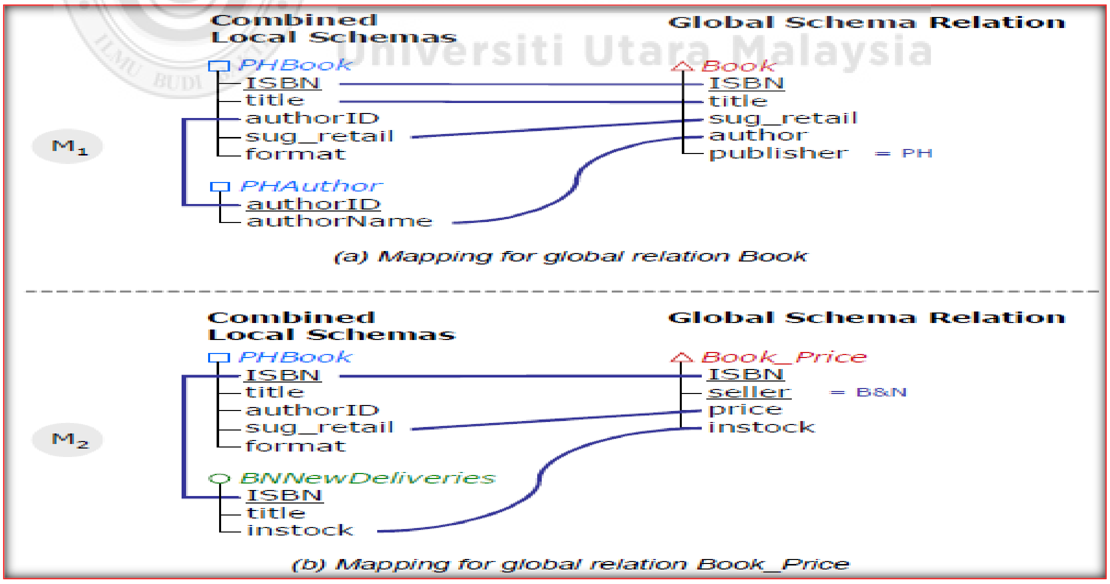


Figure 2.3. The Example of GAV Mapping

Source: (Katsis & Papakonstantinou, 2009).

2.2.1.3.2 Local As View

In Local As View (LAV), the data integration system is constrained to the set of tuples in the sources while the set of tuples expressible over the global schema can be much larger. Therefore, LAV systems must often deal with incomplete answers. Figure 2.4 illustrate an example of LAV.

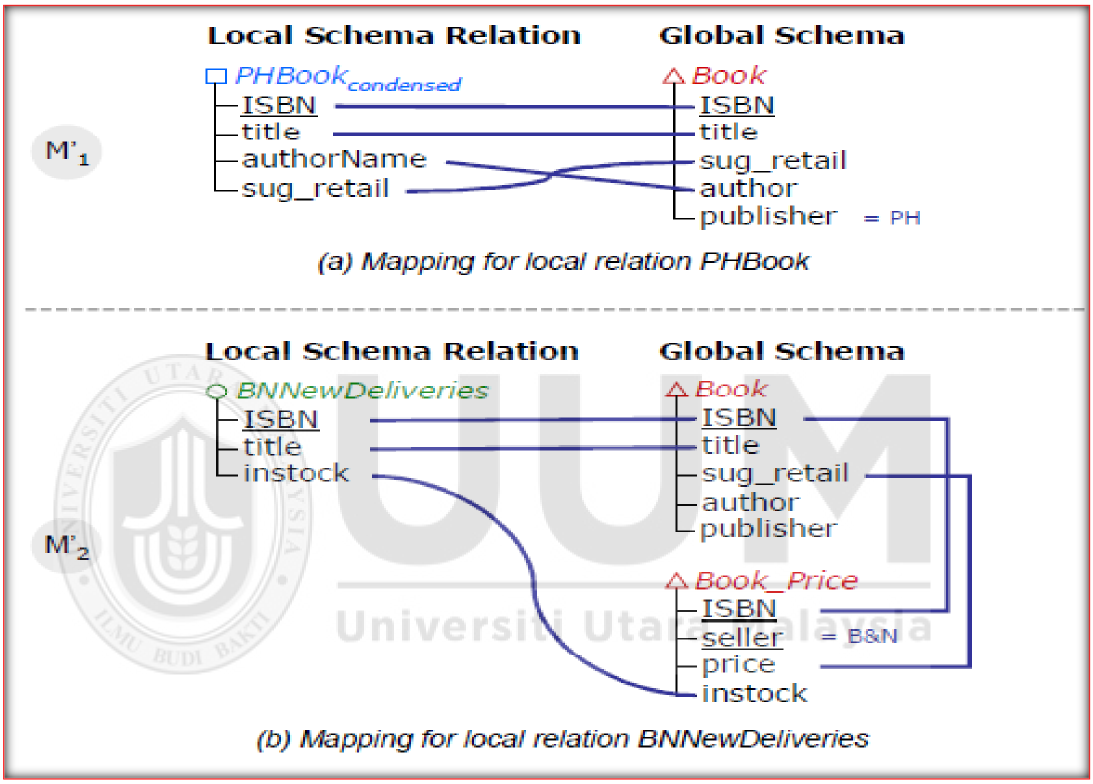


Figure 2.4. The Example of LAV Mapping

Source: (Lavery, Cockton, & Atkinson, 1996).

In line with the above situations, it has been realized that the difference between GAV and LAV is GAV mapping from entities in the mediated schema to entities in the original sources while LAV mapping from entities in the original sources to the mediated schema.

2.2.1.3.3 Global and Local As View

To overcome the limitations of both GAV and LAV, a new category of mapping languages called Global and Local As View (GLAV) is proposed by (Yu & Popa, 2004). The data integration system with GLAV is a generalization of both GAV and LAV, and is illustrated in Figure 2.5.

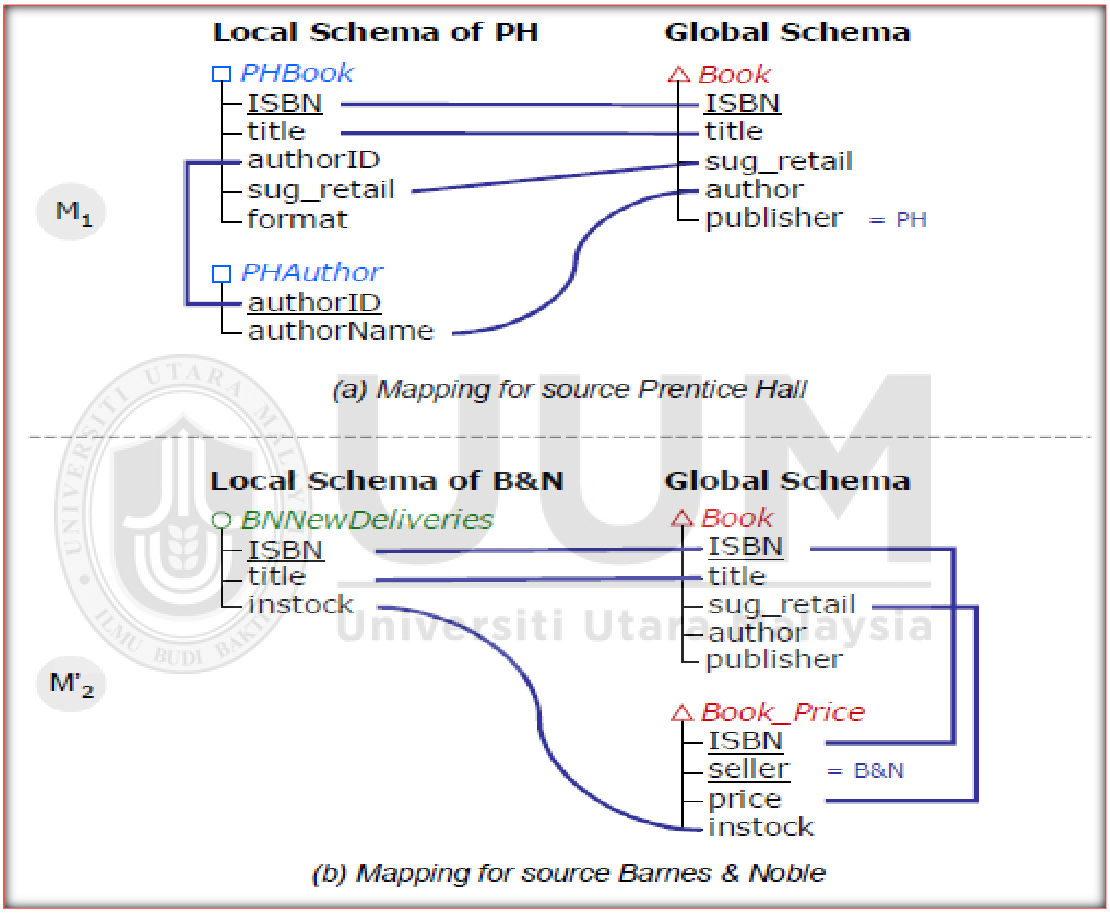


Figure 2.5. The Example of GLAV

Source: (Yu & Popa, 2004).

2.2.2 Decision Making Theory

The core purpose of the DSS is to support the process of decision-making in an organization. This assumes that the way in which decision-making actually takes place in the organization is fully understood. There are many models of decision-making. People with a background in the quantitative analysis would typically have been exposed to rational decision-making methods, such as Simon's (1960) four-step decision model that incorporates intelligence, design, choice and review as visualize in Figure 2.6. This process is often accompanied by the calculation of the subjective expected utility (SEU) or another way of ranking alternatives to facilitate choosing the best option (Newell, Allen, Simon, & Alexander, 1972; Simon & Herbert, 1960).

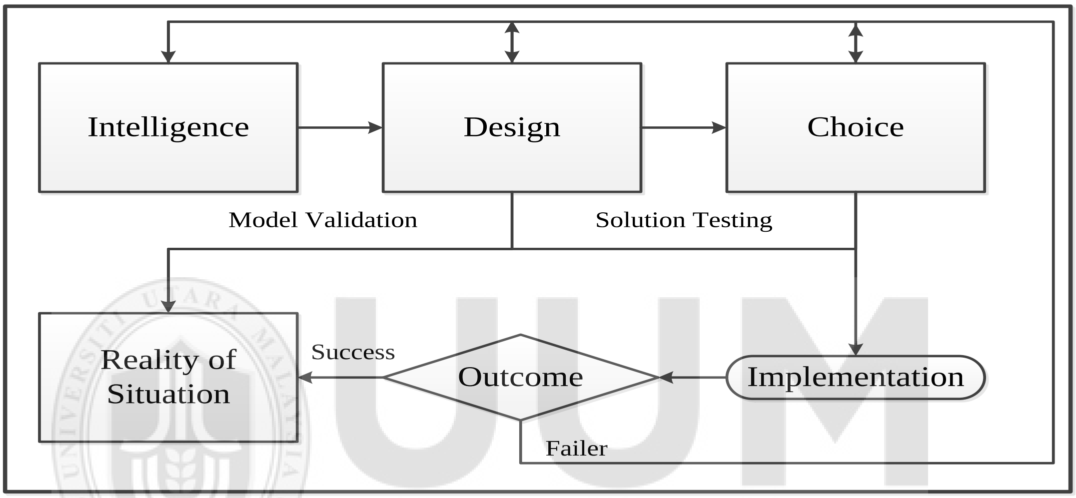


Figure 2.6. Simon's Decision Model

There are many views and theories related to and support decision making process can be found in the literature and can use to support this research, are presented in Table 2.1.

Table 2.1
Decision Making Models

	Model Name & Author	Description
1	The rational model for decision-making (Craik & Leibovich, 1976).	The process of rational decision making comprises a number of steps: ✓ Intelligence: finding occasions for making a decision; ✓ Design: inventing, developing and analyzing possible courses of action; ✓ Choice: selecting a particular course of action from those available; and ✓ Review: assessing past choices.
2	The multiple perspectives approach (Churchman, 1971).	This approach classifies perspectives as either being technical, organizational or individual in nature.

Universiti Utara Malaysia

Table 2.1 Continued

3	The organizational procedures view (Cohen, March, & Olsen, 1972).	The organizational steps view required to grasp decisions as the output of standard operating procedures mentioned by organizational sub units.
4	The individual differences perspective (Keen & Morton, 1978).	This model focuses on the attention to the problem-solving behavior of the individual manager, as influenced by the manager's decision-making manner, background and personality. It tries to demonstrate how managers may use various means or come to various outcomes because of differing personalities.

It can be seen from Table 2.1, the multiple perspectives view classifies perspectives as either being technical, organizational or individual in nature. Analytical models that collect data as a basis for understanding the system would all fall under the technical perspective. Different analysts or modeling projects will come up with different technical views, even if these projects claim to present an objective or rational picture of the situation. Thus, it is encouraged that more than one technical view of a system is obtained. In order to cover the organizational and individual perspectives, as many as possible of the role players and stakeholders should be investigated. Data collection is also to follow the "sweeping in" approach, and especially the organizational and technical perspectives data need to be gained in multiple modes and from as many sources as possible.

2.2.3 Work System Theory

Work system theory (WST) is the collection of opinions that determine the basis of the work system method (WSM) for analyzing and designing systems. WST is a lens for thinking about any system in an organization. Generally, WST consists of three core ingredients (Bostrom & Heinen, 1977):

i. The definition of work system: a system in which human participants and/or machines perform work (processes and activities) using information, technology, and other resources to produce specific products and/or services for specific internal and/or external customers.

ii. The work system framework: a static view of the work system as it exists during a particular time interval when it retains its identity and integrity even though it may change slightly through small adaptations, workarounds, personnel changes, and even unintentional drift as shown in Figure 2.6.

This framework is a pictorial representation of a work-system in terms of nine elements included in a basic understanding of the work system's form, function, and environment during a period when it is relatively stable, even though incremental changes may occur during that period (Bostrom & Heinen, 1977).

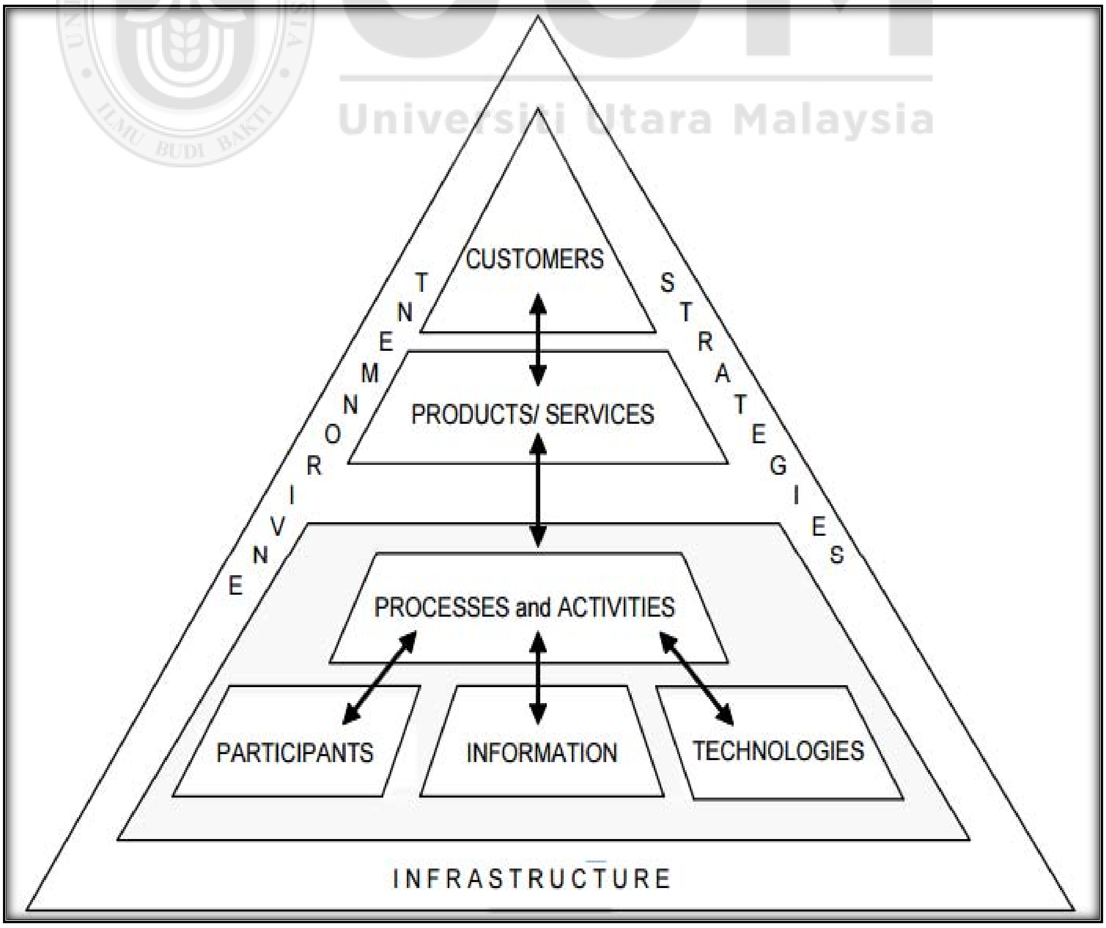


Figure 2.6. The WST Framework
Source: (Bostrom & Heinen, 1977).

iii. The work system lifecycle model: Bostrom and Heinen (1977) have defined work system lifecycle as a dynamic view of how work systems change over time through a combination of planned and unplanned change. Figure 2.7 illustrate WST lifecycle.

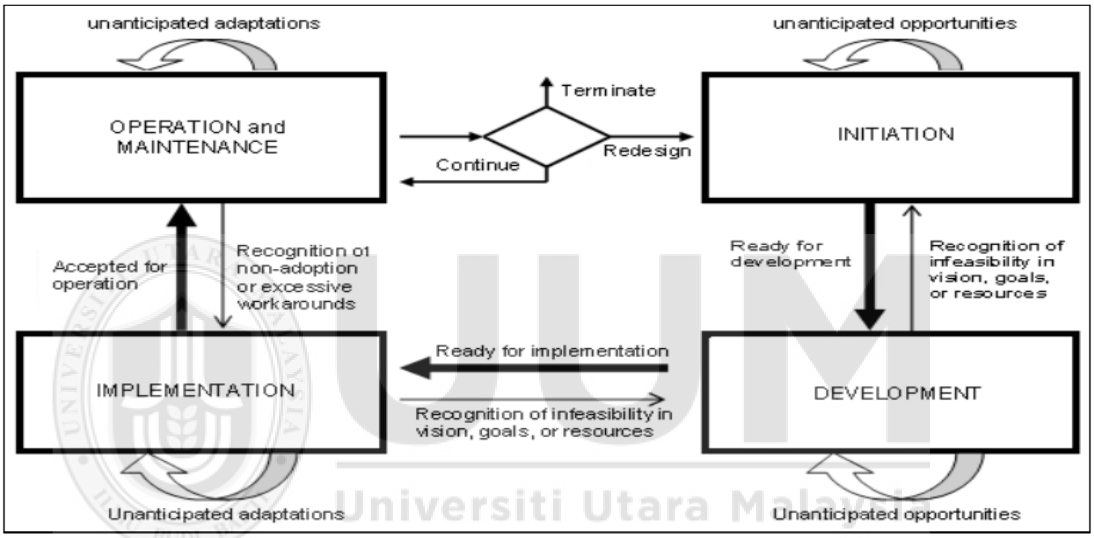


Figure 2.7. The WST Lifecycle

Source: (Bostrom & Heinen, 1977).

2.2.4 Implication of Research Theory to Study

There are three research theories adopted in this study, view-based data integration theory, decision-making theory, and work system theory. Since the study is aimed to propose DV design model for near real-time decision making in BI environment, then all of these the are suitable theories. They explain how data integration can be well-designed and how it should be applied in BI environment in supporting decision-making process. Besides, the principles used in these theories will be useful for the proposed DVDeM model and determine the main model components. Accordingly, all the mentioned theories are important and have been applied in the proposed model. They have become the guideline to propose the DV design model for near real-time decision making in BI environment. Furthermore, in the context of this study, an overall overview of BI environment in identifying the lifecycle, goals, architecture, process stages and phases of the BI have been reviewed. The overall overview of BI is described in the next section.

2.3 Business Intelligence and Decision Support Systems

BI and DSS are two terms that are related and are used interchangeably. In some cases, these terms are used to describe an evolutionary stage in software concepts that is used to support organizational decision making. The field of DSS started more than 60 years ago by Simon (1960), Anthony (1965), and Gorry and Scott Morton (1971). Today, decision support process is considered as one of the core research in the fields of information systems (Burstein & Holsapple, 2008).

Universiti Utara Malaysia

As mentioned in the previous paragraph, the concept of BI emerged more than five decades ago. In 1985, Rhines proposed idea to transfer artificial intelligence (AI) techniques from laboratories to businesses. He argued that computers can emulate some aspects of human intelligence and simulations can be used in business. Rhine further noted that both expert systems and human experts possess the same way to capture knowledge and is capable to use this knowledge to provide solutions in specific situations. From the same perspective, this idea has been supported by Herring (1988), saying that AI techniques have ability to be fruitful in the industry, commerce and government to help achieve competitive advantage in those areas.

The term BI appeared for the first time in an IBM journal article in 1958. It has been used to describe the automatic systems, which will distribute information to any sections of the organization, such as the scientific and industrial sections. BI can be defined as “a broad category of software and solutions for gathering, consolidating, analyzing and providing access to data in supporting decision makers to make better business decisions”(Karim, 2011b).

As mentioned in Chapter 1, there are two views to describe BI, as shown in Figure 2.8. The first perspective is called technical perspective which includes all the applications and tools that are used to support the functionality of BI for any organization. The second one is business perspective which includes customers, competitors, and competitive environments.

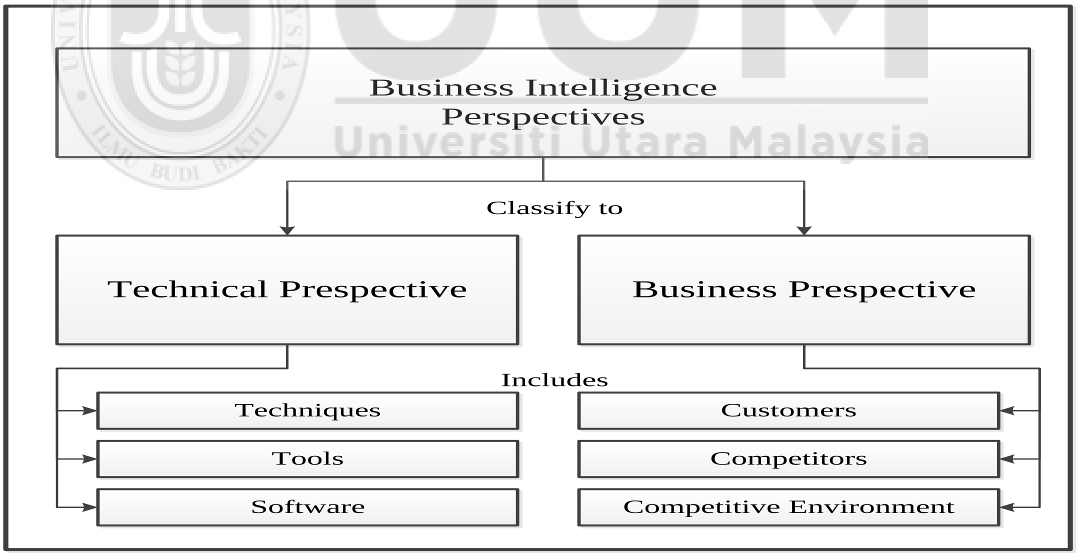


Figure 2.8. Business Intelligence Perspectives

Both technical as well as business perspectives are important in BI. While business perspective of BI remains important, it is not within the scope of this research. This research focuses is on the technical perspective of BI.

BI is used extensively in different areas in an organization (such as management and human resources, marketing and finance). Through the literature reviewed, we can infer that the main aim of building and applying BI system is to get meaningful information through the extraction of data from diverse sources and turning it into information. The information will then be transformed into knowledge to enhance the decision-making process in any organization. Due to changes in the needs of researchers and the requirements of organizations, new definitions for BI have emerged, and in spite of that, all the definitions put into consideration the importance of the information; therefore, new definitions of BI may emerge with the developments in the IT industry as well. Table 2.2 shows some of these definitions.

Table 2.2
The Business Intelligence Definitions

	Business Intelligence Definition	**Authors**
1.	To constantly monitor and assess the emerging market, the players, and the instruments and forces of change.	(O'Brien & Fuld, 1991).
2.	It's a systematic process that collects, analyzes, and organizes the flow of critical information, focusing it on important strategic and operational issues.	(Thomas Jr, 2001).
3.	Business intelligence is a series of systematic techniques to collect, validate, analyze, and deliver public information and expert insights about the competitive environment to those in your firm who can act upon it.	(Fiora, 1998).
4.	BI is neither a product nor a system. It is an architecture and a collection of integrated operational as well as decision-support applications and databases that provide the business community easy access to business data.	(Moss & Atre, 2003).
5.	Business intelligence is the rational application of the principles of intelligence services to business. It is simply the collection, analysis, and application of strategic information to business decisions.	(Marren, 2004).
6.	Business intelligence systems combine operational data with analytical tools to present complex and competitive information to planners and decision makers.	(Negash, 2004)
7	The integration of core information with relevant contextual information to detect significant events and illuminate cloudy issues for management decision-makers. It includes the ability to monitor business trends, to evolve and adapt quickly as situations change and to make intelligent business decisions on uncertain judgments and contradictory information.	(Hill & Scott, 2004).

Table 2.2 Continued

8.	The decision making using data warehousing and online analytical processing techniques (OLAP).	(Herschel & Jones, 2005).
9.	The acquisition of information, often by the use of technology, about one's own operations and those of one's competitors.	(Management Today ,2006).
10.	The mixture of the gathering, cleaning and integrating data from various sources, and introducing results in a model that can enhance business decisions making.	(Karim, 2011b).
11.	knowledge about the business environment (e.g. markets, customers and competitors) and the organization (e.g. business performance and economic issues)	(Lönnqvist & Pirttimäki, 2006).
12.	A generic term to describe leveraging the organization’s internal and external information assets for making better business decisions	(Kimball & Ross, 2011).
13.	Business Intelligence is a strategic approach for systematically targeting, tracking, communicating and transforming relevant weak signs into actionable information on which strategic decision-making is based.	(Rouibah & Ould-Ali, 2002)
14.	Business intelligence means using your data assets to make better business decisions. It is about access, analysis, and uncovering new opportunities.	(Chang, Hussain, & Dillon, 2006)

Universiti Utara Malaysia

Accordingly, from the BI definitions aforementioned as seen in Table2.2, in the context of this study, BI have defined as “BI is an application and technologies that can take the data are derived from structured, unstructured, internal and external data that is stored in a large variety of data sources, integrate it gain the meaningful information, and deliver it to BI end users in supporting the decision-making process”.

It can be concluded from the definitions aforementioned as seen in Table 2.2, that information (knowledge) is the basic building block of the BI and can be considered the main requirement for the construction of BI. Therefore, BI can be described as knowledge acquiring activity in business. Likewise, Dobbs et al. (2002) elucidate that meaningful information (knowledge) it is the fundamental building block of BI. Having information and knowledge in hand will enable BI in decision-making process of the business, rather than using a basic structure with assumption in the decision-making process. The knowledge can be obtained by extracting from data and information from various sources. Knowledge may also be drawn from the experience (Loebbert, 2011; Tiwana, 2000). As such, it can be described as the amount of knowledge to take advantage of the information (Dwivedi, Papazafeiropoulo, & Metaxiotis, 2009; Loebbert, 2011).

In business, the torrent of information for the business aspects is supplied by data. The business aspects include people, products, operations and location. Then, this information is used to answering business inquiries like, “Which product is the best?”, and “How are my transactions and my customers doing?” Eventually, the decision-makers make decisions based on knowledge learned from the feedback they gathered. From above it can be concluded that, business information is the input to business knowledge, which is the input to business intelligence. Figure 2.9 illustrates BI input and output.

Universiti Utara Malaysia

Source: (Loebbert, 2011).

Figure 2.9. Business Intelligence Input & Output

In the context of this study, the technical perspective is related with the main aim. In the same aspect, BI refers to the process of extracting, transforming, managing and analyzing business data, in order to support decision making. This process is mainly based on huge data sets, particularly DW and is aimed at disseminating intelligence or knowledge across the whole organization. Therefore, businesses across different industries have focused on the utilization of information as an important aspect in generating BI systems.

As discussed extensively in the previous paragraphs, the main aim of this research is to propose DVDeM model in BI environment, and in order to obtain an overall overview about BI environment such as: BI goals, BI architecture, BI lifecycle, and BI stages should take into consideration. Consequently, the overall overview about each of them is given are further detailed in the next paragraphs.

2.3.1 Business Intelligence Goals

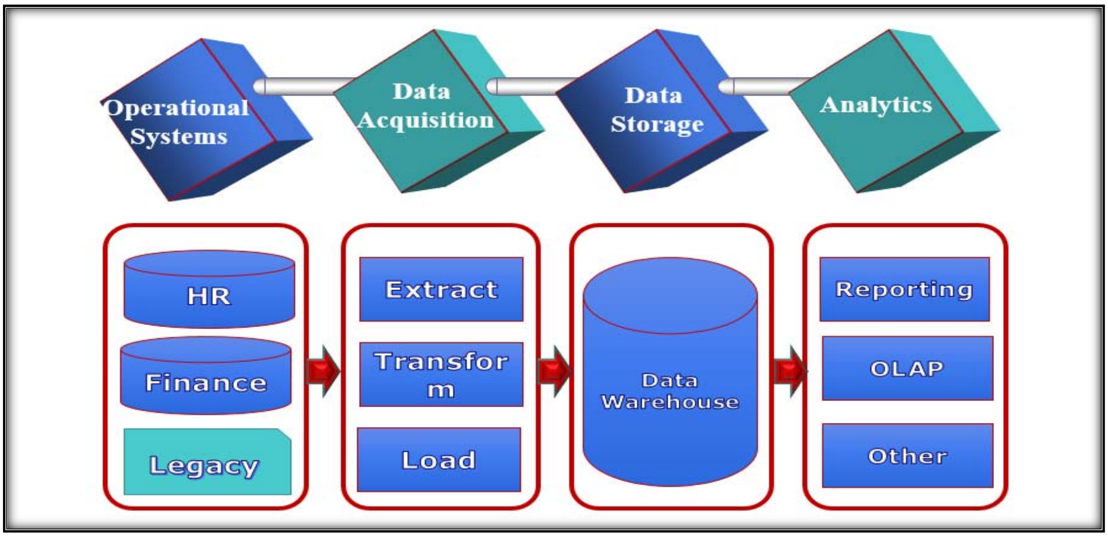
According to Thomas (2001), the key goals of BI are:

i. To makes the organization's top management expect predictions about their organizations to avoid surprises
ii. To make the organizations   can identify challenges and opportunities.
iii. Gives organizations a full understanding in the case of a deficit as well as shows the weaknesses of the organizations.
iv. Decrease the time for reaction.

In line with the above situations, it can conclude that data integration that is used as inputs to BI environment are mainly concerned with creation knowledge through information gathering and processing. Therefore, in the next section, the overall overview for all the tools that support BI environment were discussed, in order to identify the main components, steps, and techniques related to design BI model, also to find out the obstacles that the data suffered from the beginning used as input for business intelligence. Accordingly, this research aims to propose a DV design model in supporting decision making process by delivering a near real-time decision-making in BI environment; hence, the identification of BI components is vital for the development of the proposed model in this research.

2.3.2 The Architecture of a Business Intelligence System

Traditionally, BI systems consist of four levels of components, namely Operational Systems Level, Data Acquisition Level, Data Storage Level, and Analytics Level. Figure 2.10 visualizes the BI Architecture. These various components cooperate with each other for the purpose of facilitating BI functions These functions includes data extraction from data sources, whether these data homogeneous or heterogeneous sources, store this data in data warehouses, and eventually retrieve data stored for use in various business analysis applications (delivered to the data consumers) (Codd, Codd, & Salley, 1993; Inmon, 1996, 2005; Inmon & Hackathorn, 1994). In line with the above situation, it can conclude that these functions extract data from data sources and then deliver to the data consumers.



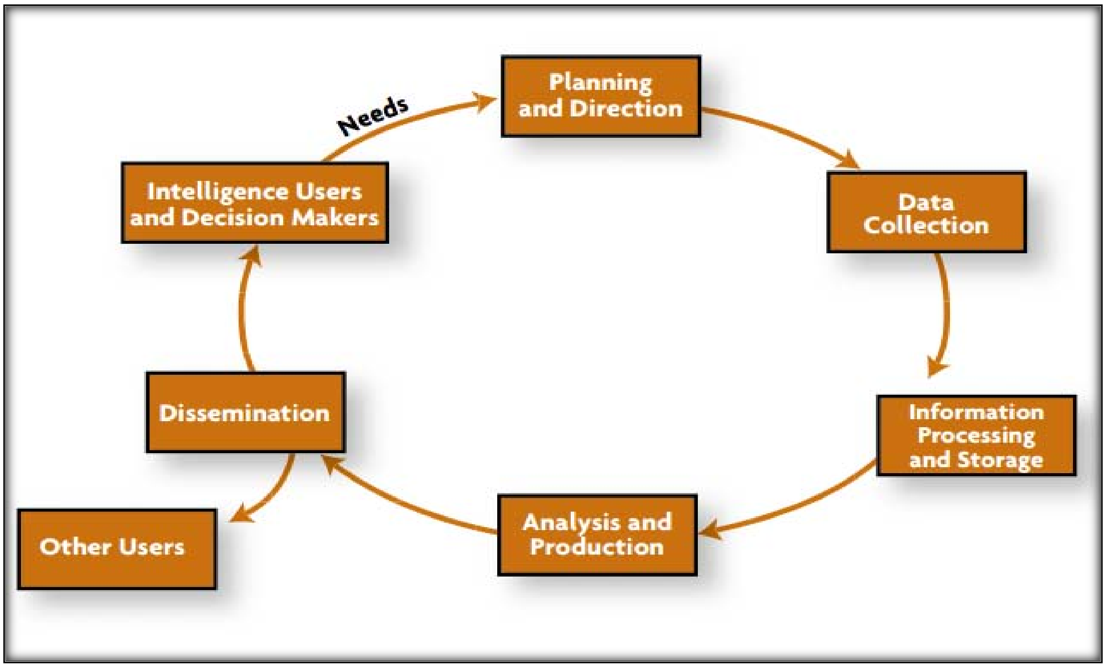
Source: (Inmon & Hackathorn, 1994)
Figure 2.10. Business Intelligence Architecture

In summary, BI architecture is a framework for organizing the data, information management and technology components that are used to build (BI) systems for reporting and data analytics. Accordingly, the principles in BI architecture will be useful for the design and develop the proposed DVDeM model and determine the main components in BI environment;

2.3.3 The Business Intelligence Lifecycle

According to Thomas (2001) and as illustrated in figure 2.11, the BI lifecycle consists of six key phases: the first phase is planning and direction phase, and it is commonly based on organization needs. The second phase is data collection where the required data is collected in an ethical manner. The third phase is information processing and storage which mean the information have to be more reliable and been extracted from valid data sources, analysis and production that's means these data have to analyze to intelligence based on the strategic planning, and business needs, dissemination which means eventually, for purpose to have intelligence value, these data should be disseminated (published) in very clear way and easy to monitor and understandable, and intelligence users and decision makers (Li, Shue, & Lee, 2008; Thomas Jr, 2001).

Universiti Utara Malaysia



Source: (Thomas, 2001).
Figure 2.11. The Lifecycle of Business Intelligence

In a nutshell, as seen in the Figure 2.12, the BI users and decision makers have connected to the project planning and directions, hence, the starting over of the BI lifecycle again. To gain best results following the BI lifecycle, BI developers may pay most attention to project planning and organizations requirement. Accordingly, the principles in BI lifecycle will be useful for the proposed DVDeM design model and determine the main phases in BI environment.

In other aspect, BI lifecycle provides an ability to input project requirements, logical entities, relationships, business rules, source attributes, target attributes and business metadata. These act as inputs to the logical model which can be defined and then reviewed by the business. Once approved, a physical model is generated from the logical model, the database objects are generated automatically from the physical model and the data integration mappings are created along with the objects in the persistence layer.

Universiti Utara Malaysia

2.3.4 Business Intelligence Process Stages

According to CSIRO (2003), BI ideally, consists of five main stages, which named: data sourcing, data analysis, situation awareness, risk assessment, and decision support as visualize in Figure 2.12. As mentioned in Chapter 1 in the context of this research, the focusing is directed at stage 1 (data sourcing). Consequently, the next paragraphs, stage 1 were described in detail while the remaining stages were discussed in a nutshell.

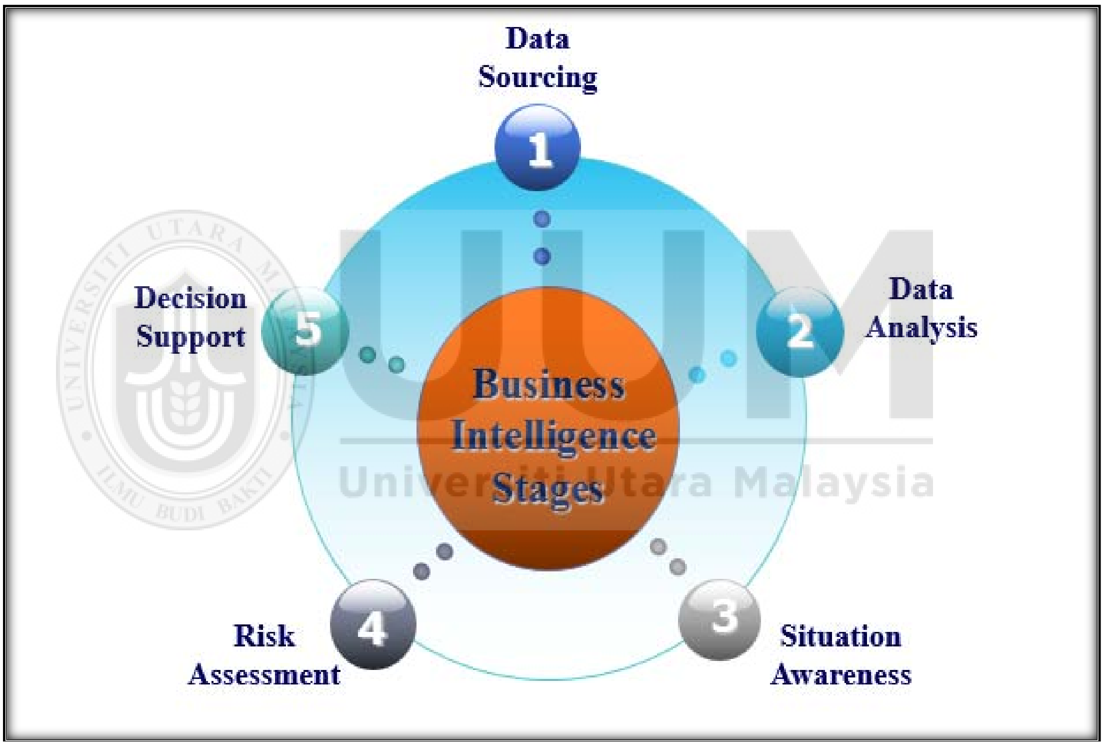


Figure 2.12. Business Intelligence Stages
Source: (CSIRO, 2003).

2.3.4.1 Data Sourcing

BI systems can extract data from various data sources. These data may include data related to finance, human resources, marketing and other data which were stored in homogeneous or heterogeneous sources and are transformed, cleansed, cleaned and integrated before they are delivered to the second stage (data analysis).

2.3.4.2 Data Analysis

At this stage, the BI systems use a different data analysis technology in order to convert data into meaningful information and knowledge. The output of this stage will help executives and managers to get a full understanding of the environment and this in turn positively affects decision making.

2.3.4.3 Situation Awareness

Situation awareness is a stage where BI systems should have full in-depth understanding and comprehension of the current decision situation. This is based on the outputs provided by the data analysis stage. On the other hand, BI systems should be able to provide assistance to decision-makers to develop rich situation awareness about their decisions situation.

2.3.4.4 Risk Assessment

Risk assessment is the stage where BI stakeholders make predictions about business future based on the output from situation awareness stage. This stage is essential as it allows businesses to identify threats, risks, seize opportunities and response accordingly. Nowadays, business complexity has increased significantly in the operation of a corporate environment. Thus, performing the risk assessment is an essential part of BI environments.

2.3.4.5 Decision Support

The aim of building a BI system is to help stakeholders (managers and executives) to make wise and timely decisions based on data derived from the data sources. Therefore, the aim of this research is to design a BI model using the data virtualization technology. Specifically, this research focuses on the first stage of BI system, which is data sourcing stage. As mentioned in the previous sections, the information and knowledge of the business are extremely valuable, and should be stored and handled in an acceptable manner. There is an urgent need for businesses to possess the capabilities for the management and treatment of information. Therefore, it is imperative that organizations adopt intelligent solutions to deal with this data (storage, processing and use).

Since, the mean aim of this research is to purpose DV design model for near real-time decision making in BI environment; which means providing near real-time data in supporting decision making process in BI environment, therefore, in the next paragraphs, the details of the data sourcing as the first stage in BI environment was highlighted; in conjunction with the first three phases of BI lifecycle which have been discussed previously.

Universiti Utara Malaysia

As mentioned early, the information and knowledge of the businesses is extremely valuable, and should be stored and handled in an acceptable manner, there is an urgent need the businesses to capabilities for the management and treatment of information. Accordingly, it has emerged the urgent needs of thinking about the intelligent solutions for how to deal with this data (storage, processing, and use).

In order to create business knowledge, information need to be dealt with at the right time, place and manner (Folkes & Quintas, 2004). Hence, the concept of databases is considered as a significant information resource that is used to create business knowledge. Accordingly, organizations tend to record information created from business transactions directly in databases, and timely access to these data. These data will then be transformed into knowledge and extracted to be used in making better decisions. Organizations and businesses also require the capabilities for managing information in order to have quick access and extract knowledge. All these have led to the design and adoption of information technology and information systems in businesses in the 1970's. Over the years, new issues have emerged to address issues such as reducing access time and improving user friendliness in future IT applications. Consequently, all information systems designed were aimed at creating an environment characterized by being easy to use and functional to meet user requirements (Jain, 2003).

After the introduction of database concepts in the 1970's, IT vendors such as SAP and IBM have shifted their focus to the provision of innovative management; this includes improving access, maintenance and retrieval of data from different databases. At that time, IS Businesses used online transaction processing application (OLTP) to enter, update, process and access data from databases. In the same aspect, the main aim of OLTP applications is considered as major facilitator uses business data stored in databases, which is one of the early facilitators in the business intelligence systems.

Besides, limitations in OLTP, are it doesn't provide a deep use of information should be linked to data from different applications, Therefore, access to information from various applications are of vital importance to supporting decision-making process as well as performance monitoring (Lau, 2005). Accordingly, the idea that standing behind of storing data outside of the operating system to enable the integration of data from several and different applications.

As a result, the need for a solution used to access and manipulates data from different applications in supporting decision-making process by providing real time or near real time data arises. Hence, the next section describes an overall overview of real-time and near real-time data.

2.3.5 Real Time and Near real Time Decision Making

In 1998, Gartner Research introduced the term Zero-Latency-Enterprise (ZLE) (GROUP, 1998). The idea behind ZLE is to try to reduce the time between a business event and the suitable action as shown in figure 2.13. ZLE generally aims to enhance the business process performance as shown in figure 2.6 that is adopted from (Nguyen & Tjoa, 2006).

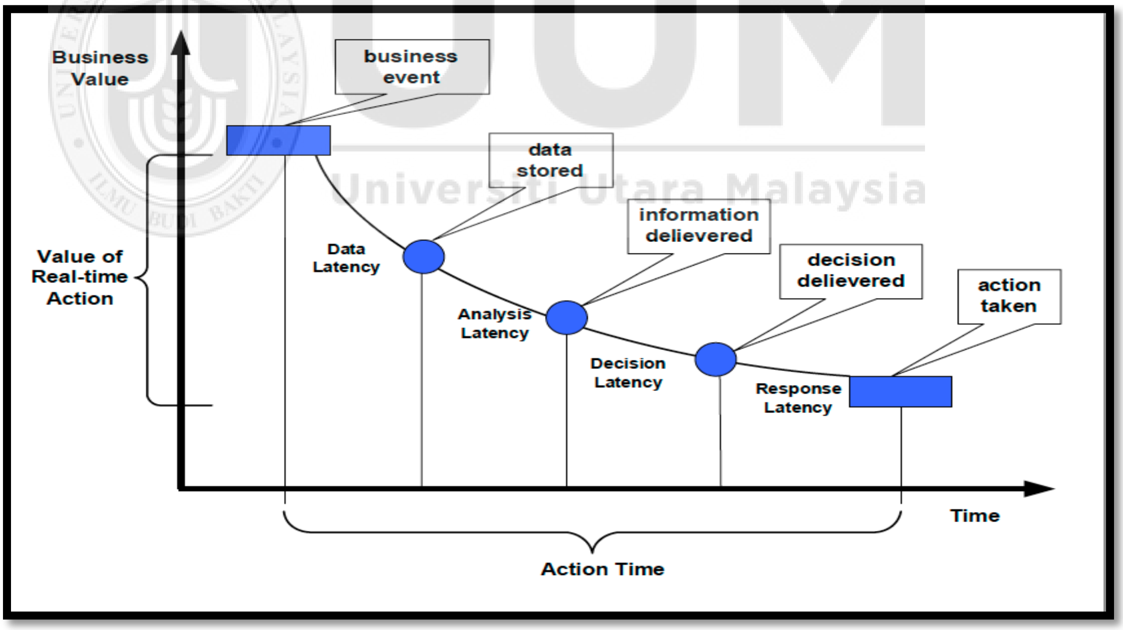


Figure 2.13. The ZLE Idea
Source: (Nguyen & Tjoa, 2006).

In BI environment, the term 'real-time' is frequently used, but what real-time means in a specific context is rarely defined. BI is usually defined based on the user's perspective and situation the user is in. For instance, a Chief Executive Officer may interpret it differently than a production worker. In the other aspect, IT/IS applications are usually adjusted to a "human schedule" for example "backup on weekends". Sometimes the term "right time" is used as an alternative to 'real-time' to emphasize an implicit relevant time scale for a problem domain.

According to Azvine et al. (2006), there is no specific accurate definition or understanding of the term of "real-time" and he introduced three different meanings to it. "Zero Latency" Processes refer to up-to-date information whenever needed by user, and key performance indicators (KPI) relate to current situation (i.e. now) (Ben Azvine et al., 2006). However, there are two justifications been given by authors for the significance of real-time in BI environment; i) the business environment, and ii) the advances in technology. The environment businesses operate in changes rapidly (e.g. share prices, sales pattern etc.), thus, there is an increasing need for live (real-time or at least near real-time) information that can be used to create significant and meaningful reports. Today's technology would look forward to the design of RT-BI systems. For example, the Internet is mentioned as a means of distributing data throughout an organization. However, according to literature, the current BI systems face two challenges in regard to providing RT-BI i.e. the conversion from data to information and from information into action or knowledge. Converting data to information is challenging because highly skilled professionals are required (e.g. expensive, limited availability). The transition from information into action is currently "manual" and BI systems provide data and reports, but these outputs are not automatically applied to the respective processes. Occasionally, the data processing systems are to slightly slower than real-time, which is called near real time. Consequently, the next paragraphs discuss the near real time.

Universiti Utara Malaysia

Today, business time is increasingly moving toward real time. As enterprises look to grow their competitive advantage, they are trying to uncover opportunities to capture and respond to business events faster and more rigorously than ever. The duration between the event and its consequent action needs to be minimized. Therefore, one of the emerging trends for data integration is the increasing demand for “real-time” or “near real-time” data integration, i.e., the refreshment of data happens very quickly after a triggering business event (Behnam Azvine, Cui, & Nauck, 2005; Bruckner, List, & Schiefer, 2002). It is important to mention that the near real-time has no constant definition, but it varies depending on the requirements of the organizations, besides, near real time totally dependent on the requirements of the organizations and the nature of their work and also it depends on to what extent the organizations interpreting the data in real time.

In the context of this research, the term near real time mean; the data processing systems are slightly slower than real-time based on the interpretation of the organization to the real-time data as well as based on the organization's requirements. As mentioned early, one of core phases of BI systems is data processing such as DW, DV. Consequently, the next paragraphs discussed all types of data processing that used in BI environment.

2.3.6 Data Warehouse

Most organizations consider data analysis as the source and generator of knowledge that eventually support the decision-making process. Knowledge provides the answers needed to make the right strategic decisions (Chu, 2004). Databases are useful within the operational systems which deal with data that are not complicated. However, they become inappropriate in the case of complex data in organization-wide systems (Ramachandran et al., 2010). Therefore, DW were developed in order to provide the necessary ability for businesses in all sectors to effectively store and analyze data; according to Baker (1993) state that: “Analysis is the number one reason organizations and executives give for wanting a data warehouse”.

Therefore, DW could be described as central information repository that is used by BI applications to access and process information in different ways than databases. The first and fundamental discussion about data warehouse (DW) appeared in 1988 by (Devlin & Murphy, 1988). They defined DW as the kind of database characterized by being read-only database. These data are configured through the collection and integration of data from various sources, to fit different types of operation and provide tools to help user to interact with and use these data. On the other hand, the definition that has been proposed by Inmon has received the largest amount of respect and attention by authors over the years. Inmon defined DW as “a subject oriented, non-volatile, integrated and time variant collection of data in favor of decision making” (Inmon & Hackathorn, 1994). In the same aspect, there are many main business measures, usually extracted, transformed and loaded from various data sources. These data are then integrated in the DW and make it ready to use by online analytical processes or any other BI tools. Based on the literature reviews and the DW definitions, it can be concluded that there are the differences between databases and DW. Table 2.3 lists these differences.

Table 2.3
Comparison between Database and Data Warehouse

	Data Warehouse	**Databases**
1.	In terms of data integration: DW includes data from different sources in different formats, and then it defines and	In terms of data integration: Database includes data from a single source which is the application where the database is

Universiti Utara Malaysia

	standardizes all the aggregated data and stores it in a separate environment.	installed.
2.	In terms of Subject-orientation, the integrated historical data is arranged and organized by facts and dimensions; each dimension represents a subject area.	In terms of Subject-orientation: the data is arranged rationally within a single subject (purpose of the application).
3.	In terms of time-variant: Represents the data flow over time. Data is added periodically.	In terms of time-variant: Represents data about current transactions. Data is added continuously.
4.	In terms of non-volatile: Data is read only and it can't be modified or removed.	In terms of none-volatile: Data may be modified or removed.
5.	In terms of purpose of data is to help planning, problem solving, prediction and decision support.	In terms of purpose of data is to control and run fundamental business tasks.
6.	In terms of size, may reach to Gigabyte-Terabyte.	In terms of size, not exceed Gigabyte.

Normally, BI systems are built around a central DW. The DW is the center for analytics and decision support activities and is populated by a process that made up of three main tasks: Extract, Transform, and Load (Castellanos, Simitsis, Wilkinson, & Dayal, 2009; Franconi & Sattler, 1999; Inmon, 1996, 2005; Mousa et al., 2014b).

As seen in Figure 2.14, the traditional architecture of DW has been visualized. The relational database is considered as input sources to the DW; and therefore, the data stored in these sources should be improved and cleaned to be ready for used in supporting decision-making process in BI environment. The data stored in these sources have not been improved and is clean to be ready for use by business intelligence tools, and decision-support systems (Kimball, 1998; Kimball et al., 2008). Accordingly, data from these sources need to extracted, transformed into a form that is compatible with DW after discrepancies are cleaned and loaded in DW (Kimball & Caserta, 2004; Mousa et al., 2014b).

Universiti Utara Malaysia

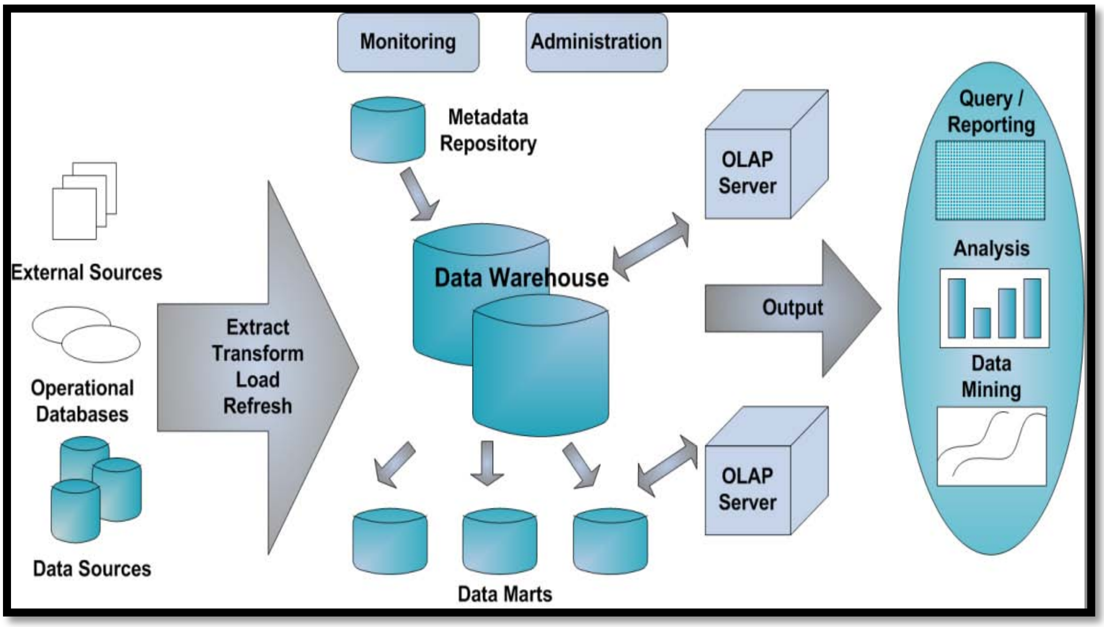


Figure 2.14. The Data Warehouse Architecture
Source: (Kimball & Caserta, 2004).

Generally, to create DW, the sources data should be passing through three core stages named extract the data, transforming from multiple sources, and eventually be downloaded to the DW. This process however, is known extract, transform, and load (ETL). A review of previous studies conducted revealed some advantages and disadvantages of DW. These findings are tabulated in Table 2.4.

Table 2.4
The Advantages and Disadvantages of DW

	Advantage	**Disadvantage**
1	Ability to access huge amounts of data that can be used to solve many obstacles faced by companies and institutions such as the trend to increase profits or to increase sales and others.	Sometimes developers do not consider the time required to process and this leads to the depletion of a large proportion of the time allotted for the construction of the system.
2	Consolidated information from various origins and merged in one location.	Generally, DW design model is complex and not flexible, causing difficulty for DW users to deal with a large amount of the organization's information, especially with changing business requirements and business needs.
3	Increased productivity and cost-effective decision-making	DW designers are forced to re-design ETL when they intend to add new sources, thus resulting in changes or alterations in rules or sources of data.

Table 2.4 Continued

4	Typically, DW store data in the facts and dimensions tables this makes it easy to understand as well as ease of study and analysis by users with limited skills.	The DW updating is normally done offline, which means that while updating the DW, all BI applications cannot access any data and deliver in real time. On the other hand, the difficulty of maintenance, the maintenance costs are high and sustained over time.
5	DW is considered as a backup for the original data, which increases the reliability and confidence of problem of data losing.	Due to the nature of the data warehouse is being considered as an extra copy of the sources data, therefore, there will be an exhaustion of the storage sources. In general, physical data integration of data from multi location data sources involves higher storage and licensing costs, longer response times, increase maintenance efforts, and a greater need for staff resources.

In line with the above situations, and a review of previous studies conducted revealed that ETL process runs at pre-defined times in addition to the execution time of ETL is too long. From this, it can be concluded that using DW in decision-making might cause lack the real-time data, which negatively affects o decision-making process. High granularity data about local demand characteristics may be lost due to some form of aggregation during the ETL process (D’Souza & White, 2006; Kemper & Baars, 2009; Negash & Gray, 2008). This leads us to the conclusion that the traditional business intelligence systems strongly do not support decision-making in real-time and business needs. This is where data virtualization is emerged, as an alternative technique of transforming available data into the form needed for reporting and analytics.

2.3.7 Data Virtualization

As mentioned early, most organizations would immensely appreciate fast decision making. Fast decision making however, is hard to implement in current BI systems because it requires dramatic redesign, this is because most business intelligence systems that were developed over the last 20 years were based on a chain of databases and transformation process, (Burstein & Holsapple, 2008; Reynolds et al., 2008). As visualize in Figure 2.15, data is transformed and copied from one database into another until it reaches an endpoint where the database is being accessed by a reporting or analytical tool (Davis & Eve, 2011; Mousa & Shiratuddin, 2015; Van der Lans, 2012). Each transformation involves a series of processes such as extraction, cleansing, integration, and transformation of the data. The transformed data is then loaded into the next database in the chain. This process continues until the data reaches an acceptable quality and form, suitable for the reporting and analytical tools. These transformation processes are normally referred to as ETL. This chain of databases and transformation processes is often long, complex, and highly interconnected. Each change made to a report or to data has its chain reaction and can lead to a myriad of changes throughout the chain. The change, though simple, might take up to days, or even weeks, before it can be implemented throughout the chain. The effect is that the BI department is unable to keep up with the speed of change required by the business. This leads to an application backlog and has a negative impact on the decision-making speed and quality of the organization. In addition, because so many transformation processes are needed and each of them takes time, it is hard to deliver operational data at an endpoint of the chain, such as a data mart.

Universiti Utara Malaysia

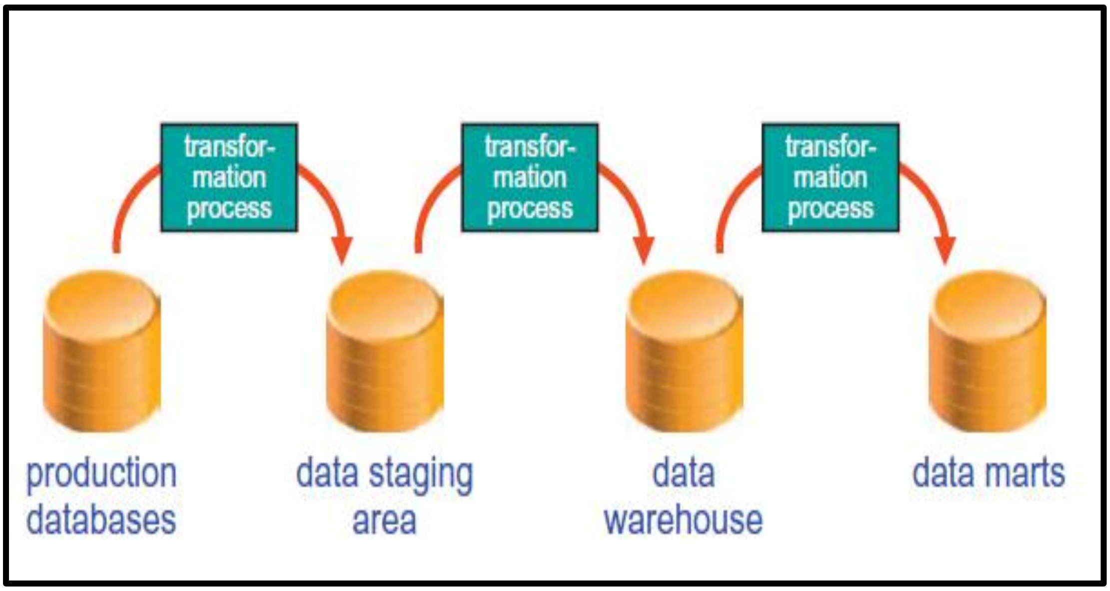


Figure 2.15. BI Systems Based on and Transformation Processes
Source: (Van der Lans, 2012).

Consequently, there is an urgent need for an agile architecture that is modular. The best way to do that is to design an architecture that consists of fewer components than other techniques. Furthermore, an architecture with fewer components means there are fewer databases being deployed and fewer transformation processes are involved. Hence, this would lead to a leaner architecture and requires only minimal changes. In addition, fewer components simplify the architecture, which also increase the agility level.

DV has been defined by several authors. In this study; the definition by Lans (2012) is adapted. Generally, data virtualization refers to a set of data stores which offer users to query access and manipulate data in a unified, abstracted and encapsulated manner, regardless of the data location. Data virtualization hides the fact that the data is being integrated to form that unified view (Ben Azvine et al., 2006; Bucher et al., 2009; Marjanovic, 2007; Van der Lans, 2012). Figure 2.16 shows DV Architecture.

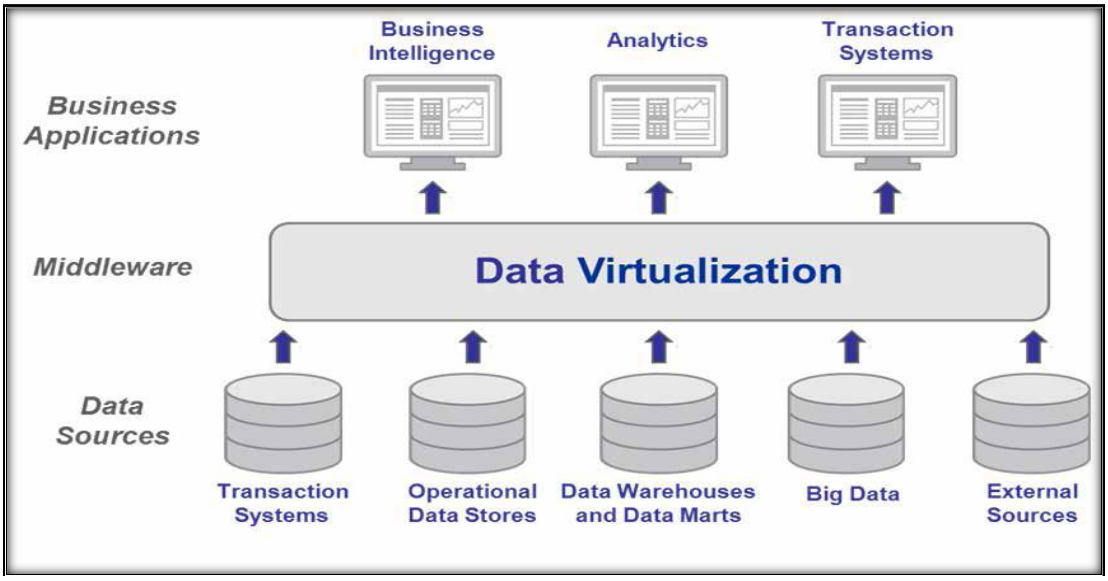


Figure 2.16. Data Virtualization Architecture
Source: (Van der Lans, 2012).

As clearly shown in figure2.16, with regards to the design issues, the DV architecture will be applicable guidelines with the intention to adapt it in the DVDeM proposed model.

Since this study focuses on using DV technique on providing near real-time data in supporting the decision-making process; and due to being DV technique plays an important role to ensure that the design model mapped with all features that are necessary to cater for of needs decision-making process. Therefore, it's important known the advantages and disadvantages of DV technique. Accordingly, the next section has discussed the advantages and disadvantages of DV technique.

2.3.8 The Advantages and Disadvantages of Data Virtualization

Like any technique, DV has its advantages and disadvantages. Consequently, many studies have shown that some advantages and disadvantages of DV (Ahuja, Kumar, & Singh, 2012; Davis & Eve, 2011; Richter et al., 2012; Van der Lans, 2012). Accordingly, it can be elicited from the literature were conducted, the advantage and disadvantage of DV. Table 2.5 lists the advantages and disadvantages of DV.

Table 2.5
Advantages and Disadvantages for DV

	DV Advantages	DV Disadvantages
1.	Provide real-time and near real time data for stakeholders.	During extracting data from multiple heterogeneous sources may facing difficulty in term of validity and accuracy.
2.	Launches data integration process after the process of extracting and merging data.	Requires the construction of an additional layer of software between the data store and stakeholders.
3.	Using DV technique cost will be less.	Results may not be used more than once to the fact that DV processed data on demand.
4.	DV be less complexity than DW regarding to data integration.	May lead to performance degradation.
5.	The time to implement DV is less.	All data sources must be online during extracting data to DV.
6.	Support data quality.	Not support recording and keeping historical data.
7.	Be appropriate in dealing with huge amounts of data.	CPU consuming time.

As described in Table 2.5, the importance of data virtualization comes from its ability to help organizations deliver complete, high-quality and actionable information more quickly and with fewer resources than traditional data integration approaches. This faster time-to-solution advantage translates into the faster realization of the business benefits such as increased revenue, improved customer service, and retention, enhanced competitive responsiveness and better regulatory compliance that are the business drivers behind new and real-time information requests.

2.3.9 Comparison between DW and DV

In this paragraph, the comparison between DW and DV was conducted and tabulated in Table 2.6. The three key purposes of this comparison are; i) to assist researchers in this field to know the pros and disadvantages of each technique, ii) to help developers to take the appropriate decision regarding any of the techniques that suitable to use, and iii) under the current state of this study, both of DV and DW are input for BI environment. Accordingly, it’s important to understand the differences between them in line with the motivation that stand behind this research is to design a new or enhanced version of the “traditional” BI concept.

Table 2.6
Comparison between DW and DV

	Characteristics	Data Warehouse	Data Virtualization
1	Outcome.	✓ Physical data model. ✓ Dimension model (facts and dimension tables. ✓ Difficult to deliver real time data in supporting decision making process.	✓ Virtual data model. ✓ Virtual tables and virtual data marts. ✓ Deliver real time and near real time data in supporting decision making process.
2	Data integration process.	✓ No connection with data sources. ✓ Three main functions to integrate data are (extract, transform and load). Accordingly, the result is a new physical copy of data sources. ✓ No isolates BI tools from the details for data sources. ✓ Sources data not mandatory to be always online.	✓ There is a connection with data sources. ✓ Three main functions to integrate data are (imported data sources, create wrappers, and define mapping). Accordingly, the result is a new virtual copy of data sources. ✓ Isolates BI tools from the details for data sources. ✓ Sources data are mandatory to be always online.
3	Data refreshed	✓ Lack of timely data updating, daily, weekly, or monthly updating.	✓ Timely data updating, Data on fly and data on demand.
4	Complexity	✓ The process of design is complex. ✓ Re-design in the case of add new sources of data.	✓ The process of design is less complex than DW. ✓ No need for the redesign in the case of adding new sources of data.
5	Design model	✓ The conventional DW system focuses on design a centralized repository of aggregated or summarized historical data. ✓ Lack of flexibility design model.	✓ DV designed as a direct connection to multiple disparate sources systems and deliver a virtual environment for accessing integrated information. ✓ More Flexible design model than DW.

Universiti Utara Malaysia

As clearly shown in Table 2.6, the comparison between DW and DV based on their characteristics was conducted. In summary, the feature of DV technique lies in its ability to provide decision makers with a single high-level view of data as well as real-time or near real time data that may be spread throughout the enterprise. This capability can dramatically simplify access to data for BI end users and supporting decision making process.

Since this study is intended to propose DV design model in supporting near real time decision making in BI environment, accordingly, the next section has discussed many existing models that relate to the development of the proposed model. Accordingly, in section 2.4, twelve existing models and approaches are discussed.

2.4 Related Work

According to Buchanan, (1990), design can be defined as "the conception and planning of the artificial". In general, the reality is excessively difficult to be duplicated, so a model should be simple and only capture the abstraction of reality. On the other hand, the complexity is unsuitable in problem solving (Turban, Forret, & Hendrickson, 1998).

Since this study is intended to propose DV design model for near real time decision making in BI environment, and to further understand design model, this section discusses twelve (12) existing BI models, and approaches that adapt traditional data integration (DW) as data integration technique, as well as 10 real-worlds DV solution that uses DV as virtual as data integration technique. Additionally, two comparative analyses of existing studies have been carried out to confirm the research gap as well as identifying common phases and components of the models/ approaches in BI environment. The DVDeM proposed model was developed based on the outcome gathered from them. Accordingly, Section 2.4.1 has highlighted the physical data integration models / approaches; while section 2.4.2 highlighted the virtual data integration Models/approaches.

2.4.1 The Physical Data Integration Models and Approaches

In proposing DV design model for near real-time decision making in BI environment, analysis on existing BI models and approaches are important. It should identify common components, and elements, as well as theories and approach that are adapted in the study.

In response to that, this study analyzes twelve (12) existing studies on design BI application uses physical data integration (DW).

i. The Generic CRISP-DM Reference Model (Wirth & Hipp, 2000).
ii. The DWARF Technique (Paim & De Castro, 2003).
iii. The CoDMODS Model (Shahbani & Shiratuddin, 2011).
iv. The IIHS Model (Khraibet, Mousa, Bakar, & Shahbani, 2013).
v. Academic Business Intelligence Model (Ta’a, Bakar, & Saleh, 2006).
vi. A Data Warehouse Model for University HRM (Zhang, 2009).
vii. Business Intelligence Model for Unstructured Data Management (BIMUDA) (Abdullah & Ahmad, 2015).
viii. Knowledge-Based Model for Real-Time Business Intelligence (AlSuwaidan & Zemirli, 2015).
ix. A Service-Oriented Architecture for Business Intelligence (Wu, Barash, & Bartolini, 2007).
x. A Conceptual Framework for Business Intelligence as a Service (SaaS BI) (Liyang, Zhiwei, Zhangjun, & Li, 2011).
xi. A Knowledge Management Approach for Rear-Time Business Intelligence (You, 2010).
xii. A Five-Layered Business Intelligence Architecture

Finally, the result of the comparison will be used to construct the DVDeM proposed model. In accordance, to the analysis of the twelve BI studies that have been carried out to identify their common components and phases on designing BI model.

In the context of this study, the approaches and models that are reviewed above relate to the proposed DVDeM model as basis of components in BI environment. Based on the above discussion, it is concluded that most of the models and approaches share similar phases in proposing the design model for BI, which are divided into some phases, sub-phases, and its components. Some of the discussed models focus on data integration while the others focus on the requirements gathering phase. In the same aspect, some of them include detailed phases; but the main steps are still similar. More importantly, data integration in real time process is required to ensure the BI applications meet the user requirements. Therefore, in proposing DVDeM, this study extracts the applicable phases and processes of the discussed BI models and approaches. Based on the analysis, the comparisons and limitations of each of the model are tabled and exhibited in Table 2.7.

Table 2.7
The Comparative Analysis for current BI Models

	Studies/ Author	Description on Comparisons	Limitations
1	The Generic CRISP-DM (Wirth & Hipp, 2000).	The study highlights a comprehensive model for carrying out data mining process. Furthermore, this model starts from understanding business problem and understanding data and ends with delivery of these data for data mining engine.	The provided elements in this model do not cater out for the universal approach for data integration. On the other hand, it does not detail to which type of data integration (physical or virtual). It lacks clarification in terms of data integration theories that are applicable to the study as well as the target users of this study is limited for data mining users.

Table 2.7 Continued

2	The DWARF Technique. (Paim & De Castro, 2003).	This approach highlights a series of stages. Each of these stages follows levels of abstraction from the application in depth, where they are grouped as requirements for projects and forms the basis requirements. On the other hand, this study caters out the gathering requirements phase only in BI environments.	The data management is not discussed and the requirements in this technique focus on physical data integration only.
3	The CoDMODS Model (Shahbani & Shiratuddin, 2011).	This study has been proposed to develop BI application by focusing on operational data. In this model, a two-phase requirements process, at the organizational and subject area levels. Additionally, two phases of requirements process have been proposed in this model, between these two phases there is requirement elicitation process that includes an approach of the requirements gathering in addition of community collaboration method.	This study does not address how to design and implement data integration. Instead, this study has focused on how to gather requirements and use it in developing business intelligence by using physical data integration. In nutshell, Clarifications about BI model design were not specifically described. This means that this model was not reflecting universal BI users. Technically, the focusing of this study is limited to the operational data store (ODS).
4	The IIHS Model (Khraibet et al., 2013).	This model consists of four key functions. The first one is data sources followed by data integration using ETL tool and then data stored using DW technique, and finally, analytical reports are produced. These analytical reports take into consideration the final results for this model.	This study has focused on designing business intelligence model without paying any attention to the requirements phase. Instead, this study focused on technical aspects only in designing and developing business intelligence environment.
5	Academic BI Model (Ta'a et al., 2006).	The aim of this BI model for academic sector in public university. Besides, this model focuses on collecting and analyzing user requirements and uses these requirements to build a data warehouse.	The model is not comprehensive and is used for the specific organization. On the other hand, it focuses on the physical side of data integration only.

Table 2.7 Continued

6	The Data Warehouse Model for University HRM (Zhang, 2009).	This model begins with a full understanding of the data source, because it is the basis for building a data warehouse. Next, the process of understanding the data system will be ready to analyze this data based on user requirements. It remains to say that the data warehouse in this model includes complex business functions.	The main focusses for this study is on physical data integration only and not a comprehensive model. On the other hand, the requirements phase in this model has been neglected.
7	Knowledge-Based Model for Real-Time Business Intelligence (AlSuwaidan & Zemirli, 2015).	This model took into consideration that all latencies (data, analysis, and action) are maintained in zero latency, which is the constraint of RTBI. While developing this model and surveying the literature on integration models and approaches for knowledge management and RTBI, on the other hand this model noted the limited attempts to improve the models in the business domain.	This model is tries to deliver the data in real time but uses the traditional methods of data integration (ETL). However, the ETL is normally a complex process and it operates at pre-defined times and requires time to process and transfer data.
8	Deploy Data Warehouse in BI Environment (Ghosh, Haider, & Sen, 2015).	This approach has been focused on presenting an integrated architecture to manage and design BI environment by coordinating several associated entities to achieve business agility. This architecture is deployed in distributed environment, and this architecture could be deployed in different types of business applications.	On this model cannot be in the delivery of real-time data because it uses DW technology because data warehouse have been used to extract, transfer, and load data.
9	A Service-oriented Architecture for Business Intelligence (Wu et al., 2007).	In this study, the architecture of SOA in BI has introduced. In order to describe a service-oriented architecture (SOA) for BI that makes possible a seamless integration of technologies into a coherent BI environment, thus enabling simplified data delivery and low-latency analytics.	Provide decision-makers and BI stakeholders with data in a timely manner will be difficult for the fact that these approaches because these approaches focus on the practical side and it does not resolve the problems of delay in the data sources.

Universiti Utara Malaysia

Table 2.7 Continued

10	A Conceptual Framework for Business Intelligence as a Service (SaaS BI) (Liyang et al., 2011) .	This framework consists of a unified five layers for Business Intelligence as a Service (SaaS BI), with each pluggable component delivered as a service. On the other hand, business users can tailor on-demand SaaS BI deployment based on their business processes and application requirements. The SaaS BI conceptual framework generally includes Infrastructure Layer, Data Service layer, Business Service layer, User Interface Service Layer, and Operational Service Layer.	Despite this framework attempted to provide the data upon request, but he focused on the conceptual side did not only focus on the practical side and design.
11	Conceptual Framework to Organize Large Volume of Data for Business Intelligence (Anusha & Krishnan, 2012).	The proposed framework use for organizing the enormous volume of data having business information using data mining techniques to retrieve information and knowledge useful in supporting complex decision-making processes. On the other hand, the crucial concept on this conceptual framework is that the business analyst preserves the capability of thinking and generating hypothesis that can be automatically tested by posing queries to the system. These queries are supported by a variety of data mining techniques that allow the system to obtain the pieces of evidence required for the query.	This framework focuses on the conceptual class without go down to the design and presentation layers. On the other hand, it failed to address the quality of the data and the type of data integration technology used.
12	BI Analytics Method without Conventional DW (Haque & Demerchant, 2010).	This method has been proposed to minimize both the ETL and DW components of the solution and allows for agile development and incremental adoption, this is achieved by taking advantage of the organization’s current legacy reporting structure as the basis and then building a BI reporting layer on top for a high-level view of the information.	This method cannot cater BI requirements and Business need because it doesn't provide any real-time data and dependence on previous reports data prepared in advance.

Universiti Utara Malaysia

2.4.2 The Virtual Data Integration Real-Word Case Studies

In the context of this study, in proposing a DVDeM model, analysis of existing real-world case studies is important. It should identify common components, and elements, as well as theories and approach that are adapted in the study. In response to that, this study analyzes ten existing real world case studies on design and implementation DV in BI environment. The main reason stands behind reviewing of these real-world case studies is to demonstrate the significant value and tangible benefits, to DV technique. The proof of success of any technique or technology lies in its adoption, and these case studies will provide all detailed information about how a wide range of different companies in several industries are achieving significant business benefits from the use of data virtualization. The key purpose of this review of these real-world DV solutions is to understand the challenges facing the data virtualization, formulating appropriate solutions to these challenges and, identify the benefits to be derived using data virtualization from the user's perspective. Besides, based on the premise that learning from the experiences of peers is often the most productive way to gain the wisdom and confidence required for success. Finally, the details of the real-world DV solutions are provide developers a tremendous amount of information about how to design, implement and manage data virtualization successfully.

It can be concluded that the main goals were (i) to understand the challenges of DV technique, (ii) how solutions were developed, (iii) what’s the benefits achieved from the user’s perspective and the best practices advice these experienced implementers could offer to other organizations with similar business problems. Table 2.8 tabulates an overview of these DV case studies adopted from (Davis & Eve, 2011).

Table 2.8
Data Virtualization Case Studies Overview

	Organization Name	Industry Type	Domain	DV development
1.	Comcast	Communications	Directory Services	Ownership change processing.
2.	Compassion International	Not-for-profit	Enterprise wide	Ministry Information Library
3.	Fortune 50 Computer Manufacturer	Technology	Procurement	Integrated procurement reporting system
4.	Fortune 50 Financial Services Firm.	Financial Services	Wholesale Bank	Support for mergers and acquisitions, new business opportunities.
5.	Global 50 Energy Company	Energy	Upstream Operations	Virtual data warehouse to support BI reporting and analytics
6.	Global 100 Financial Services Firm	Financial Services	Investment Bank division.	Data Vault
7.	Northern Trust	Financial Services	Corporate and Institutional Services business unit.	Investment Operations Outsourcing client reporting platform
8	NYSE Euronext	Financial Services	Enterprise wide.	Virtual data warehouse for post-trade reporting and analysis
9.	Pfizer	Health Care	Worldwide Pharmaceutical Sciences (R&D)	Project portfolio database
10	Qualcomm	Communications	Enterprise wide	Multiple applications

Source: (Davis & Eve, 2011).

As clearly indicated in Table 2.8, it can be concluded that the main purpose of the use of DV in these organizations is to enhance business agility. According to Davis & Eve, (2011) the benefits that have been achieved from the use of DV in those organizations were impressive and this another motivation to encourage other organizations to go and take advantage of the possibilities of DV. Besides, these real-world DV solutions that reviewed above are relating to proposed DVDeM model.

Universiti Utara Malaysia

In the context of this study, ten real-world DV solutions that reviewed above are relating to proposed DVDeM model. Accordingly, the study adopted most of the components and sub-components as well as the flow between phases and components which have reviewed in above. Based on the analysis, the comparisons and limitations of each of the real-world case studies are tabled and exhibited in Table 2.9.

Table 2.9
The Comparative Analysis for Real-World DV Solutions

	Case Studies	Description on Comparisons	Limitations
1	Comcast	In this real-world DV solution, the DV technique was adopted in to meet the organization needs supporting decision making process for this organization and deliver data in real time.	The data virtualization solution is not comprehensive and the requirements phase is neglected. Additionally, it does not detail the steps that used for developing DV solution to be a guideline to BI developers in designing and implementations DV solution in BI environment.
2	Compassion International	In this real-world DV solution, the data virtualization solution focus on Ministry Information Library, this organization has developed a data integration system called it the Ministry Information Library. To provides an enterprise-wide, single version of the truth in virtual manner to answer all questions and provide corporate alignment of numbers and metrics as well as definitions of core business entities.	This solution use for specific purpose. On the other hand, the requirements phase in this solution is neglected. Additionally, it does not detail the steps that used for developing DV solution to be a guideline to BI developers in designing and implementations DV solution in BI environment.
3	Fortune 50 Computer Manufacturer	In this real-world DV solution, the data virtualization to create Procurement Reporting System in order to deliver a global view of inventory data to procurement users. On the other hand, this company looking forward to building a reporting system which can use to facilitate data accessing, managing, analyzing, and integrating by the users in flexible manner.	This data virtualization solution focus on deliver reporting service, therefore, it considers as specific purpose. On the other hand, this study does not mention any about the requirement phase. Additionally, It does not detail the steps that used for develop DV solution to be a guideline to BI developers in designing and implementations DV solution in BI environment.

Table 2.9 Continued

4	Fortune 50 financial services firm	In this real-world DV solution, the organization applying data virtualization technology has become is not impossible to integrate all companies flexible and inexpensive as is the case in other technologies such as data warehouse. Additionally, this solution consists of three phases: pre-merger, merger and post-merger.	In this DV solution focus only about how to overcome the difficulties that faced during ETL and tries to reduce the cost by using data virtualization technology. On the other hand, in this case study the focusing on technical side only, without pay any attention about the requirement side.
5	Global 50 Energy Company	In this real-world DV solution, the proposed solution is a data virtualization layer implemented with the Cisco Data Virtualization Suite that deliver a virtual data warehouse and virtual data marts in order to give fully support BI systems tasks such as analysis, reporting and decision-making needs.	This real-world DV solution has done for specific purpose to enhance data integration in energy sector. The requirement side during the design and implementation such data virtualization solution have been neglected.
6	Global 100 Financial Services Firm	In this real-world DV solution, the data virtualization solution for such investment bank creates Data Vault (DV). The DV is an operation data store (ODS) to provide access, the bank data at real-time and on-demand, in order to make the stakeholders access to the bank data wherever and whenever needed. On the other word, this solution provides a single version of the truth between data consumer and data sources.	The data virtualization solution in this case study has focus on deliver real time access for data by the data consumers. There is no any guideline for other developer to use for other company. Noteworthy, the requirement part has been neglected in this solution.
7	Northern Trust	In this real-world DV solution, the data virtualization solution is design and implementation data warehouse in a virtual manner with the Cisco Data Virtualization Suite with consideration to new client reporting front end.	This data virtualization solution focusing on physical data integration. Noteworthy, the requirements side for this designing and implementation has been neglected.
8	NYSE Euronext	In this case study, The NYSE Euronext data virtualization solution is an enterprise-wide DV layer, built using the Cisco Data Virtualization Suite that functions as a create a virtual view of (DW) to deliver unified read-only to post-trade data for access in order to manage these data,	The data virtualization solution in this case study has focus on deliver real time access for data by the data consumers. There is no any guideline for other developer to use for other company. Noteworthy, the requirement side has been neglected in this solution.

Table 2.9 Continued

9	Pfizer Inc.	In this case study, the data virtualization for this company is the PharmSci Portfolio Database (PSPD), a federated data delivery framework implemented with the Cisco Data Virtualization Suite. This solution enables the integration and manages of all data sources into a single reporting schema of data to access and use all front-end tools and users.	The solution has done for specific purpose to enhance data integration in energy sector. The requirement phase during the design and implementation such data virtualization solution have been neglected.
10	Qualcomm Incorporated	In real world solution, the data virtualization solution for this company is an enterprise-wide data virtualization layer built with Cisco's Data Virtualization Suite. To overcome all the challenges faced by this company and deliver data in real time and without the loss of any other storage as is the case in ways that traditional data integration.	The data virtualization solution in this case study has focus on deliver real time access for data by the data consumers. There is no any guideline for other developer to use for other company. Noteworthy, the requirement side has been neglected in this solution.

2.4.3 Implication of Comparative Analysis to the Study

In this research, two comparatives analysis were conducted. The first analysis, described in Section 2.4.1 compares existing BI models to know the challenges and past trends of the data integration in BI environment. In Section 2.4.2, the comparative analysis is restricted to real-world DV solutions that have used the DV as a data store that was created during the life cycle of BI. This comparison was conducted considering ten DV solutions which highlight issues related to challenges faced by data storage, and data delivery at an acceptable level. Moreover, the comparative analysis also covered most reviewed models and subsequently produced guidelines to ensure it matches with their target user. However, it was also discovered that some models are highly lacking in providing live and right data to be used near real time manner for decision making in BI environment. The study adapted some of the design principles, phases, BI theories and BI approaches in the studies that were found appropriate. Nevertheless, this study also analyzes on the existing DW designed approaches as well as DV real-world solutions, in proposing the BI model to serve as providing real-time data integration using DV technique. It is important to consider the identified common components, and phases, as well as theories and approaches that are adopted in the study. Thus, in Section 2.2 have been discussed the research theories that are applicable to this study. Finally, the result of the comparison will be used to construct the proposed model. In accordance to the analysis twelve of studies related to BI existing models as well as ten real-world data virtualization solutions that have been carried out, to identify their common components, phases, BI theories, and data integration approaches. Based on the comprehensive review of related literature the research gape was analysis and identified. Hence, the details about gaps analysis were described in the next paragraph.

Universiti Utara Malaysia

2.5 Gap Analysis

In view of this, the gap in the literature is identified and fulfilled in this research by design and develop DVDeM proposed model. Therefore, the efforts of this research are focused on providing a near real-time data to support decision-making process in BI environment. Hence, the following research gaps are extracted:

i. Some restrictions and aspects that should be considered when developing BI models are not clearly specified in the current models and approaches in terms of data integration.

ii. Due to DW update in an offline manner, therefore, most BI models lack tools for data delivery in real time or near-real time.

iii. Due to the dynamic change of business needs, the traditional BI approaches do not fully support the emerging business needs of operational and local decision support.

iv. Due to the complexity and cost of BI design models, thus, it is not easy to rebuild BI especially due to the growing amount of data dramatically.

v. There is limited data integration in a virtual manner in the current available BI environment design components.

vi. There are limited guidelines in the form of a design model for developing BI application using DV technique.

In view of this, there is a need for a comprehensive model for BI data that not only serves the decision-makers process in the delivery of near real-time, but this model to be a guide for developers in this area. Hence, in to fulfill these gaps, the DVDeM model was proposed. Therefore, the efforts of this research are focused on providing a near real-time data to support decision-making process in BI environment.

2.6 Business Intelligence Modeling

BI modelling is used to describe progress of all activities and how they interact and dealing with all resources in order to achieve goals (OMG, 2009; Rumbaugh, Jacobson, & Booch, 2004). According to (Eriksson & Penker, 2000), there are six different reasons to lead us to create a model of BI, which is: (i) to understand all the underlying mechanisms of action exist; (ii) to guide the establishment of information systems necessary and appropriate to provide full support for the work; (iii) to implement the improvements needed for the development of the current models to show the architecture of the innovators business; (v) to experience and apply new concepts to BI, and (vi) to determine all the elements of the business are not considered part of the core, which can be delegated to the external supplier (Eriksson & Penker, 2000). The growing interest in BI modeling has led to the invention of a different set of BI modeling languages (BPMLs). Currently, there are a wide range of notations that are commonly used in the BI modeling (List & Korherr, 2006), such as UML 2.0 Activity Diagram (OMG, 2009), Business Modeling Notation (OMG,2006), Event Driven Process Chain (EPC) (Scheer, 1999), and others. The main reason for that is a variation to the needs and differing views for designers and modelers (Mendling, Neumann, & Nüttgens, 2005; Zur Muehlen & Rosemann, 2004).

2.7 Requirement Gathering and Analysis

In the field of requirements engineering, the requirements gathering is the process of collecting the requirements of the projects or systems from the stakeholders, users or other customers (Sommerville & Sawyer, 1997). The terminology 'elicitation' or 'gathering' have been used in requirements engineering area interchangeably.

Bourque and Fairley, (2014) reported that the gathering and analysis requirements is significant to lead the systems, projects, or software to successful completion. The process of collection requirements should be workable, measurable, testable, and reflects the needs or opportunities of users or stakeholders in terms of business and determines the level of detail sufficient for system design.

A requirement is a statement regarding what should be done or how a task should be performed for a product or for intentional software. In requirements engineering, requirements gathering is an important and necessary component of any project or project management endeavor. The complete understanding of a project output is critical to the success of the project in question, and requirements gathering can be considered a cornerstone to the success of any project. The primary reason for the failure of many data integration projects is improper planning and inadequate project management (Newell, A. F., et al. 2006). Moreover, requirements must be clear, unambiguous, and specific to meet the related needs of various projects.

Requirements can be divided into three general types: business requirements, which include the sponsor's point of view, project scope, and business goals and objectives; user requirements, including the user perspective, goals, and inputs and outputs; and system requirements. This third type is further subdivided into two types: the first one is functional requirements, which answer the question "what does the system do, The second consists of non-functional requirements that answer the question "how well does the system perform the task, Figure 2.17 illustrates the three types of requirements.

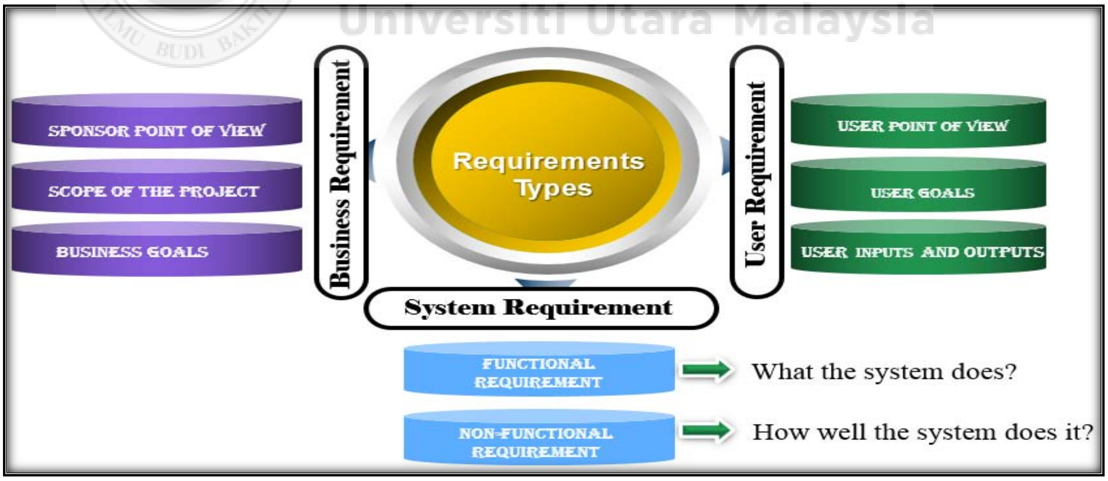


Figure 2.17. The Requirements Types
Source (Mousa, Shiratuddin, & Bakar, 2015b).

In the context of this study, the gathering and analysis the requirement is important, therefore, one of the DVDeM Model main phases is DV requirement gathering (refer to Chapter 4), which is focuses on collecting and analyze necessary requirements to develop BI applications using DV technique.

2.8 Goal-Oriented Requirement and Business Intelligence

Generally, the organizations have used BI based on strategic information applications frequently in the organizations due to the competition in the business environment and the necessity for the forecast, predictive analytics, and reporting (Schläfke, Silvi, & Möller, 2012; Viaene & Van den Bunder, 2011). This typical information system also has the capability of entailing comprehensive analysis that supports decision making or devoid of multifaceted statistical models. The system also identifies the actionable insights of organization through the processes of management; from planning, to operation and evaluation. In doing this, high cost would be reduced and the organizational resources would be well-utilized (Viaene & Van den Bunder, 2011).

For the decision-making process, BI tools utilize model dimensions models of the data, with data models that are designed with respect to the organizational goal specifications (Pourshahid, Richards, & Amyot, 2011). The sophistication of BI is determined by the correctness of the data, the appropriateness of the data warehouse schema and the precision in their OLAP technologies or BI reporting services tools (Pourshahid et al., 2011; Ramanigopal et al., 2012). Therefore, it is important to design the appropriate components for the goal-oriented data integration analysis and design, and identify sources and nature of the operational data.

This typical information system also has the capability of entailing comprehensive analysis that is supporting decision making amidst or devoid of multifaceted statistical models. The system also identifies the actionable insights of organization

Universiti Utara Malaysia

through the processes of management; from planning, operating, to evaluation. In doing this, high cost would be reduced and the organizational resources would be well-utilized (Viaene & Van den Bunder, 2011).

There are many requirement analysis techniques such as: Knowledge Acquisition in autOmated Specification (KAOS), Non-Functional Requirements, Goal-Based Requirements Analysis Method (GBRAM), and Tropos. Also are Unified Modelling Language (UML) as goal-driven approach, and Scenario-based Requirements Analysis Method (SCRAM) as scenario-based approach. Table 2.10 listed and explained these approaches.

Table 2.10
Modelling Approaches in Requirement Analysis

	Approach Name	**Author & Proposition year**	**Brief Description**
1	Knowledge Acquisition in autOmated Specification. (KAOS)	Van Lamsweerde,2009)	Formal framework based on temporal logic and AI refinement techniques where all terms such as goal and state are consistently and rigorously defined. The main emphasis of KAOS is on the formal proof that the requirements match the goals that were defined for the envisioned system.
2	Non-Functional Requirements	(Mylopoulos et al.,1999)	This approach is based on the notion of soft goals rather than (hard) goals. A soft goal is satisfied rather than achieved. Goal satisfying is based on the notion that goals are never totally achieved or not achieved.
3	Goal-Based Requirements Analysis Method (GBRAM)	(Anton, 1996).	This approach defines a top-down analysis method refining goals and attributing them to agents starting from inputs such as corporate mission statements, policy statements, interview transcripts etc.
4	An agent-oriented software development methodology (Tropos).	(Bresciani, Perini, Giorgini, Giunchiglia, & Mylopoulos, 2004).	This approach utilizes the concept of agent goal, and related notions are used to support all software development phases, from early requirement analysis to implementation. Tropos differs from other goal-oriented methodologies since it moves the notions of agent and goal to the early stages of software development.

Table 2.10 Continued

5	Scenario based Approach: (SCRAM Scenario).	(Uygun, Öztemel, & Kubat, 2009).	This approach based Requirements Analysis Method (SCRAM) concern on scenario modelling. Scenarios are the representations of the real world. From DSS point of view, this method is suitable for simulation type of research where the elicitation approach were made through a series of iterative different scenarios.
6	GRAnD	(Giorgini, Rizzi, & Garzetti, 2008).	A goal-oriented approach to requirement analysis for data warehouses based on the Tropos methodology. Two different perspectives are integrated for requirement analysis: organizational modeling, centered on stakeholders, and decisional modeling, focused on decision makers. Our approach can be employed within both a demand-driven and a mixed supply/demand-driven design framework.

Based on the above explanation on requirement analysis, it is identified that goal driven approach is more suitable for modeling purposes due to scalability of features offered by the technique that ranges from early requirement to detailed design. In the context of this study, the different direction has tended. The goal-oriented approach based on data virtualization technology has been proposed. (refer to Chapter 4).

2.9 Software Testing in Business Intelligence Environment

Usability testing is an extremely important element of software development. It focuses on how well users can understand and utilize a product to fulfill their intended goals (Clemmensen, Hertzum, Hornbæk, Shi, & Yammiyavar, 2009; Frandsen-Thorlacius, Hornbæk, Hertzum, & Clemmensen, 2009). In the world of software development, most software functions tend to be complex and to ensure that such software will satisfy users, it is extremely important for this software is to have high degree of usability (Benbunan-Fich, 2001; Nielsen, 1994). In the same aspect, the optimal use of BI applications depends on various factors including the usability of the product (Jooste, van Biljon, & Mentz, 2013).

According to Nielsen, 1994, usability, to some extent, is the question of whether the system is good enough to satisfy all the needs and requirements of the users and other potential stakeholders, such as the users, clients and managers. On the other hand, usability can have defined as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency. In the same aspects, Dix, (2009) states that the process of usability evaluation includes three main goals ; i) to measure the extent and accessibility of the system's functionality, ii) to measure users' experience of the interaction, and iii) to identify any particular issues within the system (Jooste et al., 2013). There are many attributes have been identified by (Nielsen 1993), and we list them in the next paragraph.

From the definition of usability, four factors can be determined. Those factors which are users, tasks, environments, and technologies can strongly influence the usability of application (Baharuddin, Singh, & Razali, 2013). Many studies have been conducted to develop models, frameworks, and approaches to assess the usability of application and determine it is contextual factors (Coursaris & Kim, 2011; Yen & Bakken, 2012).

2.10 Usability in Business Intelligence

Usability can be considered as one of the factors in terms of determining of the best of use and ultimate benefit obtain from BI application. As we mentioned in the beginning of this chapter, the purpose of BI is to help and support the massive warehouses and flow business data in, out, and around the organizations by identifying, processing the information into meaningful information and valuable managerial knowledge and intelligence (Hou, 2012; Lönnqvist & Pirttimäki, 2006). In 2013, Jooste studied the usability of BI in the coal mining organization and he proposed a guideline for measure the BI usability. This guideline classify the functional grouping into many attributes in supporting usability in BI environment (Jooste et al., 2013).

In order to gain from the actual users about the prototype usability and according to (Barnum & Dragga, 2001), a usability testing should be conducted (Barnum & Dragga, 2001).

In the same aspect, Jooste et al. (2013) state that the decision-making data for BI system can be measured as one of the usability attributes of this system, and based on the items that spread over such attribute. Accordingly, in the context of this study, the near real-time data for decision making are considered as one of the Q-U instrument attributes as well as the other usability attributes, and therefore, it is measured based on six items spread over such attribute. Based on the existing studies as well as Jooste guideline, an instrument (Q-U) to test the usability of the proposed prototype was developed. The process of instrument development is described in Chapter 3, Section 3.4.3.1.1.

2.11 Chapter Summary

This chapter reviewed key literature in the fields of BI, DSS, Database, DW and DV and presents the relevant knowledge base for this research. Each of these topics presents a current and highly interesting area in IS/IT research that is of particular relevance for businesses.

BI is the term that describes the current state of decision support systems. This concept emerged at the beginning of 1990 with the intention to support the work extraction and processing large amounts of data from the heterogeneous sources and turn it into meaningful information. However, any improvement or change in the business intelligence environment requires these systems to deal with more data in less time for the purpose of supporting the complex decision-making processes. Review of literature shows that researchers in the area of BI argue repeatedly that these systems currently do not support or deal with these demands appropriately.

DV technique was presented as a new data integration technique with characteristics that allow the integrate data from multiple and heterogeneous data sources for use in BI environments. This technology is used in this research as an important component in our proposed model to deliver near-real time data for decision making process in BI environments to that overcomes some of the issues that were identified in the literature. On the other hand, an overall overview of existing models and frameworks of BI have been presented in this chapter. Besides, ten real-world DV solutions have been presented and discussed, for knowing businesses and organizations that have used the data virtualization technology as well as knowing the key components that can use in data virtualization modeling.

Especially in industries field with a rising number of decisions, IT systems can be supported by analyzing large amounts of data. The huge number of decisions and the various influence factors makes the decision-making process an extremely complicated and should be able to show some of the aspects and pros of the proposed model.

In the context of this study, the research gaps have been identified and addressed, in order to provide an assistance to develop BI applications and therefore, the concept can leverage available data (and the investments made in technologies) and better support businesses and their decision makers in their work. However, the outcomes is explained to emphasize the benefits of the research and whole finding that will contribute all the objectives.

UUM
Universiti Utara Malaysia

CHAPTER THREE
RESEARCH APPROACH AND METHODOLOGY

3.1 Introduction

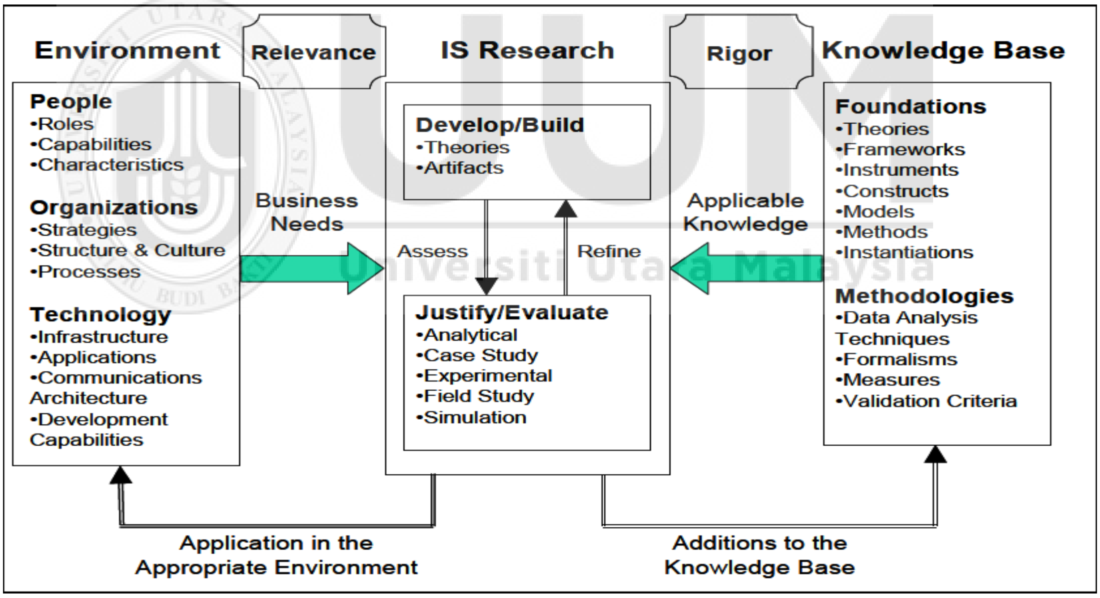
The selection of appropriate research methodology is based on the main aim of this study, which is to the develop DVDeM design model. When considering a suitable methodology to be adopted, the methodology should guide this study towards achieving the targeted aims. However, the main objective of this chapter is the full description for research approach undertaken in conducting this research study to achieve the outlined objectives. A methodological approach that consists of all the processes is adapted from design science research in information system. This approach will demonstrate the whole flow of a project by detailing each phase involved as well as the relationship between the outcomes and research objectives. A brief description on the design science approach is provided in Section 3.2.

3.2 Research Approach and Methodology Selection

Most researches in the IS discipline is characterized by two paradigms: design science and behavioral science. Design-science is a research paradigm that extends the boundaries of human and organizational capabilities by creating new and innovative artifacts. The behavioral science paradigm "seeks to develop and justify theories (i.e., principles and laws) that explain or predict organizational and human phenomena surrounding the analysis, design, implementation, management, and use of information systems" (Hevner, March, Park, & Ram, 2004).

Once the research questions that drove this study were defined, it became apparent that a design science paradigm is the suitable research approach for this study as addressing the research questions will be a contribution by improving system performance.

Figure 3.1 presents the IS research framework proposed by Hevner et al, (2004). In this framework, the environment defines the problem space in which includes people, organizations and technology. The research problem perceived by the researcher (derived from business needs) transpires from the environment. The business needs are influenced by people and are assessed and evaluated within the context of an organization (organizational strategies, structure, culture, and existing business processes). Accordingly these needs are shaped with relation to the existing technology.



Source: (Hevner et al., 2004).

Figure 3.1. Information Systems Research Framework

Having identified the business needs, research is conducted in two complementary phases - design science (develops and builds) and behavioral science (justifies and evaluate). During the design science phase, the researcher addresses the research through the building of artifacts designed to meet the identified business needs. The applicability of the designed artifacts to the business needs is then evaluated during the behavioral science phase. The knowledge base provides foundations (theories, frameworks, instruments, constructs, models, methods, and instantiations) and methodologies (data analysis techniques, formalisms, measures, and validation criteria) that have resulted from prior research studies. Foundations are used during the design phase, while methodologies are used for evaluation during the behavioral science phase. Based on this framework, the contributions of design science and behavioral science are assessed by (i) applying to the business needs in an appropriate environment, and (ii) the value added to further research and practice as the result of adding to the contents of the knowledge base (Hevner et al., 2004).

The research design for this study is influenced by the IS framework proposed by Hevner et al, (2004). As it is based on both behavioral science and design science paradigms, the research design for this study utilizes the design science paradigm as a research approach to innovate new artifacts.

Universiti Utara Malaysia

3.3 Design Science Paradigm

The design science paradigm is a problem-solving paradigm (Hevner et al., 2004) which aims to provide answers to design problems (Andriessen, 2006) . The design science methodology, in general, consists of a process (set of activities) and a product (artifact) (Hevner et al., 2004; March & Smith, 1995; Walls, Widmeyer, & El Sawy, 1992). It means the design process is a sequence of activities that produces an innovative artifact. The design science methodology by March (1995) identifies two design processes: build and evaluate; and four artifacts: constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems). Constructs are the conceptual vocabulary in which problems and solutions are defined. A model is a set of propositions that presents the relationship between constructs. A method is a set of procedures used to perform a task. Instantiations are operationalized constructs, models, and methods - an instantiation is the realization of the artifact in an environment (March & Smith, 1995).

Appropriate process in design science research has been suggested by many researchers. Peffer, Tuunanen, Rothenberger, and Chatterjee (2008) reviewed and evaluated the process for conducting design science research in IS. March and Smith (1995) and Vaishnavi and Kuechler (2009) also carry out the design research processes in their work. On the other hand, Hevner and Park (2004) have proposed seven design research guidelines to assist researchers to understand the requirements and structuring suitable process for the approach.

Universiti Utara Malaysia

3.4 Research Methodology Phases

The main idea of this study is to design a model in BI environment using data virtualization technique. However, in order to achieve this, systematic tasks should be planned for achieving the research objectives (Dwolatzky, Kennedy, & Owens, 2002). Design science research methodology involving three main phases, which are theoretical study, artefact development and empirical testing were adopted to accomplish the research objectives.is adopted for accomplishing the research objectives. These three main phases can be divided into five phases; (i) awareness of problem, (ii) suggestion, (iii) construction, (iv) evaluation, and (v) conclusion.

i. **Theoretical Study**: identify gaps, literature research and comparative analysis.

ii. **Development:** design artifacts (design model), model validation via experts' reviews and focus group discussion, and additional literature research.

iii. **Empirical Testing:** prototype implementation in case studies, prototype usability testing, and data analysis.

In the context of this study, the overall research methodology steps and its activities are illustrated in Figure 3.2.

Universiti Utara Malaysia

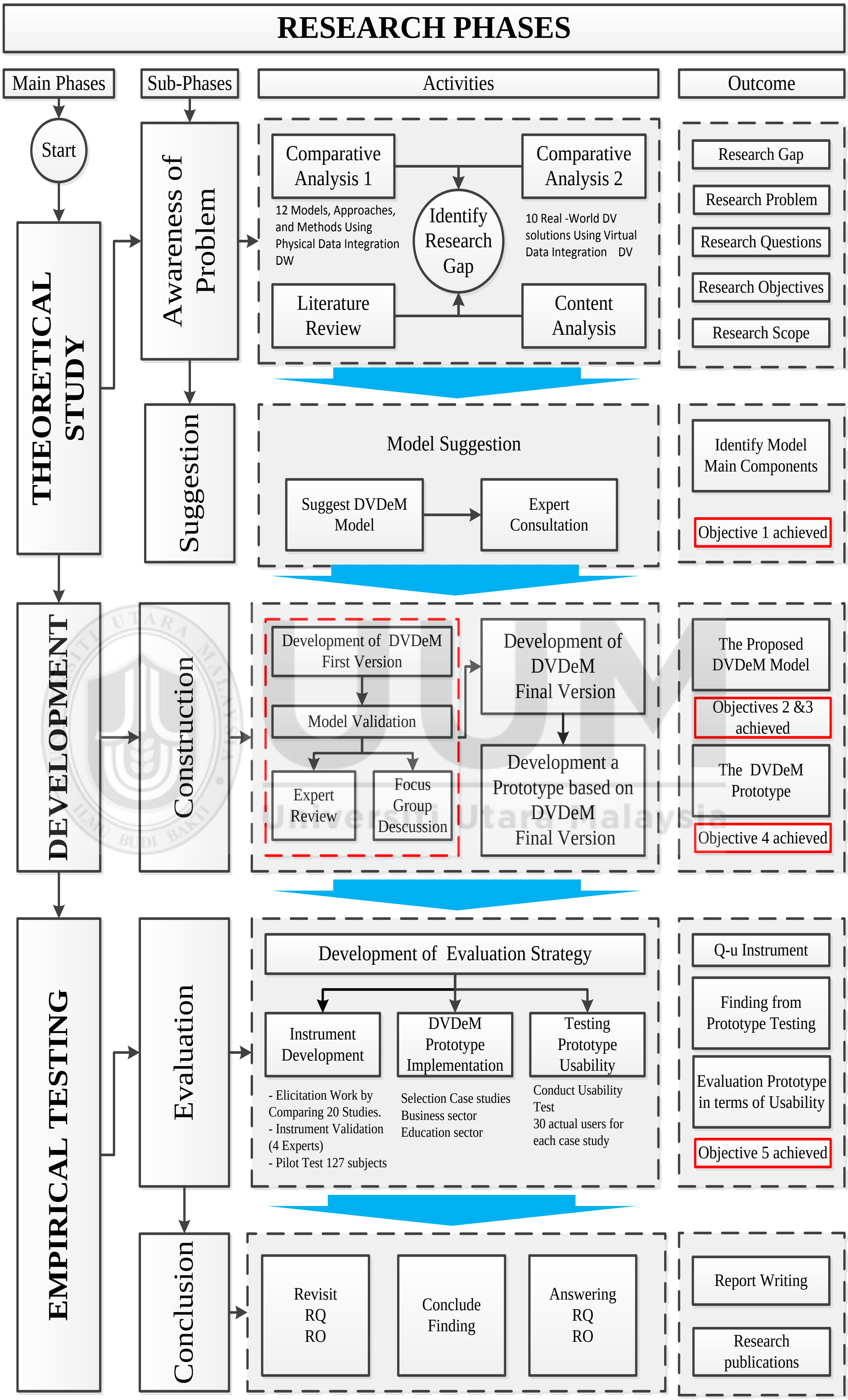


Figure 3.2. Research Process Phases

3.4.1 Theoretical Study

Theoretical study is carried out to obtain the major issues to be studied as well as to get the research gaps and scopes. The main purpose of this phase is to review prior and relevant literature in order to create a firm foundation for advancing knowledge by facilitating theory development (Gacenga, Cater-Steel, Toleman, & Tan, 2012; Offermann, Levina, Schönherr, & Bub, 2009; Webster & Watson, 2002) as well as uncovering areas where research is required. Content analysis, relevant literature, and comparative analysis are required in understanding the concepts and theory related to this study to confirm the outcomes for the first objective. However, to implement all these activities, two sub-phases that belong to design science research are involved which are named as awareness of problem and suggestion. The following subsections will cover them in detail.

3.4.1.1 Awareness of Problem

The main processes involved in this phase have been summarized in Figure 3.3. This phase consists of two activities, namely identifying research gap and literature research, and content analysis. Accordingly, this stage starts with the review of the relevant literature to i) establish the context of the research that will present in this thesis; ii) provide a context for describing and elaborating the problem being identified. During this stage, the research problem, research questions that need to be addressed, and research objectives to provide a solution for the identified research problem are articulated this stage. The outcomes from this phase will be the research problem, research questions, and research objectives. The research problem, research questions as well as research objectives were discussed in Sections 1.3, 1.4 and 1.5 respectively.

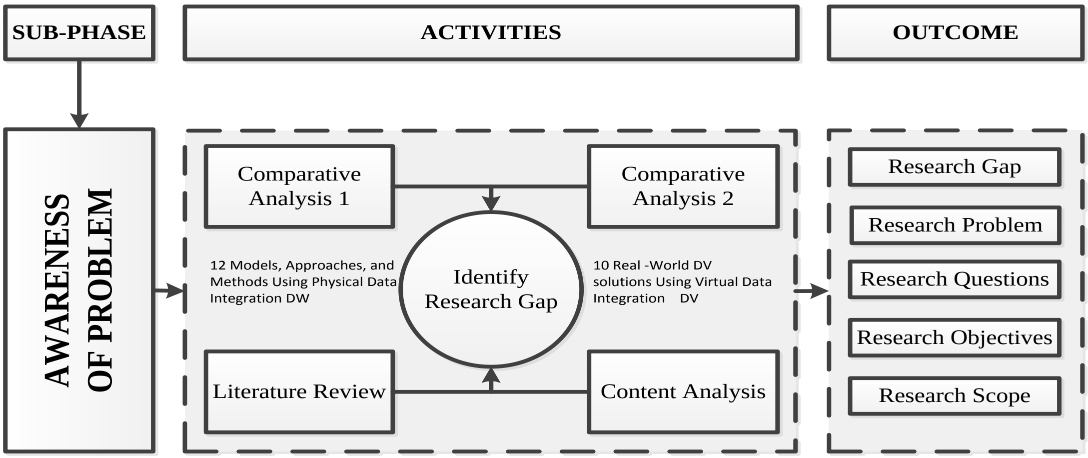


Figure 3.3. Awareness of Problem

3.4.1.1.1 Literature Research and Content Analysis

Content analysis is a process of extracting sufficient knowledge about the proposed study. It was performed to seeking for information and facts from various sources including books, articles. Moreover, the content can be found in various forms such as text, audio, and other components which are stored in different databases (Ariffin, 2009; Preece, Rogers, & Sharp, 2002). In this study, the main aim of the content analysis is to gather the concept and theories in proposing a DV design model in supporting decision making process in BI environment. On the other hand, theoretical framework is one of the criteria that needs to be fulfilled to find research gaps (Norshuhada & Shahizan, 2010). Other than that, the theoretical gaps as well as practical gaps in current scenario in BI environment are recognized. Current issues on BI development have been discussed in Chapter Two, where the findings are integrated with a number of theories concerned with the problems. The problems that have been identified generally should be of interest and relevance to more entities (Gacenga et al., 2012; Offermann et al., 2009). Finally, these existing theories are used as a basis to identifying the major issues, phases, flow, and components in

Universiti Utara Malaysia

developing the proposed model. The literature and content analysis are discussed in Chapter Two.

3.4.1.1.2 Comparative Analysis

This study adopts the technique applied by Ariffin (2009) and Syamsul (2011), in identifying the main components and entities from the existing models. Every design model has different special elements based on the application to be developed (Ariffin, 2009). In the context of this study, two comparative analyses were conducted; the first one specializes in comparing models with respect to the existing BI models in order to know all the obstacles and past trends of data integration process in BI environment. The comparative study included models proposed by (Aguilar, Ruiz, García, & Piattini, 2006; McGregor & Kumaran, 2002; McGregor & Scheifer, 2003; Pourshahid et al., 2008). While the second type of comparative analysis is restricted to real-world solutions that used the DV as a main store of data that is created during the building their BI systems, the purpose is to learn all the required components and all the elements that need to be saved and to understand the process of designing BI model. Moreover, the comparative analysis for all the reviewed models produced certain guidelines to ensure it matches with their target users. However, it was found highly lacking in providing the live and right data to be used in BI environment.

3.4.1.2 Suggestion

In this phase, the outcomes from awareness of problem are utilized in accomplishing the first objective. This phase includes two activities which are model suggestion and expert consultation. In the next paragraphs, we will cover them in detail. Figure 3.4 illustrates this phase.

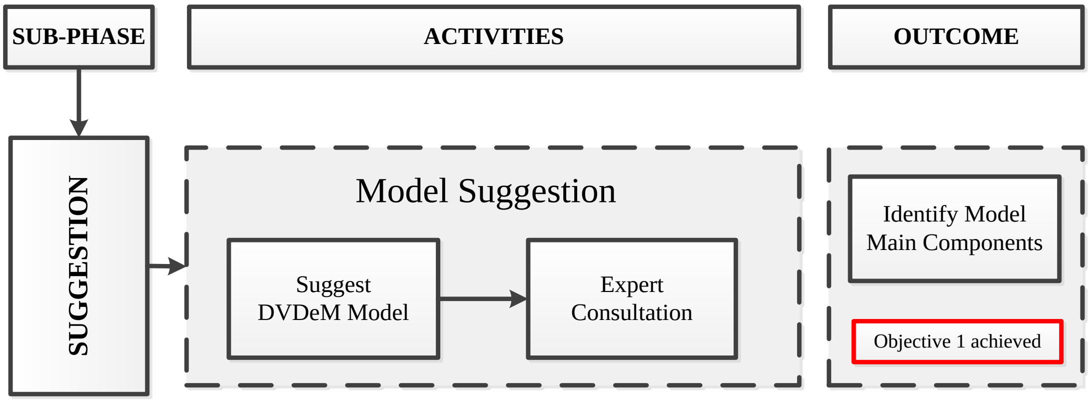


Figure 3.4. Suggestion Sub-Phase

3.4.1.2.1 Model Suggestion

The main purpose of this activity is to determine the appropriate elements of proposed model. At this stage, all components gathered that are related to BI environment were compiled and integrated into DV design model.

Universiti Utara Malaysia

3.4.1.2.2 Expert Consultation

Expert's consultation is one of the practices to identify the relevancies of the addressed problems. This consultation comprises of suggestion of ideas and concepts, current trends of BI environment, technologies involved, as well as reviews on research material. In this phase, experts from both the academia (6 respondents were obtained among the PhD. students at University Utara Malaysia who have either previously be a student or lecturer in software development / computer science as well as three IT lecturers. and industry (2 respondents from BI development company) were consulted to review and validate the suggested model and its elements. the demographic of experts that participated in the expert's consultation session are presented in Table 3.1.

Table 3.1: Demographic Profiles of Experts Consultation

No.	Field of Expertise	Experience (Year)	Location
2	Developers	15 Years	Malaysia
1	Academician	13 Years	Iraq
3	Academician	8 Years	Malaysia

3.4.2 Construction

The construction phase is considered as most challenging part of this study because it leads to the main contribution which is the DV design model process for near real time in BI environment and its prototype. The design process brings out the outcomes of proposed DVDeM Design Model. This phase involves the development of design artifacts and additional literature research (Figure 3.5). A solution in the form of artifact is designed and developed to overcome the research problems that have been identified. The artifact also addresses the research questions that were identified in in the first phase. The main outcome from this phase is a DV design model for near real-time decision making in BI environment (first version) that is validated by two methods (expert review and focus group discussion). In order to verify and evaluate the DVDeM proposed model (first version), the model validation process was performed; two validation method were used, expert review and focus group discussion. Chapter four, details the validation process.

Universiti Utara Malaysia

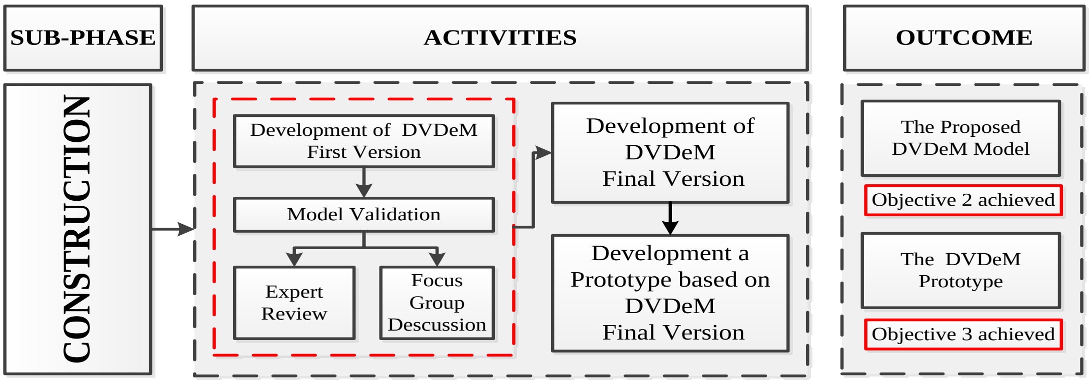


Figure 3.5. Construction Sub- Phase

3.4.2.1 Development of DVDeM

The construction of DV design model for near real-time decision making in BI environment is one of the specific objectives for this study. This model acts as the general framework through the presented functionality (Jones & Malik, 1992; Mantel, 1994). This design model contains components which were derived from the existing BI models, approaches, and methods. Many existing BI models, approaches, and methods as well as many real-world DV solutions have been reviewed in order to identify required phases, components, and model's functions for this model. Johnson and Henderson (2002) and Preece, Rogers, and Sharp (2007) state the design model as objects and relations between them answer the question how a system is organized and operated. The aim in proposing the DVDeM model is to provide near real time data in supporting decision making process in BI environment based on the organizations needs as well as to assist and guide the BI developers in designing BI systems. In the same aspect, the main outcome from the DVDeM model is to create virtual tables. Hence, the virtual table consists of the huge amount of data, in order to make a virtual table contain only the relevant data to minimize the access time the data as well as reduce the data retrieval time, therefore, in this study, the GODV approach was proposed. The proposed GODV approach consists of three steps: i) build goal structure, ii) matching the indicators with virtual tables, iii)

Deliver a new virtual table to the BI end users. The three steps of such approach have discussed in detail in Chapter 4. Accordingly, the proposed DVDeM model was validated in an iterative manner through experts review and focus group discussion. Hence, the next section highlighted on the model validation process

3.4.2.2 Model Validation

In verifying of the proposed DVDeM model as well as in order to revise and modify it, the validating process was conducted. There are different forms of approaches in evaluating and validating the research outcomes (Ariffin, 2009; Siti Mahfuzah, 2011; Syamsul Bahrin, 2011). However, two types of model validation methods were used in this study; the first one is expert review, while the second one is focus group discussion. In the context of this study, the experts in expert review method are not the same the participants in focus group discussion method, to obtain the reliable result in the model refinement and validation. Furthermore, this validation process is an extensive and iterative process, and is discussed in the next sections.

3.4.2.2.1 Expert Review

As mentioned in previous section, one of the model validation methods that has been used in this study is expert review. Expert review includes reviewing and validating the proposed DVDeM model and it has been recognized as a significant way to improve the quality of the developed application (Wiegers, 2002). In regard to this, the review was done iteratively for the DVDeM (first version) (refer to chapter 4, section 4.5.1). In this study, experts were chosen and the selection is based on two considerations. The main instrument used for this review was a questionnaire and the format of the questionnaire is adapted from Siti Mahfuzah (2011). It contains six questions asking about the: (i) terminologies used in the conceptual design model, (ii) relevancy of proposed elements in each component, (iii) connections and flows of the components, (iv) how usable is the proposed DVDeM to the development of BI environment, (v) how useful is the proposed DVDeM model to the implementation of BI environment and (vi) readability of the design model. Along with that, few demographic questions were also asked like name and gender, and years of experience. Experts were also encouraged to write their further comments in the provided instrument. The first consideration took into account selection of experts from the industry; experts were selected as they possessed sufficient experiences in developing DV applications. The second type of experts considered were academic experts who have experience in BI, DW, DV, and data integration, these experts has no less than eight years' experience in this field and from that pool, nine experts were chosen. Twelve experts are more than sufficient for this study as supported by Folch-Lyon & Trost (1981), Kitzinger (1995), and Morgan (1996).

Universiti Utara Malaysia

Meanwhile, the reviewing process was conducted in the following strategy; (i) prepare the review form by listing selection assessment attributes (ii) conduct the review sessions, (iii) and analyze the findings. In fact, evaluation form and e-mail were utilized as the media of communication with the experts. Finally, all data that were gathered from the expert review are tabulated, and later it was used to refine a DVDeM first version, and accordingly, the DVDeM proposed model (final version) was produced. Hence, the details on the DVDeM proposed model will be described in Chapter four. The data is documented as in frequency of responses of the expert review to the questions asked in the instrument. More detail regarding expert review form, questions is in Appendix A.

3.4.2.2.2 Focus Group Discussion

According to Morgan (1996), focus group is one of the accepted techniques used to gather qualitative data by means of group interaction on a matter determined by the researcher. Hence, this study uses the focus group discussion technique to evaluate and verify the proposed DVDeM model. The discussion was joined by six members in the field of BI, data integration, and database solution. These numbers are sufficient for the focus group discussion members of this study as supported by Folch-Lyon and Trost (1981), Kitzinger (1995), Morgan (1996), and Nielsen (1997). Selected participants evaluated the proposed model based on the questions provided in the review form presented to them. A question and answer session was conducted in order to deliberate further into the issues being discussed. In chapter 4, details and results from the focus group discussion are presented. Details regarding focus group discussion form and questions are appended in Appendix B.

Based on the responses from expert reviews as well as focus group discussions, the results were analyzed. Finally, all data that were gathered from the focus group discussions are tabulated, and later it was used in associated with the findings of the expert review together in refining a DVDeM first version, and accordingly, the DVDeM proposed model (final version) was produced.

3.4.2.3 Development of DVDeM Prototype

Dix et al. (2004) defined prototyping as the process of translating system's specification into a tangible outcome in order to gain users' feedback. The DVDeM prototype should give opportunity to the users to explore the benefits of using DV to improve BI and overcome the difficulties and problems caused by using DW or other data store techniques. Therefore, inclusive modeling approach will be utilized in this stage.

Inclusive modeling is known as user centered approach for agile software development. The key practices in Inclusive modelling are Active Stakeholder Participation. Stakeholder as defined by Scot (2002) is any person who is involved either directly or indirectly to the system development (direct, indirect user, manager, and operation staff member, funder of the project and support staff). Stakeholders also include developers who work on other system that interrelates with the project under development, as well as maintenance professionals that potentially affected by the development of a software project.

in the context of this study, the inclusive model approach is important, It is embodied in how to communicate with the BI stakeholders, collect the organization requirements and analyze organization goal.

Universiti Utara Malaysia

However, most of them will not recognize the complex diagrams used by most of the system developers. In this regard, inclusive model was adopted to help them in capturing and analyzing the systems requirements by using simple tools as well as simple techniques. Figure 3.6 shows the inclusive model used in system development where the stakeholders (product owner) will continuously inform and update from early process.

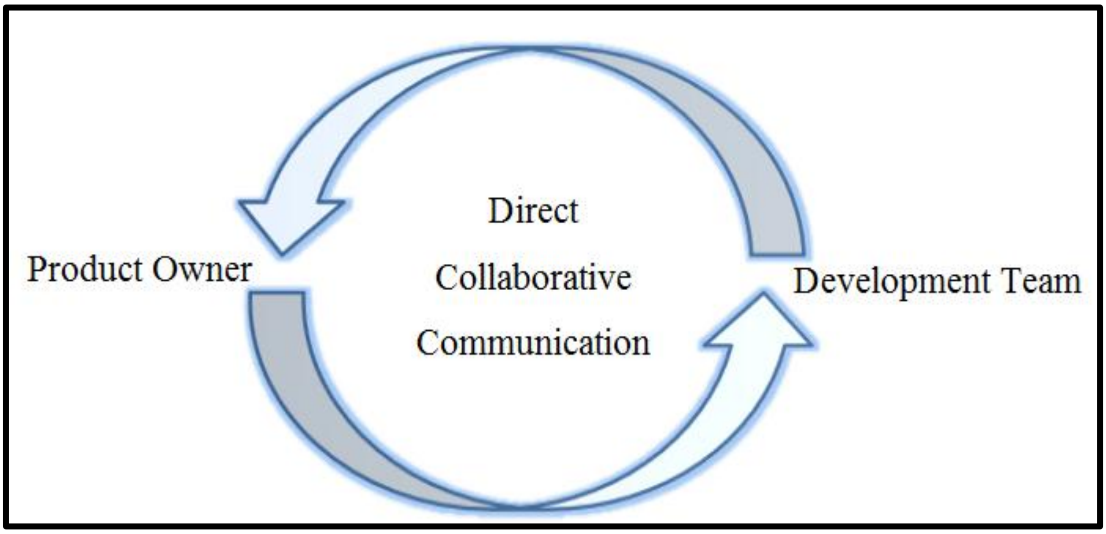


Figure 3.6. Inclusive Model for Prototype Development

Thus, the prototype was refined to produce a high-end prototype, which is evaluated and serves as a feeder to fulfill objective 3 in the next stage, evaluation phase. In general, and according to Summerville (2001), prototyping can be used as a tool and technique for analyzing risk and in turn reducing it. Figure 3.7 demonstrates the overall prototyping development processes and the activities associated with it (Baxter & Sommerville, 2011). In particular, the developed prototype demonstrated how such prototype can be implemented on standard hardware and software (like SQL Server and visual studio). Also, the DVDeM prototype demonstrates an implementation where the decentralized and autonomous nature of the virtual data integration supports the overall design objectives.

Universiti Utara Malaysia

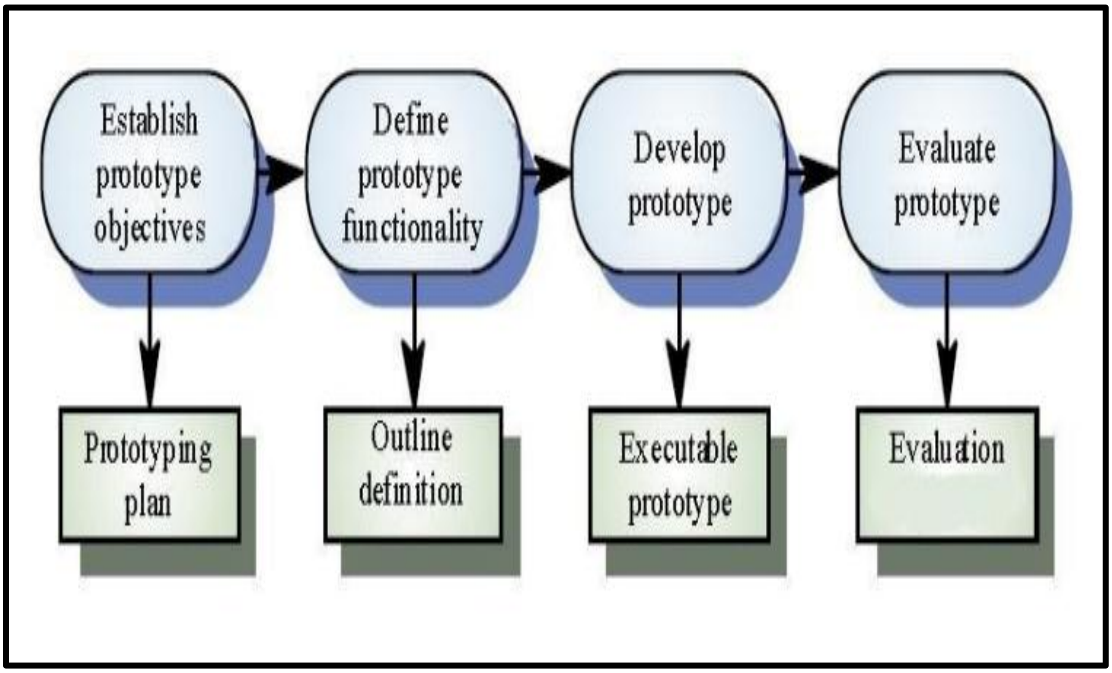


Figure 3.7. Prototyping Process (Baxter& Sommerville, 2011)

The proposed DVDeM model roles are as guideline to develop the prototype. Therefore, this phase is conducted to validate the proposed model. Later, the prototype was implemented, tested, and examined it and accordingly, use it by BI developers by guiding them in developing a BI system in BI environment. Accordingly, If the prototype is able to be achieved that means the proposed DVDeM model is applicable in developing BI applications to be used by BI developers in BI environment. Chapter 4, Section 4.8 detailed the DVDeM prototype development.

3.4.3 Empirical Testing

Empirical testing is the important part in this study as it ensures the proposed model is usable. Empirical testing includes two sub-phases which are evaluation and conclusion. The next paragraphs explain these sub-phases in detail.

UUM Universiti Utara Malaysia

3.4.3.1 Evaluation

Evaluation phase will be conducted when a proposed solution design reaches an adequate state. Ellis and Dix (2006), Dix (2002) states that three main goals should be derived from evaluation which are; (i) to assess the extent and accessibility of the system's functionality, (ii) to assess user's experience of the interaction , and (iii) to identify any specific problems with the system (Ellis & Dix, 2006) and (Dix, 2002). On the other hand, there are various types to evaluation and validation for the suggested solution, namely experimentation, demonstration, using metrics, simulation, benchmarking, logical reasoning, and mathematical evidence. These approaches vary in terms of their appropriateness and strength (Ariffin, 2009; Vaishnavi & Kuechler Jr, 2007).

In this study, the evaluation stage consists of many sub-phases. As mentioned in the development phase, the ultimate mission in the development phase is to develop a prototype based on the proposed DVDeM model. In the evaluation phase, the prototype should be tested before delivered to the end user. As mentioned in Chapter 2, the software testing in BI environment includes testing the usability. However, the usability strategy includes, reviewing existing models and approaches for BI usability testing, prepare the usability testing instrument by selecting assessment attributes, conducting the usability testing, and finally, analyze the finding. This strategy is illustrated in Figure 3.8.

Universiti Utara Malaysia

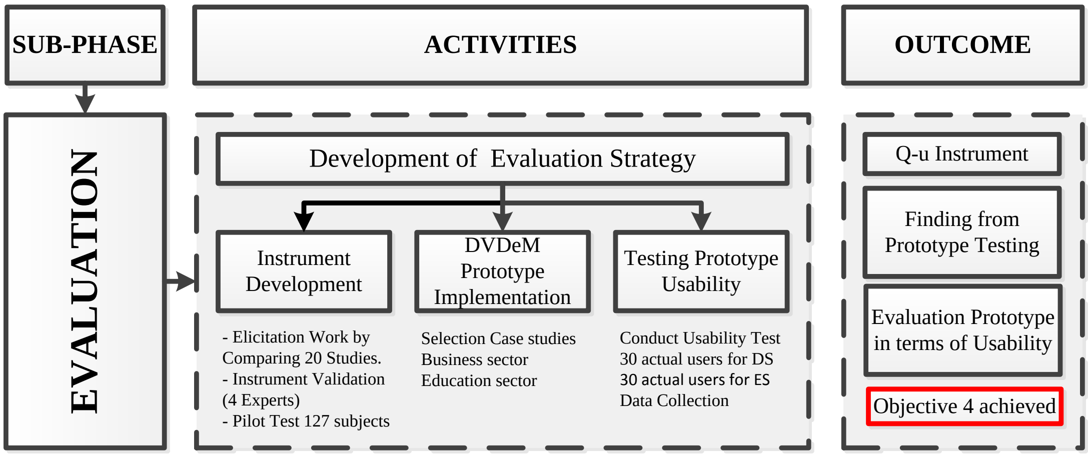


Figure 3.8. Evaluation Sub-Phase

3.4.3.1.1 Instrument Development

This study utilizes a questionnaire as the main evaluation instrument to investigate whether the DVDeM is effective and workable in BI environment and make BI developers feel motivated to using DVDeM in terms of usability. Generally, Oppenheim (1983) has described questionnaire as a widely research instrument that is reliable in terms of data collection. Throughout the review process, two mediums of communication were used email review and face to face review. Consequently, in validating the proposed DVDeM model, a questionnaire was adapted by considering both design and implementation of BI model. To design the evaluation instrument, a systematic approach was adapted as suggested by authors Sekaran (1992), Creswell (2003), Ariffin (2009), Siti Mahfuzah (2011), Jooste (2013), and Nurulnadwan (2014). In which the rigor process of evaluation instrument development is illustrated in Figure 3.9.

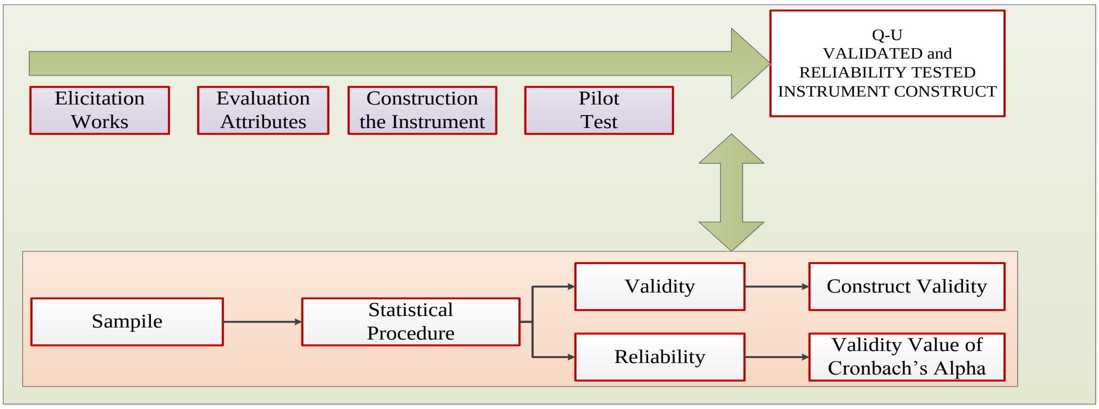


Figure 3.9. Summary of Instrument Design

As illustrated in Figure 3.9, the design of the instrument began with the elicit works to determine the dimensions and items of the evaluation instrument. Then the drafted instrument was piloted for a test of validity and reliability. In this pilot study, 67 respondents were obtained among the postgraduate students at Universiti Utara Malaysia who have either previously be a student or lecturer in the field of computer science, software engineering, or IT. The respondent numbers who have participated in the pilot study is enough to achieve a reliable outcome in the statistical test as described by Sekaran (1992). Accordingly, the instrument is ready to measuring usability of DVDeM prototype.

A. Elicitation Work

In this study, six attributes of usability are considered, which are visibility, flexibility, learn-ability, application behavior, error control and help, and near real-time decision making. They have been proposed to test the usability of DVDeM prototype. The proposition of the attributes was elicited from previous studies related to usability evaluation of BI applications. A total of 20 evaluation works were reviewed particularly on the attributes used in the study. The findings are tabulated and summarized as shown in the Table 3.1.

Table 3.1
Summary of Usability Attributes

	Dimensions / Authors	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	Total
1	Visibility	√		√	√	√		√	√	√	√	√	√		√	√	√	√	√			15
2	Flexibility	√	√		√	√	√		√	√	√	√			√		√	√		√	√	13
3	Learnability	√	√	√	√		√	√	√		√	√	√	√		√	√	√	√		√	15
4	Application Behavior	√		√	√		√			√	√		√		√		√		√		√	11
5	Error Control and Help	√		√		√		√	√	√	√	√				√		√	√			11
6	Near real-time Decision Making		√		√	√	√				√			√	√	√	√	√			√	11
7	Helpfulness							√	√	√		√	√			√			√			7
8	Efficiency			√			√		√				√		√		√					6

Note :(√) mean the attribute is utilize in this study.

Where: studies A is (Jooste et al., 2013), B is (Rouhani et al., 2012), C is (Nagl & Marquardt, 2008), D is (Ghazanfari, 2011), E is (Gould & Lewis, 1985), F is (Gulliksen et al., 2016), G is (Hwang & Salvendy,2010), H is (Hou, 2012), I is (Lewis, 1995), J is (Lavery et al., 1996), K is (Poppe et al., 2007),L is (Lutsch, 2011), M is (Moczarny, 2011), N is (Omerali, 2012), O is (Karahoca, 2008), P is (Bak et al., 2008), Q is (Rogers et al., 2007),R is (Ssemugabi & De Villiers, 2007), S is (Lin et al., 1997), and finally, T is (Amor, 2014).

Universiti Utara Malaysia

Based on the meta-analysis in Table 3.1, this study decided to select attributes which indexed (1 to 6) to measure the usability of the prototype based DVDeM as they have the highest frequency.

B. Proposed Operational Definition

Based on the meta-analysis and its findings shown in Table 3.1, the operational definition of selected dimensions is listed in Table 3.2.

Table 3.2
Operational Definition on Selected Dimensions

	Dimensions	Operational Definition
1	Visibility	Display of information and interface design of the system. Is the ability to accurately and completely view the processes,

		transactions and other activities operating via using BI application within an enterprise.
2	Flexibility	Customizable and user control of the system. It's the ability of software to change easily in response to different user and system requirements.
3	Learnability	Easy to learn. The ease with which a user can learn to operate, prepares inputs for, and interprets outputs of a system or component.
4	Application behavior	Measuring the system behavior in terms of enhance user efficiency through a consistently rapid response rate.
5	Error Control & Help	System ability in error prevention and error recovery.
6	Near Real-Time Decision Making	System ability to deliver live data, relevant data, and near real-time data to support decision making process; for users, who have never been exposed to a system or at least have little exposure to the system, is increasing as well as system ability in how quickly the task performance.

As stated previously, this study adapts the existing instruments to measure prototype usability. Therefore, to gather relevant items for the constructs, various existing questionnaires were studied and come out with the first draft of the instrument.

Universiti Utara Malaysia

C. Instrument Validation via Face Validity

The first draft of the instrument was then validated through Content Validity and Inter-Item Consistency Analysis. Sekaran and Bougie (2010) consider face validity as the fundamental measuring approach for Content Validity. This is to ensure that the measuring instrument includes a sufficient and present a set of measuring items of intended concept. Consequently, this study engaged four experts in various expertise which are: BI Developers, Software Engineering, Information Technology, and Multimedia, through e-mails as well as face to face consultation to review the items in terms of Content Validity. This is inline with the suggestion of Schneiderman (1992), Zainuddin, Zaman and Ahmad (2011), and Cole and Lewis (2012); where three to five experts were employed for their content validation. From the feedback of the experts, it was found that some of the items were not good enough to use and not fit well with the intended constructs. This led to some modifications to the first draft. The modifications involved rewording, repositioning, and discarding some contents in the instrument. The expert’s comments and suggestions are shown in Table 3.3.

Table 3.3
Comments from Face Validity

	Experts Comments and Suggestion	Applying Comments and Suggestion
1	The question words and question marks are stated in the sentences.	All question words and question marks were deleted from the questionnaire to make the sentence straight forward.
2	The sub- question "do you think" is stated in the sentences.	All questions which contain "do you think" have been modified.
3	The word “application" should be replaced with word "system".	All word “application" has been replaced with word "system".
4	The word “during use" should be removed.	All questions which contain "during" have been modified.
5	Some of the grammar errors need to pay attention.	Correct all the grammar errors.
6	The attribute "Business Intelligence and Decision making " should replace with term “near real time decision making "	The attribute "Business Intelligence and Decision making” was replaced with term “near real time decision making".

The instrument was then named as questionnaire to measuring the DVDeM prototype usability (Q-U), that contains attributes that test the proposed DVDeM prototype (Usability: visibility, flexibility, learnability, application behavior, error control & help, and near real-time decision making).

Schreiber and Asner-Self (2011) has suggested that there is no debate on whether a middle or neutral choice is acceptable. This study used "neutral" in the middle category, to avoid forced the respondents to choose a side, while neutral choice is appropriate with these constructs of this study. Besides, Schreiber and Asner-Self (2011) has suggested that there is no debate on whether a middle or neutral choice is acceptable. In the context of this study, the instrument has 5 scales, Hence, a 5-point Likert scale ranging from strongly disagree (denoted by 1), to strongly agree (denoted by 5) is used in the study. As mentioned earlier, the instrument is partially used as a measure of outcome; hence, scale sensitivity becomes an important concern (Cummins & Gullone, 2000). The overall (Q-U) instrument after refinement is shown in Table 3.4. When concerns with scale reliability, Cicchetti et al. (1985) reported that using response options beyond 5- point do not significantly alter the scale reliability. However, difficulties might arise in generating categorical names as the scales expanded (Cicchetti, Shoinralter, & Tyrer, 1985; Cummins & Gullone, 2000).

Table 3.4
Draft of Q-U Instrument

	Q-U Usability Instrument	Scales (1-5)					Sources
1	**VISIBILITY**						
	1. This system can display the information (Virtual Tables, Virtual Data Mart, and BI Reports) in an uncluttered and well-structured manner.	5	4	3	2	1	I
	2. All instructions are visible and self- explanatory.	5	4	3	2	1	I
	3. Navigation options in this application such as (Links, shortcuts, home, back, forward, etc.) are displayed in visible manner.	5	4	3	2	1	I
	4. The system is able to communicate the status at all times (whether resting, processing etc.).	5	4	3	2	1	A
	5. In this system, the data is concisely presented.	5	4	3	2	1	I
	6. This system has all the functions and capabilities I expect it to have.	5	4	3	2	1	I
	7. I like using the interface of this system	5	4	3	2	1	J
	8. The interface for this system is pleasant.	5	4	3	2	1	I
	9. The organization of information in the system screen was clear.	5	4	3	2	1	I
2	**FEXIBILITY**						
	1. I felt fully in control when using this system.	5	4	3	2	1	S
	2. This system has customizable feature.	5	4	3	2	1	I
	3. In this system the design for data entry is flexible.	5	4	3	2	1	I
	4. The data can be used in used, manipulated, and/or processed in easy manner.	5	4	3	2	1	S
	5. This system handles user-specified windows.	5	4	3	2	1	S

Table 3.4 Continued

3	**LEARNABILITY**						
	1. In this system the data grouping reasonable for easy learning.	5	4	3	2	1	I
	2. In this system promotes learnability to make it accessible for infrequent usage.	5	4	3	2	1	S
	3. In this system, the grouping of menu options is logical.	5	4	3	2	1	S
	4. I believe I could become productive quickly using this system.	5	4	3	2	1	I
4	**THE APPLICATION BEHAVIOR**						
	1. This application enhances user efficiency through a consistently rapid response rate.	5	4	3	2	1	A
	2. The application behavior is consistent.	5	4	3	2	1	A
	3. The information was effective in helping me complete the tasks and scenarios.	5	4	3	2	1	A
	4. It was easy to find the information I needed.	5	4	3	2	1	I
	5. Whenever I made a mistake using the system, I could recover easily and quickly.	5	4	3	2	1	J
	6. I was able to complete the tasks and scenarios quickly using this system.	5	4	3	2	1	A
5	**ERROR CONTROL & HELP**						
	1. This system has ability for error prevention and error recovery.	5	4	3	2	1	A
	2. The information (such as online help, onscreen messages and other documentation) provided with this system was clear.	5	4	3	2	1	A
	3. The system gave error messages that clearly told me how to fix problems.	5	4	3	2	1	A
	4. The application provides a help on demand.	5	4	3	2	1	I
	5. Overall, I am satisfied with this system.	5	4	3	2	1	I
	6. Overall, I am satisfied with how easy it is to use this system.	5	4	3	2	1	A
6	**NEAR REAL-TIME DECISION MAKING**						
	1. In this system, the knowledge sharing is allowed.	5	4	3	2	1	A
	2. This System provides the information visualization functionality (comparison charts, graphs to reveal trends etc.) to assist in decision making.	5	4	3	2	1	A
	3. In this system, the breadth and depth of the data provide sufficient coverage for all data resources.	5	4	3	2	1	A
	4. In this system, data is received on time to take suitable actions and decisions.	5	4	3	2	1	A
	5. In this system, the data is always live data and up to date.	5	4	3	2	1	A
	6. In this system, the data was cleaning, cleansing, and profiling, so there are almost no errors in the data.	5	4	3	2	1	A

Universiti Utara Malaysia

Note: Interpretations of the scales: 5= Strongly Agree, 4= Agree, 3= Neutral, 2= Disagree, and 1= Strongly Disagree.
A- (Jooste et al., 2013). T- (Amor, 2014). I- (Lewis, 1995).
S- (Lin, Choong, & Salvendy, 1997).

In addition, the BI usability testing instrument consists of two general questions asked to the respondents.

i. I agree that the system based on DVDeM can help users to create the intended reports in right time.
ii. I agree that the system based on DVDeM can be used in business intelligence environment.

Both of general questions use a nominal scale, in which a respondent is required to provide either 'yes' or 'no' answer. Considering the interest of this study, which is to examine how the prototype based on the proposed DVDeM model, is workable in BI environment. Therefore, these general questions are important to know the majority feedback towards the proposed model.

Besides, the following demographic profiles of respondents which are name and gender were also gathered and assessed towards measuring the BI usability of the prototype based on DVDeM design model. The demographic questions also use the nominal scale as means to gather intended information. Next, the Q-U instrument was pilot tested to examine its goodness of measures.

D. Pilot Study: Testing Goodness of Measures of Q-U Consistency

In order to ascertain that the instrument indeed measures the required variables or constructs, a pilot study has been conducted to measure their consistency (Sekaran & Bougie, 2011). Accordingly, the instrument was used to test the usability of the prototype based on DVDeM proposed model. It is important for this study to determine the consistency of instrument to be used, because it is an adaptation of other works. Consequently, subjects and result of consistency analysis are detailed separately by the results of validity and consistency test.

In this pilot study, 67 respondents were involved among the postgraduate students who have either previously be a BI developer, lecturer in computer science, IT, or software engineering. The respondent numbers who have participated in the pilot study is enough to achieve a reliable outcome in the statistical test as described by Sekaran (1992). According to Hair et al. (2006); the sample size for construct validity test should at least have 100 responses to get a reliable significant outcome. Hence, this study adopted the approach of Siti Mahfuzah (2011) by adding up the data from the main study with the data from the pilot study, in order to meet the required number of respondents for construct validity via factor analysis and having a more accurate reading. From the users' experiment on DVDeM prototype via two case studies, 60 total numbers of respondents participated in the study. It gives a total number of 127 respondents' altogether (with pilot study respondents). Then validity is assured and the factor analysis can be considered for recommending items to deletes and retains.

i) Factor Analysis (Validity)

The objective of determining factor analysis was to verify the degree of significance of each item and which are most suitable for each dimension (Sekaran & Bougie, 2011). Therefore, the test was run and guided for accepting each item based on utilizing Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity, Measure of Sampling Adequacy (MSA), and Factor loading.

Universiti Utara Malaysia

KMO test was investigated to determine sampling adequacy that suggests whether the partial correlation among variable are small. Bartlett's test of sphericity was investigated to determine whether the correlation matrix is an identity matrix that would indicate that the factor model is appropriate or inappropriate and worth continuing with the factor analysis as there is a relationship to investigate, while the factor load is determined to signify the correlation between the measuring item and its intended factor. In order to determine the significant level of the constructs, Hair et al. (2006) suggest the practical significance of the loadings rather than the conservative significance test, and the following condition must be met to accept the measurement items.

i. KMO test ≥ 0.50
ii. Bartlett's test of sphericity, the significant value of $p \leq 0.05$
iii. Factor loading of $\pm$ 0.30 to $\pm$ 0.40 are minimally acceptable, hence value greater than or equal to 0.50 are generally considered essential for practical significance.

Universiti Utara Malaysia

In preparing the data for factor loading analysis, KMO test was conducted and the results are tabulated in Table 3.5. It can be noticed that all the values for KMO test satisfy the condition of KMO test ≥ 0.50.

Table 3.5
KMO Test and Significant Values

	Constructs / Dimensions	KMO	Significant value of Bartlett's Test of Sphericity
1	Visibility	0.644	0.000
2	Flexibility	0.619	0.000
3	Learnability	0.621	0.000
4	Application Behavior	0.661	0.000
5	Error Control & Help	0.623	0.000
6	Near Real-Time Decision Making	0.653	0.000

In addition, the Bartlett's test of sphericity gave the significant value of 0.000 for all constructs, which shows the second condition also met and satisfy (significant value of $p \leq 0.05$). Hence, this evidences that the data are ready for factor loading analysis test. Therefore, it was executed and the results of the test are detailed in Table 3.6.

Table 3.6
Factor Loading for Each Item in Q-U

Q-U Usability Attributes and Items	Factor Loading
1 VISIBILITY	
1. This system can display the information (Virtual Tables, Virtual Data Mart, and BI Reports) in an uncluttered and well-structured manner.	0.617
2. All instructions are visible and self- explanatory.	0.690
3. Navigation options in this system such as (Links, shortcuts, home, back, forward, etc.) are displayed in visible manner.	0.698
4. The system is able to communicate the status at all times (whether resting, processing etc.).	0.713
5. In this system, the data is concisely presented.	0.699
6. This system has all the functions and capabilities I expect it to have.	0.695
7. I like using the interface of this system.	0.729
8. The interface for this system is pleasant.	0.747
9. The organization of information in the system screen was clear.	0.699
2 FLEXIBILITY	
1. I felt fully in control when using this system.	0.742
2. This system has customizable feature.	0.630
3. In this system, the design for data entry is flexible.	0.701
4. The data can be used in used, manipulated, and/or processed in easy manner.	0.697
5. This system handles user-specified windows.	0.714
3 LEARNABILITY	
1. In this system, the data grouping reasonable for easy learning.	0.675
2. In this system promotes learnability to make it accessible for infrequent usage.	0.715
3. In this system, the grouping of menu options is logical.	0.658
4. I believe I could become productive quickly using this system.	0.723
4 THE APPLICATION BEHAVIOR	
1. This system enhances user efficiency through a consistently rapid response rate.	0.678
2. The system behavior is consistent.	0.723
3. The information was effective in helping me complete the tasks and scenarios.	0.687
4. It was easy to find the information I needed.	0.698
5. Whenever I made a mistake using the system, I could recover easily and quickly.	0.652

Table 3.6 Continued

	6. I was able to complete the tasks and scenarios quickly using this application.	
5	**ERROR CONTROL & HELP**	
	1. This system has ability for error prevention and error recovery.	0.751
	2. The information (such as online help, onscreen messages and other documentation) provided with this system was clear.	0.673
	3. The system gave error messages that clearly told me how to fix problems.	0.685
	4. The system provides a help on demand.	0.754
	5. Overall, I am satisfied with this system.	0.659
	6. Overall, I am satisfied with how easy it is to use this system.	0.716
6	**NEAR REAL-TIME DECISION MAKING**	
	1. In this system, the knowledge sharing is allowed.	0.662
	2. This system provides the information visualization functionality (comparison charts, graphs to reveal trends etc.) to assist in decision making.	0.713
	3. In this system, the breadth and depth of the data provide sufficient coverage for all data resources.	0.657
	4. In this system, data is received on time to take suitable actions and decisions.	0.663
	5. In this system, the data is always live data and up to date.	0.725
	6. In this system, the data was cleaning, cleansing, and profiling, so there are almost no errors in the data.	0.658

ii) Result of Consistency Analysis (Reliability Test)

Reliability of a measure is an indication of consistency. In the pilot study, the measure of consistency is examined through the interim consistency reliability test. The value of Cronbach's coefficient alpha was computed and should indicate the value of alpha to be accepted as reliable (Sekaran, 1992). Thus, the reliability of a measure signifies the level at which the measure is without bias and therefore offers dependable measurement across different items of the instrument (Cavana, Delahaye, & Sekaran, 2001). Hair et al. (2010) and Byrne (2010) stressed that the major measures for selecting past instruments revolves around the personal internal coherence obtained through the Cronbach's Alpha reliability coefficients calculation, in which the strength is listed in Table 3.7 (Hair, Black, Babin, Anderson, & Tatham, 2006).

Table 3.7

Criterion for acceptability of Alpha Coefficient

	Alpha Coefficient Range	Strength of Association
1	< 0.6	Poor
2	0.6 to < 0.7	Moderate
3	0.7 to < 0.8	Good
4	0.8 to <0.9	Very Good
5	0.9	Excellent

Source (Hair, Black, Babin, & Anderson, 2010; Hair et al., 2006).

This study ran the Cronbach's alpha test and set ($\alpha > 0.6$) to be significant as suggested in Table 3.7 (Hair et al., 2006; 2010). Table 3.8 shows the results of reliability test of measurement items, they were found consistent and significant, hence, this measurement items are able to use for data collection in the main study.

Table 3.8

Reliability Test of Measurement Items

	Constructs / Dimensions	Cronbach's Alpha	No. of Items before factor loading	No. of Items after factor loading
1	Visibility	0.709	11	9
2	Flexibility	0.719	6	5
3	Learnability	0.725	7	4
4	Application Behavior	0.771	8	6
5	Error Control & Help	0.723	7	6
6	Business Intelligence	0.745	7	6

While this section extracts important part of the results, the detailed results of the pilot study are attached in Appendix C.

As seen in Table 3.6, all the items in Q-U are found valid and can be used to represent respective dimensions. As stated earlier, factor loadings ≥ 0.50 are considered practically significant and well-defined structure (Hair et al., 2006). Thus, all the items that show loading values less than 0.50 are not included in the test of DVDeM prototype usability. Since the items proposed in questionnaire were elicited from various previous works, hence it is important to seek confirmation (through

Universiti Utara Malaysia

factor analysis) to see if these items underlie that proposed dimensions in questionnaire. The overall Q-U instrument is available in Appendix C.

3.4.3.1.2 DVDeM Prototype Implementation

The prototype needs to be test before it is used. The test process involves a thorough examination and evaluation. The justification for embedding the case studies in this study is (i) case study is used widely in both organizational studies as well as in the social sciences; (ii) case study can be exciting and provides its users with abundance of data; (iii) there is a large and growing confidence in the case study as a strategy of rigorous research in itself (Eisenhardt, 1989; Iacono, Brown, & Holtham, 2011; Voss, Tsikriktsis, & Frohlich, 2002). In the same aspect, the prototype implementation in case studies can show the significant value and tangible benefits that can be achieved through the applying DVDeM proposed model. With regards to this, the use of DVDeM proposed model is illustrated with two case studies: the first one in education sector, while the other one in business sector.

Universiti Utara Malaysia

A. Business Intelligence in Business Sector

BI is a concept that usually involves the delivery and integration of relevant and useful business information in an organization. in business sectors, companies use BI to detect significant events and identify/monitor business trends to adapt quickly to their changing environment and a scenario. Hence, the effective use of BI in the business sector can improve the decision-making processes at all levels of management and improve tactical strategic management processes and finally support decision-making process by providing near real-time decision-making data. Furthermore, there are many reasons for applying Bi in business sector:

i. Gain insights into consumer behavior. One of the main advantages of investing in BI and skilled personnel is the fact that it will boost the ability to analyze the current consumer buying trends. Once understand what consumers are buying, the decision makers can use this information to develop products that match the current consumption trends and consequently improve your profitability.
ii. Boost productivity. Through traditional data gathering methods, users need to compile and analyze data and write related reports. This can be incredibly time-consuming, especially for small businesses that may not have the employees to do it. With a BI program, the stakeholders can pull data and create the reports at the click of a button thus freeing up time and resources allowing employees to be more productive on their own tasks.
iii. Return on Investment (ROI). Through better strategic awareness, faster reporting decreased operating costs / lower overheads and access to better quality data and information, BI can positively influence a company's ROI.

Universiti Utara Malaysia

In line with the above situations, in order to reap all the benefits of an effective BI system and strategy for business sectors, it is most important to develop BI system that has the ability to provide near real-time data in supporting decision-making process and BI system should be designed for analytical efficiency and accessibility.

B. Business Intelligence in Education Sector

BI attempts to extract meaning from mountains of bits and bytes. Developments in transmission, collection and storage of data and the ability to analyze it allow leveraging that data to achieve the business or policy objectives. In Education sector, knowledge workers in today's marketplace must have access to the right information at the right time to make the right decisions, and all the while avoid "drowning" in information overload. BI is the principal discipline that helps them get there. For many educational institutions in the education sector, it has become essential to have large-scale automated data monitoring and report generating functionality incorporated into existing operational systems to manage a multitude of data. BI system provides a comprehensive software application for all education analytics and reporting requirements of administrators and scholars alike.

In line with the above situations, BI in the education sector can support in delivering near real time data to improve tracking and reporting of learning management system, student registrations, examination performance, and workforce effectiveness. Track leading indicators to adjust operations to market changes and resource requirements. Deliver improved marketing return on investment by enhanced visibility of recruitment performances. Reduce operational costs through improved financial reporting and procurement efficiency.

In the context of the education and business sectors, the current business environment is constantly evolving. The global economic scenario is providing opportunities as well as challenges. The factors affecting the business environment are consumer needs, globalization, and government policies. In such a business environment, organization basically has four action steps. The organization can be reactive, anticipative, adaptive, or/and proactive. For this, an organization can develop a new strategy, get into partnership. Today most of the businesses are having a computerized business support. This support is in form of decision support system, business analysis. The main objective of BI is to bridge the gap between organization status and its desired position. as well as BI helps the organization achieve commercial success along with sound financial management.

C. Selection of Case Studies

As mentioned early, the DVDeM prototype was implemented in two case studies; the education and the business sectors. These two sectors were selected because the data integration system (data sources and data processing) in these sectors are dynamically updated and the lack of providing near real-time data will affect negatively the decision-making process. The importance of near real time each sector is highlighted in the next paragraphs.

In the education sector, providing the near real-time data will enable the tracking and predicting of lecturers' performances through their data traces (data integration). Accordingly, finding a new data integration approaches is an effort to better govern and manage education globally. Undoubtedly, timely decisions require enhancing data integration system and providing live data in a right time.

In the business sector, near real-time data is important in supporting the decision-making process, because it enable loan management system to do near real-time monitoring, analytics, and data refresh method which includes importing user loan data, validating the imported loan data, and loading the validated loan data into a loan queue under control of the user.

In line with the above situations, it can be concluded that there is enough and strong evidence to justify on the selection of business and education sectors as case studies. Furthermore, in order to provide a complete and rich implementation for case studies, the case studies profile should include: organization background, business problem, existing application, the proposed DV solution, the implementation process, and the benefits (Davis & Eve, 2011).

D. The Framework of Business Intelligence

More and more businesses are moving towards BI. The reason for this movement is the business environment. Organizations are forced to capture, store and interpret data. This data is at the core of business success. Organizations require correct and near real time information for any decision-making process.

BI framework combines gathering and analyzing the organization requirements, extracting data from multiple data sources, data storing, business analytics, performance, strategy and user interface. The business receives data from various sources. This data is captured in the virtualization where it is stored, organized and summarized as per further utilization. Authorized users can access this data and work on it to get desired results. This result than are shared to executives for the decision-making process. These data results can be published through BI reports, BI dashboards or share points.

Overall, the framework of BI that used in both of business sector and education sector are same in terms of data integration issues except the differences in the organization's requirements and the related raw data. Besides, each of these case studies has different characteristics which will be explained in detail in Chapter 5.

Universiti Utara Malaysia

3.4.3.1.3 Testing Prototype Usability

As described previously, the usability testing for the prototype was conducted. The measurements were made through an instrument named Q-U (refer to 3.4.3.1.1). Meanwhile, Q-U instrument comprises of six main dimensions, visibility, flexibility, learnability, application behavior, error control and help, and near real time decision making as usability attributes and 36 items spared over these attributes (see appendix C). The instrument was handed to 60 respondents spread over two case studies, each case study 30 respondents were obtained from the actual users (BI developers, IT manager, and BI users); they were required to answer questions after examining the prototype. According to Nielsen (2012), in quantitative studies which aiming at statistics, not insights, the test at least 20 users to get statistically significant numbers; tight confidence intervals require even more users. therefore, in the context of this study, 60 respondents spread over two case studies, is sufficient as supported by (Nielsen, 2012) and (Roscoe, 1975). The details about prototype usability testing were discussed in Chapter 5.

3.4.3.2 Conclusion

In the final stage, empirical evidence from quantitative data via the questionnaire was justified through analysis of finding. Research findings are being answered here by data analysis.

In analyzing the data, the data in the questionnaire were coded and analyzed. Each of the data reviewed and coded comprehensively to highlight the views of visibility, flexibility, learn-ability, application behavior, error control and help, and business intelligence as parts of usability of the proposed DVDeM model. The quantitative data analysis software was used to carry out the data management, coding and analysis. Data obtained from the coding were used to analyze the subjects in the groups. A database consisting of texts with its associated codes were created using data analysis software. The software assisted in organizing and retrieving the portions of the data that linked to the common codes. Then, the final stage is to analyze the results of the evaluation phase by using one of the techniques of data analysis. In the context of this study, t-test and descriptive analysis were utilized. Figure 3.10 illustrates the conclusion phase.

Universiti Utara Malaysia

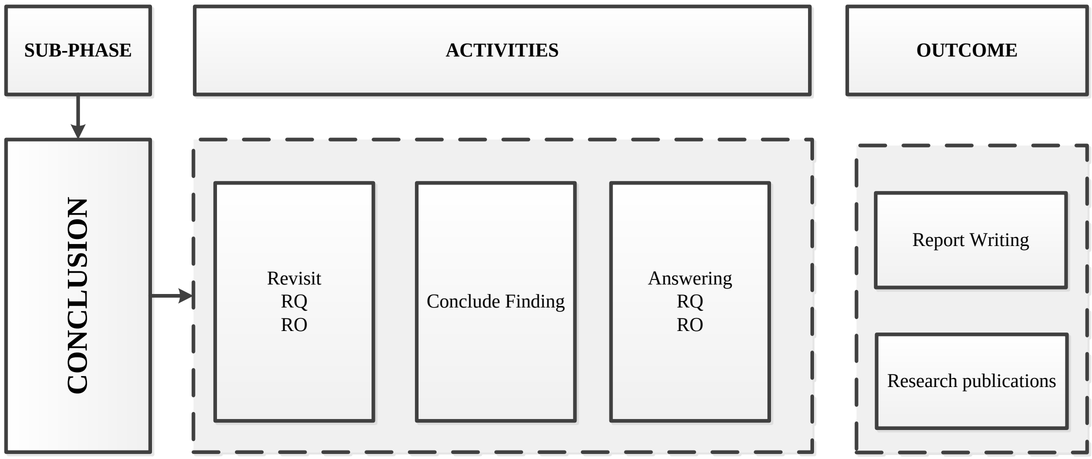


Figure 3.10. Conclusion Phase

In the other aspect, in interpreting these scores, the gap classification of interval scales is considered. As explained in Karlin and Altschul (1990), Zhang and Skolnick (2004), Zulkarnain (2001), and Siti Mahfuzah (2011), the gap classification of interval scales used in research instrument follows the formula below:

Gap = (Highest score – Lowest score) / Number of scales. The details described in the Chapter 5.

Universiti Utara Malaysia

All the findings gathered in each of the previous phases were concluded through revisiting and answering all the research questions and research objective. Finally, this study produces the full thesis and several publications.

3.5 Summary

In a nutshell, this chapter deals with the research approaches that are going to be adapted in this study. This chapter provided a detailed discussion of the research approach and methodology followed in this research.. Design science research has been adapted which includes three main phases namely, theoretical study which includes two sub-phases which are awareness of problem and suggestion. The next phase is development where construction or development the proposed model is performed. Finally, an empirical study which includes two sub-phases which are evaluation and conclusion are carried out. Each of these phases is described in detail with the all activities that are performed throughout this study. The next chapter proposed the DVDeM model and the components and phases of DVDeM are determined in detail.

Universiti Utara Malaysia

CHAPTER FOUR
DEVELOPMENT OF MODEL

4.1 Introduction

This chapter mainly details out the proposed DVDeM model which includes phases, components, activities, and deliverables. It explains development and validation process of the proposed model. On the other hand, this model is proposed intentionally for guiding developers in developing BI applications. A better DVDeM application delivery is also expected by implementing the proposed model. Expert review and focus group discussion methods have been considered which serve to validate this model.

The development process of the model was based on the problem and solution that have been discussed in chapter 1. Comparative analysis for the purpose of determining the characteristics of the proposed model have been discussed in Chapter 2. In this chapter, expert consultation and focus group discussion with BI experts and developers will be described.

In general, the methods have been adopted to validate the proposed model include the review the literature that consist of BI studies, DW, DV, DSS, requirements gathering, and experts consultation that involved experts review.

Some of the activities have been conducted prior to proposing the model phases, model flows, model component, and model activities. These activities are tabulated in Table 4.1.

Universiti Utara Malaysia

Table 4.1
Activities Prior to Proposing DVDeM Model

	Elements	Activities Details	Chapter No.
1.	Model Main phases	Experts consultation Comparative study of existing BI models.	CH 2 & CH 4
2.	Model Components	Content analysis of the literature Comparative study of real-world DV solutions to develop BI models.	CH2 & CH4
3.	Model Flows	Experts review & focus group discussion Comparative study of real-world DV solutions to develop BI models. Comparative study of existing BI models.	CH 2 & CH 4
4.	Model Activities	Experts review. focus group discussion Comparative study of real-world DV solutions to develop BI models.	CH 2 & CH 4

In the previous chapter, those selected models and approaches have been discussed in detail and deeply analyzed which include their concepts and limitations, in order to seek the research gap and scope. Therefore, the next paragraph shows the list of unique reasons that justifies the selected studies (refer to the studies listed in tables 4.2 and 4.3), with the objective to figure out their generic components which used as components in proposed DVDeM design model.

i. These studies clarify the implementation and design of BI.
ii. These studies highlight the details of design approaches that are attractive to BI environment.
iii. These studies detail the structural of BI design and data integration theories to support BI approach.
iv. These studies recommend specific guidelines to be considered in designing BI applications, which is good for this study.
v. These studies cover the whole BI design model steps.
vi. These studies take into consideration the Gathering requirements, which is an important and essential stage in the design and development of BI model.

Universiti Utara Malaysia

vii. These studies take into consideration the concern for providing data in real-time or near real-time.

As shown in the list above a number of unique reasons why all selected studies are significant to be considered in this study. Some of the studies are considered based on their data integration approaches, design guidelines, deliver real-time data, some with data integration theories and some with BI model's components. It is emphasized again that this comparative analysis was done in order to discover generic components and phases for design BI model. Accordingly, the selected studies that used for content analysis process includes (12) studies in physical data integration and (6) studies with virtual data integration as clearly shown in tables 4.2 and 4.3.

This shows that the suggestions of BI components are to improve the data integration of the existing approaches and also consider the suggestions from experts and actual users. Therefore, the next paragraph, a content analysis was conducted and the selected studies with the objective to figure out their generic components in specifying the proposed DVDeM components and phases.

4.2 Content Analysis of Existing BI models

In order to get the full imaging for components, phases, and data flow in the BI environment, an analysis of the contents of the existing models was conducted. This analysis relies upon the design and implementation the proposed DVDeM model. Therefore, Table 4.2 lists the core components and main contribution of existing models, approaches, or methods that serve as the main pillars for developing the proposed model. Table 4.3 lists the core components and main contribution of existing DV solutions that were used when developing the proposed model. In this section, the naming of the existing BI models based on the data integration technique will be conducted. However, the BIDW refers to existing BI models that use DW technique as a data integration (as mentioned in chapter 2 section 2.4) while the BIDV refers to BI models that use DV as a data integration (as mentioned in chapter 2 section 2.5).

Table 4.2
Business Intelligence Using DW (Components and Outcomes)

	Model's Name	**The Core Components**	**Outcomes / Contribution**
BIDW1	CRISP-DM (2001).	Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, And Development.	A general overview of the data mining project. This overview was carried out the lifecycle for data mining.
BIDW2	DWARF (2003)	Requirements Management Planning, Requirements Specification, Requirements. Validation and Requirements Management Control.	It is a technique based on a set of phases. Each phase follows the abstraction levels of the application in depth.
BIDW3	CoDMODS (2011).	Organization Level Requirements and ODS Design Level	The outcome of this model is to represent a requirement process and ODS design for developing BI system that is focused on ODS function, which supports operational and tactical information.
BIDW4	IIHS (2011).	Data Sources, Data Integration (DW Using ETL Tool), Reporting Service Tool	Develop business intelligence reporting application in the health sector to decision makers for the purpose of helping them making the best decisions.
BIDW5	Academic BI Model (2006).	Collect and analyze user requirements and build a data warehouse for BI.	Develop university BI using DW technique.
BIDW6	HRM (2009)	Source Data, Data Management. Data Gain, List Maintains, DW, Data Application, And End Users.	A data warehouse BI model based on university human resource management of performance evaluation.
BIDW7	BIMUDA (2015)	Data search process, data sources, data usage, data manipulation, data enrichment, data growth, and creation of new data.	BI model used to help in generating new data and information that is more comprehensive and collective to help business intelligence through advanced analysis.

Universiti Utara Malaysia

Table 4.2 Continued

BIDW8	Knowledge-Based Model for R-T BI (2015).	Data Level (data knowledge Creation), Analysis level (Data Knowledge transfer), and Application Level (data knowledge usage).	Develop BI model to deliver the real-time in business sector BI environment.
BIDW9	SOA for BI (2007).	Data sources layer. ETL layer, Physical layer, and analytic application layer.	Service Oriented Architecture BI Environment.
BIDW10	SaaS BI (2011).	Infrastructure layer, data service layer, business service layer, user interface service layer, and operational service layer.	Helping business users can tailor on-demand SaaS BI deployment based on their business processes and application requirements.
BIDW11	BI Analytics Method without DW (2010).	Reporting services, analysis services (data cube, KPIs, and data mining), and pre-generated reports (spreadsheets) or direct tie-in database.	Develop Business Intelligence Analytics Method without Conventional DW.
BIDW12	Five-Layered BI Architecture	Data source layer, ETL layer, data warehouse layer, end user layer, and metadata layer.	A Conceptual framework of five-layered BI architecture with various components.

Table 4.3
Business Intelligence Using DV (Components and Outcomes)

S. CODE	DV Solution Name	The Core Components	Outcomes / Contribution
BIDV1	DV Solution for Compassion International.	Data Sourcing Services/ And Systems, Data Transform View, Canonical Layer, And Consumer Views.	An enterprise-wide, single version of the truth in virtual manner.
BIDV2	DV Solution for Fortune 50 Computer Manufacturer	Regional procurement application system, Cisco DV layer, and global procurement reporting.	DV solution to deliver a global view of inventory data to procurement department users.
BIDV3	DV Solution for Global 50 Energy Company.	Upstream Data Sources, Data Virtualization Layer (Source Connection, Conforming Layer, Common Semantic Layer, Business Demand Layer, And Data Storage). And BI and Analytic Systems.	Deliver a virtual data Warehouse and virtual data marts in order to give fully support BI systems.

Universiti Utara Malaysia

Table 4.3 Continued

BIDV4	DV Solution for Global 100 Financial Services Firm.	Data Management Service, Data Stores, and Data Access Service.	Data Vault (DV). The DV is an operation data store (ODS) in order to provide access the bank data at real-time
BIDV5	DV Solution for NYSE Euronext.	Transactional Systems And References Data Sources, Data Virtualization Layer, And Business Intelligence Tools and Business Intelligence Applications.	Create a virtual view of (DW) to deliver unified read-only to post-trade data.
BIDV6	DV Solution for Pfizer Inc.	Data Sources, Information Abstraction, Integrated Data Virtualization, Reporting Obstruction, and Web Portal.	Federated data delivery framework implemented with the Cisco Data Virtualization.

It can be concluded from Table 4.2 and Table 4.3 that BI environment consists of three core phases (layers) namely, requirements gathering, development, and presentation, phases. Each phase is broken down into components as well as each component is broken down into several functions and processes. However, in order to analyze the modes components, the components categories were indexed. Accordingly, Table 4.4 tabulates a classification of the categories and condition of components as adopted by many studies such as Ariffin (2009), and Din et al. (2012).

Table 4.4
Components Categories Index

Index	Desecration	Condition
A	All models apply	**100 %** models apply
M	Majority of models apply	**More than 50 %** models apply
F	Few models apply	**Less than 50 %** models apply
X	Not applied in any model	**0 %** models apply

In Table 4.5 indexing categories on each of the components in the models reported in Tables 4.2 and 4.3 are tabulated. The purpose of this indexing is to find out whether each component is mandatory or optional. Table 4.5 and Table 4.6 illustrate the components categories indexing for BIDW, BIDV models respectively, while Table 4.7 represents the mapping between BIDV and BIDW.

Table 4.5

The BIDW Components Indexing

Model Phases and components	**BIDW 1**	**BIDW 2**	**BIDW 3**	**BIDW 4**	**BIDW 5**	**BIDW 6**	**BIDW 7**	**BIDW 8**	**BIDW 9**	**BIDW 10**	**BIDW 11**	**BIDW 12**	**Percentage %**	**Index**
Requirement Gathering Phase		√	√		√		√			√			41.60	**F**
Organization and Business Requirements	√	√	√		√		√			√			50.00	**M**
Data Sources Requirements		√		√		√	√	√	√	√	√	√	75.00	**M**
Infrastructure Requirements		√	√			√				√			33.30	**F**
Model development phase	√		√	√	√	√	√	√	√	√	√	√	91.60	**M**
Data connection			√	√	√	√	√	√	√	√	√	√	83.30	**M**
Data preparation			√	√	√	√	√	√	√	√	√	√	83.30	**M**
Data integration	√			√	√	√		√	√	√	√	√	66.60	**M**
Data transformation	√		√	√	√	√	√	√	√	√	√	√	91.60	**M**
Data Quality	√		√	√	√	√	√	√	√	√	√	√	91.60	**M**
Data Profiling				√	√	√		√	√	√	√	√	66.60	**M**
Validation and Control				√			√			√			25.00	**F**
Model presentation phase.	√		√	√	√	√	√	√	√	√	√	√	91.60	**M**
Reporting Service			√	√	√				√	√	√	√	58.30	**M**
Analytic Service	√		√		√		√	√		√	√	√	66.60	**M**
Data Mining Tool	√												08.33	**F**
End users	√		√	√	√	√	√		√	√		√	75.00	**M**
Validation and Control	√				√					√			25.00	**F**

Note: √ means contained in the approach

Universiti Utara Malaysia

Table 4.6
The BIDV Components Indexing

Model Phases and components	BIDV1	BIDV2	BIDV3	BIDV4	BIDV5	BIDV6	Percentage %	Index
Requirement Gathering Phase	√				√		33.33	**F**
Organization and Business Requirements				√	√		33.33	**F**
Data Sources Requirements	√	√	√	√		√	83.33	**M**
Infrastructure Requirements	√	√	√			√	66.66	**M**
Model development phase	√	√	√	√	√	√	100.0	**A**
Data connection	√	√	√	√	√	√	100.0	**A**
Data preparation	√	√	√	√	√	√	100.0	**A**
Data integration	√	√		√	√	√	83.33	**M**
Data transformation	√	√	√	√	√	√	100.0	**A**
Data Quality	√	√	√	√	√	√	100.0	**A**
Data Profiling	√		√	√	√	√	83.33	**M**
Model presentation phase.	√	√	√	√	√	√	100.0	**A**
Reporting Service		√	√	√	√		66.66	**M**
Analytic Service	√	√	√	√	√		83.33	**M**
Data Mining Tool	√		√	√	√		66.66	**M**
End Users	√	√	√	√	√	√	100.0	**A**

Table 4.7
Mapping Between BIDV and BIDW

Universiti Utara Malaysia

Models Phases and Components	BIDV Indexing	BIDW Indexing
Requirement Gathering Phase	**F**	**F**
Organization and Business Requirements	**F**	**M**
Data Sources Requirements	**M**	**M**
Infrastructure Requirements	**M**	**F**
Model Development phase	**A**	**M**
Data connection	**A**	**M**
Data preparation	**A**	**M**
Data integration	**M**	**M**
Data transformation	**A**	**M**
Data Quality	**A**	**M**
Data Profiling	**M**	**M**
Model Presentation phase.	**A**	**M**
Reporting Service	**M**	**M**
Analytic Service	**M**	**M**
Data Mining Tool	**M**	**F**
End users	**A**	**M**

Based on this content analysis, the selection of the phases and components for DVDeM adheres to the conditions listed in Table 4.8.

Table 4.8
The Conditions for Selection DVDeM Components

	Conditions	BIDW	BIDV	Compulsory	Recommended	DVDeM
1	Condition 1	A	Any	1	-	Compulsory
2	Condition 2	Any	A	1	-	Compulsory
3	Condition 3	M	M	1	-	Compulsory
4	Condition 4	M	F	1	-	Compulsory
5	Condition 5	F	M	1	-	Compulsory
6	Condition 6	X	Any	-	1	Recommended

Based on the conditions stated in Table 4.8, the DVDeM early components are listed in Table 4.9. In addition to some other components that have contributed to building DVDeM design model.

UUM
Universiti Utara Malaysia

Table 4.9
The DVDeM Early Components

Models Phases and Components	BIDV Indexing	BIDW Indexing	DVDeM Indexing
Requirement Gathering Phase	**F**	**F**	Compulsory
Organization and Business Requirements	**F**	**M**	Compulsory
Data Sources Requirements	**M**	**M**	Compulsory
Infrastructure Requirements	**M**	**F**	Compulsory
Model development phase	**A**	**M**	Compulsory
Data connection	**A**	**M**	Compulsory
Data preparation	**A**	**M**	Compulsory
Data integration	**M**	**M**	Compulsory
Data transformation	**A**	**M**	Compulsory
Data Quality	**A**	**M**	Compulsory
Data Profiling	**M**	**M**	Compulsory
Model presentation phase.	**A**	**M**	Compulsory
Reporting Service	**M**	**M**	Compulsory
Analytic Service	**M**	**M**	Compulsory
Data Mining Tool	**M**	**F**	Compulsory
End users	**A**	**M**	Compulsory

In line with above situations, the content analysis for existing BI model whether these models as used DW or DV, which given a significant for identifying the DVDeM model phases and components as well as the functions between these components. In consequence, this activity has identified the key elements of DVDeM model such as phases, components, activities, and flows for the development of DVDeM model. The detail of DVDeM development process is described in the next section.

4.3 DVDeM Development Process

This section describes the development process of the proposed model. After conducting a comparative analysis and review of related literature, the phases and components of the proposed model are formulated. All these components and phases are combined to build the first version of the proposed model. Therefore, the DVDeM proposed model consist of three core phases:

i. Data Virtualization Requirements Gathering.
ii. Data Virtualization Development.
iii. Data virtualization Presentation.

Each phase of the DVDeM model was mentioned above are contain a set of sub-phases or components. The detail of each phase is described in the next section.

4.3.1 Data Virtualization Requirements Gathering

A requirement is a statement regarding what should be done or how a task should be performed for a product or intentional software. In requirements engineering, requirements gathering is an important and necessary component of any project or project management endeavor. The complete understanding of a project output is

Universiti Utara Malaysia

critical to the success of the project in question, and requirements gathering can be considered a cornerstone to the success of any project. Gathering and analyzing the DVDeM requirements have three main components namely, organization and business requirements, infrastructure requirements, data sources requirements, and document requirement specification. In the next paragraphs, each of the components will be highlighted in detail.

4.3.1.1 Organization and Business Requirements

Before starting design and development of virtual data model, developers in this area should possess sufficient knowledge to business problem and the general background about the organization, such as the number of members and the head of the organization's capital and the type of the reports that the organization intends to present. The developers should also be aware of the goals and vision of the organization with system architectural considerations for the organization. In addition to reviewing of the current system of the organization that it intends to replace, this review includes why the current system needs to be replaced, what are the pros and cons, and what are the strategic reasons for the replacement. Therefore, the three sub-phases of this phase include organization sponsorship and involvement, business requirements, and enterprise architecture. Figure 4.1 illustrates the organization and business requirements components.

Universiti Utara Malaysia

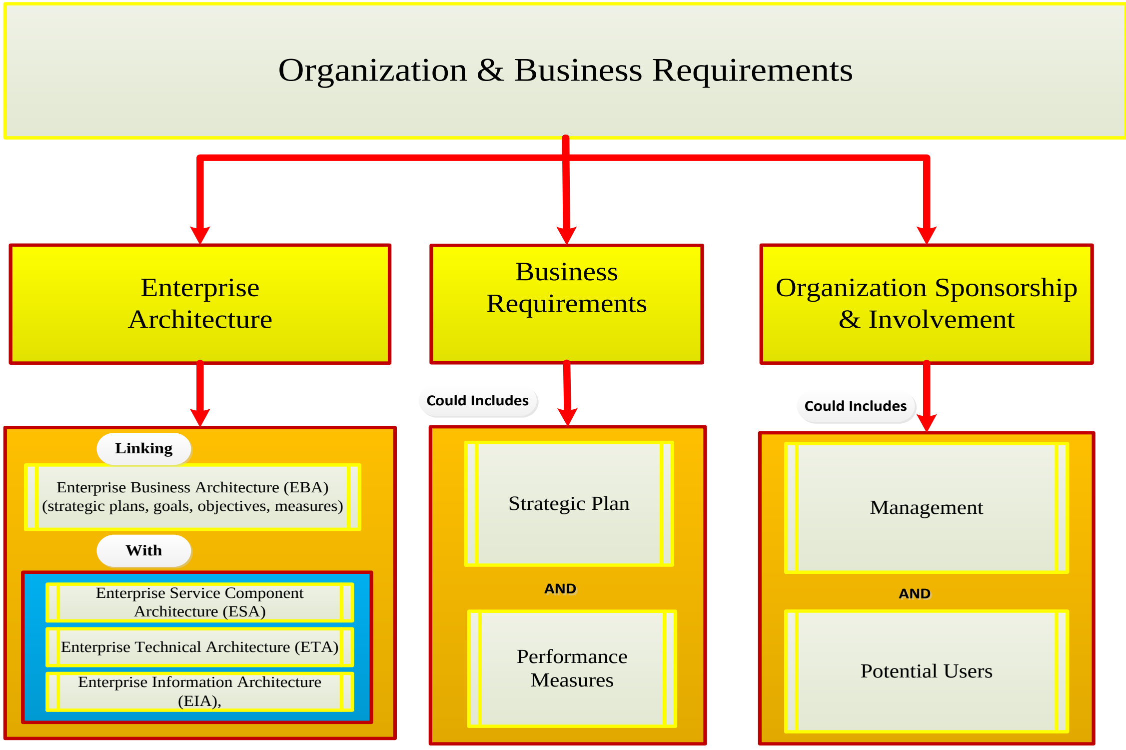


Figure 4.1. Organization and Business Requirements Components

4.3.1.1.1 Organization Sponsorship and Involvement

Organization sponsorship includes ensuring sufficient resources are available. It also means the consistent commitment to implementing data virtualization application that is the single source for corporate measurement and decision support data. While sharing means that all potential stakeholders and users of the data virtualization, even executives, from every organizational unit and level, should be actively involved in data virtualization design, development, and management. DV users will have the most influence on acceptance of the virtualization data as sources for their BI applications; therefore, it is imperative that their needs are addressed. They are also the "owners" and "stewards" of operational data and thus are the best source for subject matter expertise.

4.3.1.1.2 Business Requirements

In business requirements, developing DV applications in BI without first determining strategic business requirements is a sure recipe for failure. The best source for these requirements is the enterprise strategic plan and the performance measures identified in the plan.

i. Strategic Plan: a strategic plan outlines an enterprise's mission and purpose, goals, strategies and performance measures (business requirements). Properly used, a strategic plan is the tool with which effective managers guide their organizations and ensure corporate success.

ii. Performance Measures: establishing the right performance measures is the key to successful enterprise management. An enterprise must be able to tell whether progress is being made on its critical goals and whether stakeholder expectations are being met.

Universiti Utara Malaysia

4.3.1.1.3 Enterprise Architecture

Enterprise architecture is a process of linking between enterprise business architecture (EBA) which includes (strategic plans, goals, objectives, and measures) with its enterprise information architecture (EIA), enterprise service component architecture (ESA) and enterprise technical architecture (Dwivedi et al., 2009). This architecture is a logical organization of corporate information requirements, descriptions of application systems that support the enterprise's strategic requirements. It includes the relationships between application systems via shared software components and shared data elements. The enterprise information architecture also establishes guidelines, standards, and operational services that define the enterprise's computing technology environment. Before an enterprise can define, design, and implement the architecture for its strategic information management systems, including virtual tables, virtual data marts, decision support systems, and executive information systems, it must first document the environment in which these systems will be implemented.

After the comprehensive understanding of the organization and business requirements, the developers must jump to the investigation of all matters relating to enterprise data sources, which intends to use the DV technique to integrate. This is issue will be discussed in the following section.

4.3.1.2 Data Sources Requirements

In the world of DV, it is necessary to identify and understand the sources of data before the integration process begins. As is well known to researchers in the field of databases, there are several types of data sources and they are classified according to the way and complexity of storing such data. For the management and integration of this vast amount of data using the DV technique, developers should do some preparations that precede the integration process. These preparations process include:

Universiti Utara Malaysia

i. Identify data source.
ii. Understanding the Data source.
iii. Save results for further reuse.
iv. Knowing the organization's ability to deciding which type of data integration technique will choose.

The data sources requirement phase starts with an initial data collection and proceeds with activities in order to get familiarize with the data. Then, the phase continues with identifying data quality problems, and discovering first insights into the data.

Finally, the phases conclude with detecting interesting subsets to form hypotheses for hidden information.

The following questions need to be considered when determining the sources and costs of data for the DV:

i. Where does the data come from?
ii. What processes are used to obtain the data?
iii. What does it cost to obtain the data?
iv. What does it cost to store the data?
v. What does it cost to maintain the data?

Answering these questions requires us to classify data into three categories depending on the complexity of the process of extraction and integration of these data. In this study, a new classification of the data sources is proposed based on the complexity and the effort required of the data extraction process. This classification depends on the effort to extract the source data, which requires data to be divided into three categories based on the complexity of the data extraction. Data sources requirement is illustrated in the Figure 4.2.

Universiti Utara Malaysia

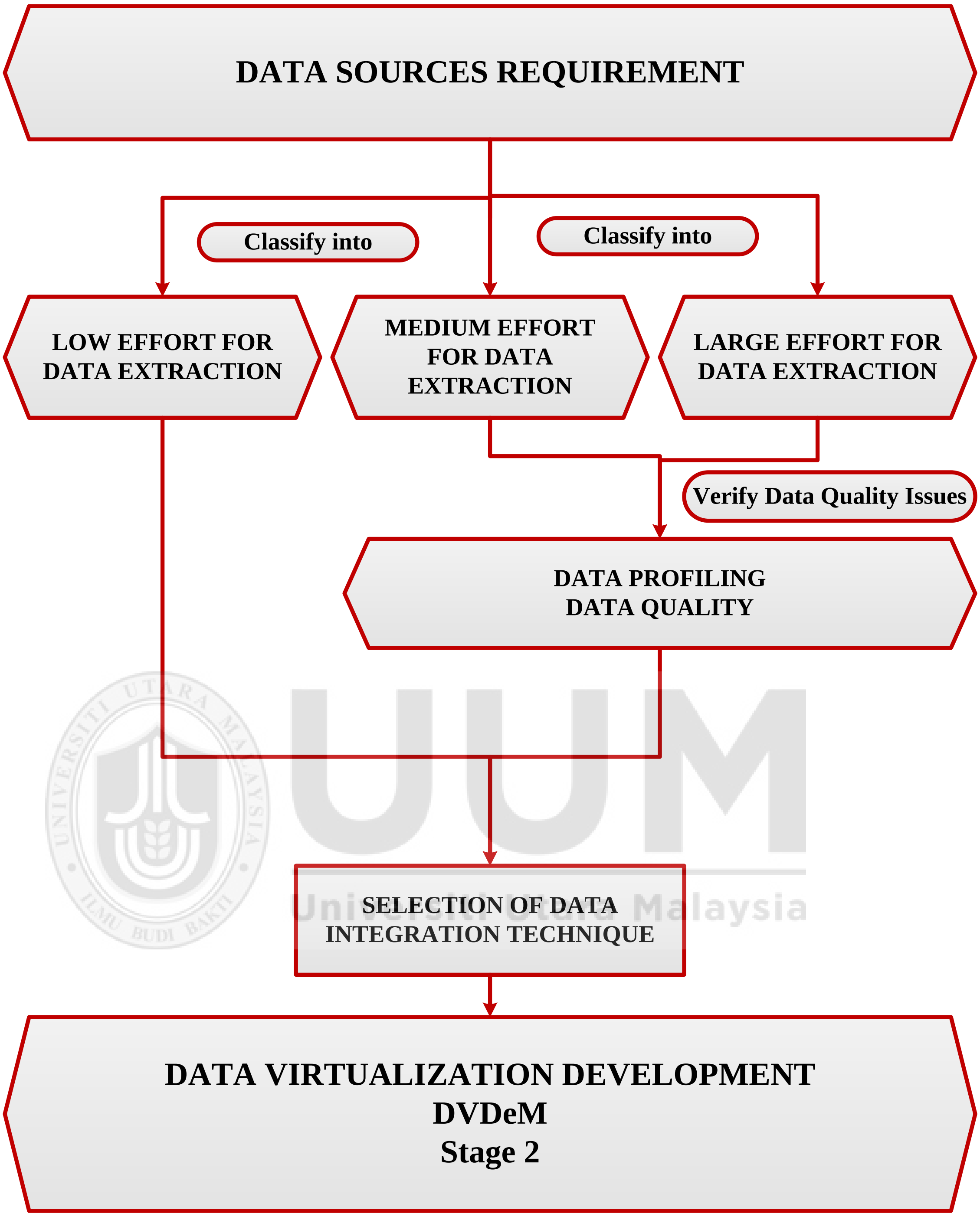


Figure 4.2. Data Sources Requirement

As discussed in previous paragraph and shown in Figure 4.2, it can be seen how data sources can be classified into three categories: the first category is called low effort data extraction, the second category is called medium effort data extraction, and the third category is called large effort data extraction. As DV has deal directly only with the data flattening such as SQL tables, therefore, developers should pass the data that are not in flattened form (large and medium effort) to the data converter as well as to pay attention to the data quality issues in order to transform these data to the standardize form before being used by DV layer.

On the same aspect, building a BI system requires effort, time and high cost. In the event of failure or a shortcoming of the use of any part in the construction of these BI systems that will lead to cost, time consuming and effort which it is not easy demolition and reconstruction BI systems. Therefore, in context of this study, in order to prevent and / or eliminate of these difficulties, a steps that used to help organizations and BI developers to quickly determine and selection a suitable data integration techniques was used (Mousa, Shiratuddin, & Bakar, 2014a).

Universiti Utara Malaysia

4.3.1.3 Infrastructure Requirements

Infrastructure Requirements in DV environments consist of two main sections; the software requirements and the hardware requirements. Infrastructure is an extremely important component of a DV environment as it provides the underlying foundation that enables the DV architecture to be implemented. It is sometimes called technical architecture and it includes several elements such as hardware platforms and components (i.e.: disks, memory, CPUs, DV/BI applications servers), operating systems (e.g.: UNIX), database platforms (e.g.: relational engines or multidimensional/OLAP engines), connectivity and networking. Several factors influence the implemented infrastructure namely, business requirements, the technical architecture, systems issues, DV team competency, policy, other organizational issues, and expected growth rates. Because the DV environment manages, addresses, and delivers an enormous amount of data in a virtual manner, developers have to pay attention to the appropriate infrastructure, both from the standpoint of hardware or software. In other words, the developers should provide at least a minimum infrastructure that allows the design and implementation of data integration system in virtual manner. A comprehensive description of the intended purpose and environment, as well as the whole model requirements for DVDeM model development should document requirements specification. The detail of how document requirements specification is described in the next section.

4.3.1.4 Documentation of Requirement Specification

The last stage in the requirement gathering phase is document requirement specification. In this phase, the requirements specification should be documented. The process of documentation should detail the whole elements and components of the applications that can be considered as fact relationships, measures, dimensions, and hierarchies. This will be contained in the technical metadata. Moreover, it is most important to involve at this phase domain expert who can help in determining business terminology for all components and in indicating whether measures are additive, semi-additive, or non-adoptive. The DV development of the DVDeM proposed model is therefore based on the outputs of requirement gathering and analysis.

4.3.2 Data Virtualization Development

Based on content analysis for existing BI models, as well as reviews of related literature, and requirement gathering of DVDeM, the components of the DVDeM model were finally formulated. The detail of DVDeM development process is described in the next sections as visualize in Figure 4.3 shows the DV development

Universiti Utara Malaysia

phase. The DV development stage is considered the core stage in DVDeM proposed model. The input for this stage is DV gathering requirements while the output are virtual tables and virtual data marts which will be input for stage 3 for DVDeM proposed model (DV presentation).

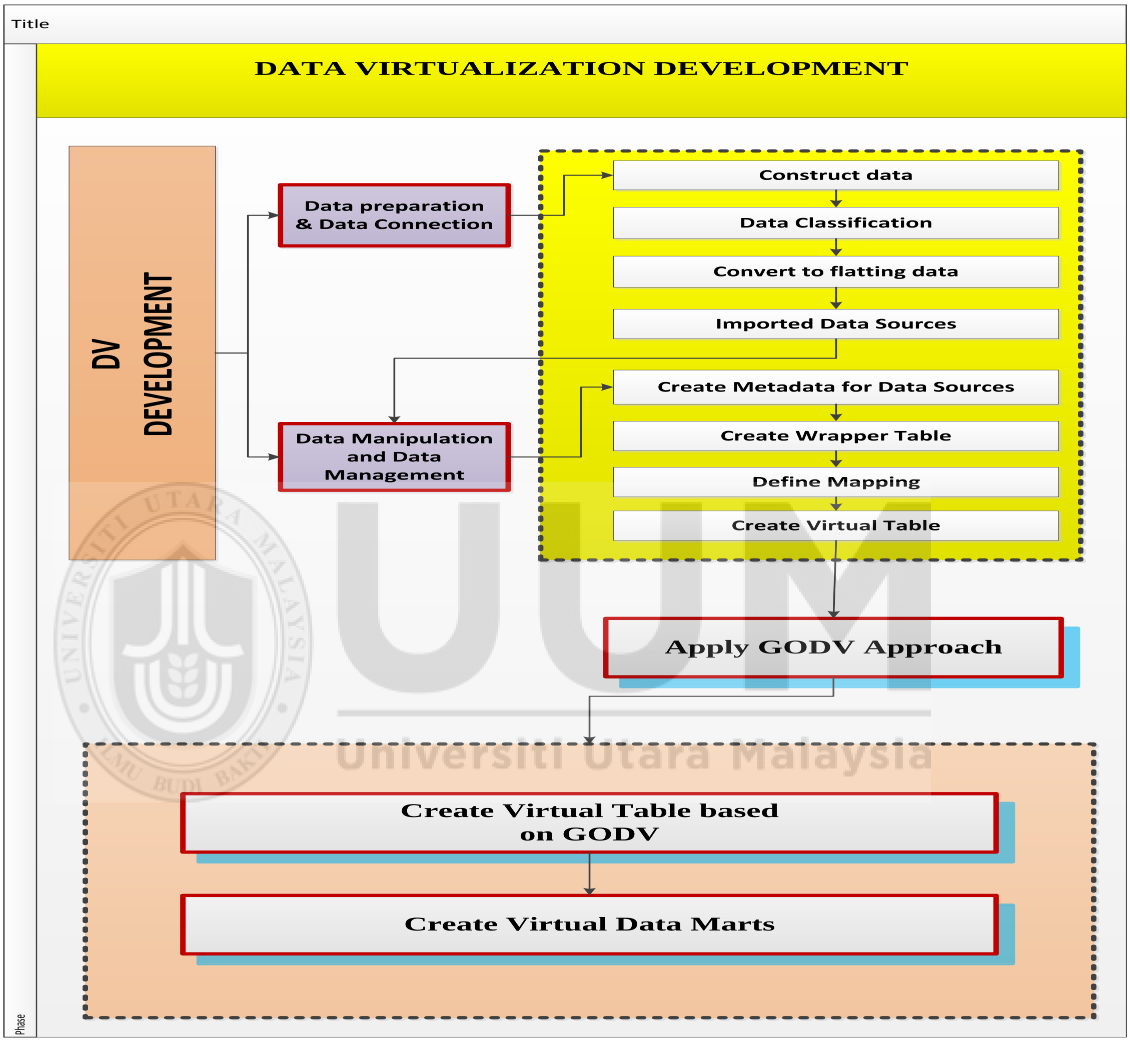


Figure 4.3. The Data Virtualization Development Phase

Data virtualization development phase consists of two major components as shown in Figure 4.3. The first one is data preparation and data connection, while the other one is data manipulation and data management. Each of these components consist of systematic steps should follow. The detail of each of them is described in the next section.

4.3.2.1 Data Preparation and Data Connection

Data preparation and data connection consist of two sub-phases namely, data understanding which includes collection of initial data, exploration and description of data, and verifying data quality while data preparation involves data selection, data profiling, and data cleaning. In the next section, the details about each of them are introduced. The output for this phase is imported data sources.

In the data gathering stage, data was collected and formatted for integration. Therefore, the first step in data preparation and data connection is to select data to be used for analysis. The selection criteria include relevance to the data, organization goals, quality and technical constraints such as limits on data volume and data types. After that data can be constructed through data preparation operations such as the production of derived attributes, creation of entire new records or transformation of values for existing attributes.

Universiti Utara Malaysia

As mentioned in Chapter Two, data sources hold various types of data format, such as SQL, XML etc. In data integration technologies, both in the physical manner or virtual manner, data integration preferably is dealt with data that is stored in flat form like SQL tables. Therefore, a classification of data stored based on the degree of complexity and effort required to extract is proposed. Based on this classification, the data that DVDeM intend to deal with should be classified into three types, namely, low effort data extraction, medium effort data extraction, and large effort data extraction. If the data type is low effort data extraction, it means that this data in flat format and there is no need to pass these data into data converter while the other two types need to go through a data converter. Finally, the data will undergo the selected of data quality method to be cleaned, cleansed and prepared before it is ready for integration.

After applying data quality and data profiling, the final step in data preparation and data connection is importing data sources. Data import means making them familiar to the server that is used to virtualize it. In the other words, importing means introducing data sources to DV table. The import process is simple in the event that the original data is stored in a database management system such as SQL. However, importing the data sources should be done data have undergone data quality and data profiling. Immediately after the selection of a data source to be imported and before the import process commences, determining if the data (values) require conversion into standard formats as mentioned earlier is essential. It can be concluded that before the data sources tables act as a source to the virtual table it has to be imported. However, during the selection of the source table for import, it is imperative to ensure that all data have been transformed in a standardized form. The justification for that is sometimes, different database management systems store values of particular data types in different manner. To be able to compare values managed by different database servers correctly, the values should be transformed to a standard form. Figure 4.4 illustrates the importing process.

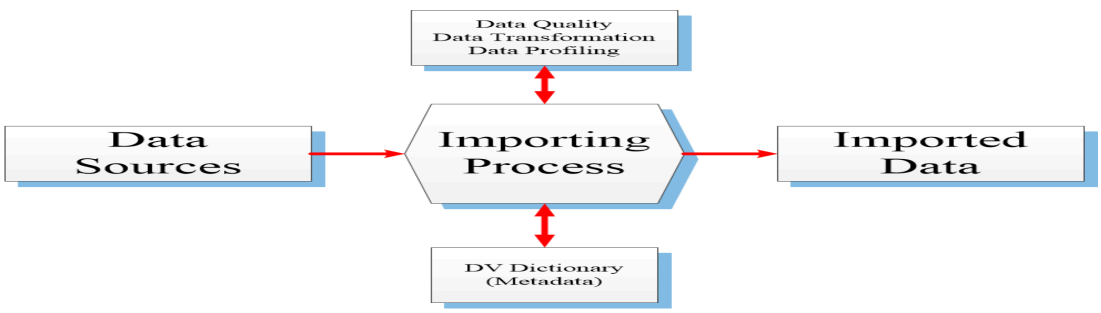


Figure 4.4. The Importing Process

It is important that during the importing process, the developer extract all the metadata relevant to the data sources (tables) and stores it in its own dictionary. The metadata might include the source table location, database connection information (log on), the database's name, owner, and date created and the data sources structure. It is also mandatory to check the primary and foreign keys on the data sources. Data manipulation and data management commence as soon as data source import completes.

4.3.2.2 Data Manipulation and Data Management

In this phase, the explanation about how to create wrapper and define the virtual table in addition of in order to enhance query performance we create virtual data marts based on analysis the organization and business and finally, we deliver the output of this phase (virtual table and/or virtual data marts) to the third main phase (data virtualization presentation).

Universiti Utara Malaysia

Upon completion of importing the source data, the next step is to a create wrapper table. The process of creating a wrapper table relies on Meta data extracted during data source importing process. These Meta data include all information related to the source table such the location of source table, the name, and owner of the source data. This information from the Meta data will then be used to create wrapper table. The relationship between the wrapper table and the imported table is a many to one relation. A wrapper table can be considered as a mirror of a source table because it has the same structure and same components as a source table. It can be concluded that the wrapper table is 100 percent similar to the imported source table with two exceptions. First, the values of particular data types are transformed to a standard data type. Secondly, the ability to directly access the data via wrappers allows the developers to determine whether these data need to be transformed or not. Figure 4.5 shows the relationship between imported table and wrapper table.

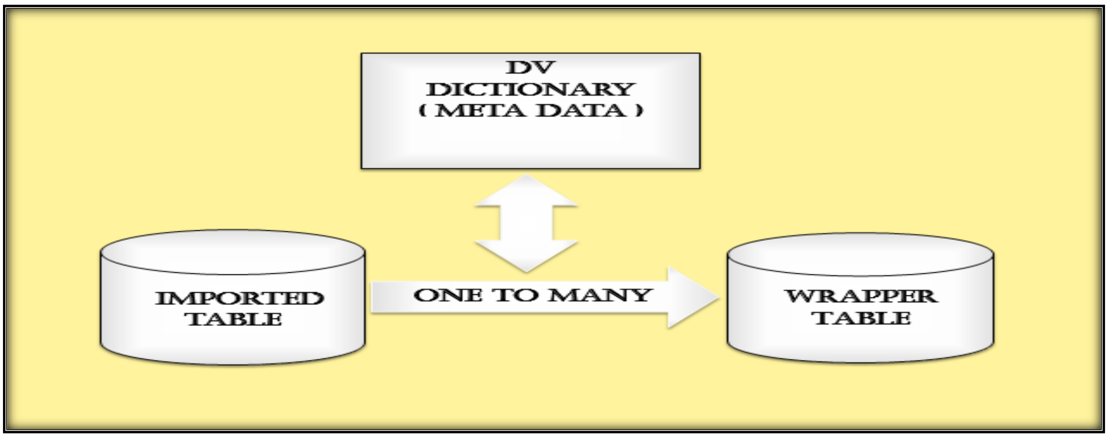


Figure 4.5. Relationship between Imported Table and Wrapper Table

As highlighted previously, the wrapper table contents and structure look like the source table contents and structure. The next step is to define and create a virtual table. The main reason stands behind creating a virtual table is to fulfil the requirements of data consumers. Some data consumers want to see the contents, while some of them prefer to see specific rows and/or specific columns. Another group of data consumers on the other hand may want to see some data in an aggregate manner, while some of them only want to see the data after applying some mathematical operations. Defining a virtual table means defining a mapping. The mapping defines the virtual table structure and how the data have transformed from data sources to the virtual table. The mapping process is usually composed of many processes such as row and column selection, column concatenations and transformations, column and table name changes, and groupings as shown in Figure 4.6.

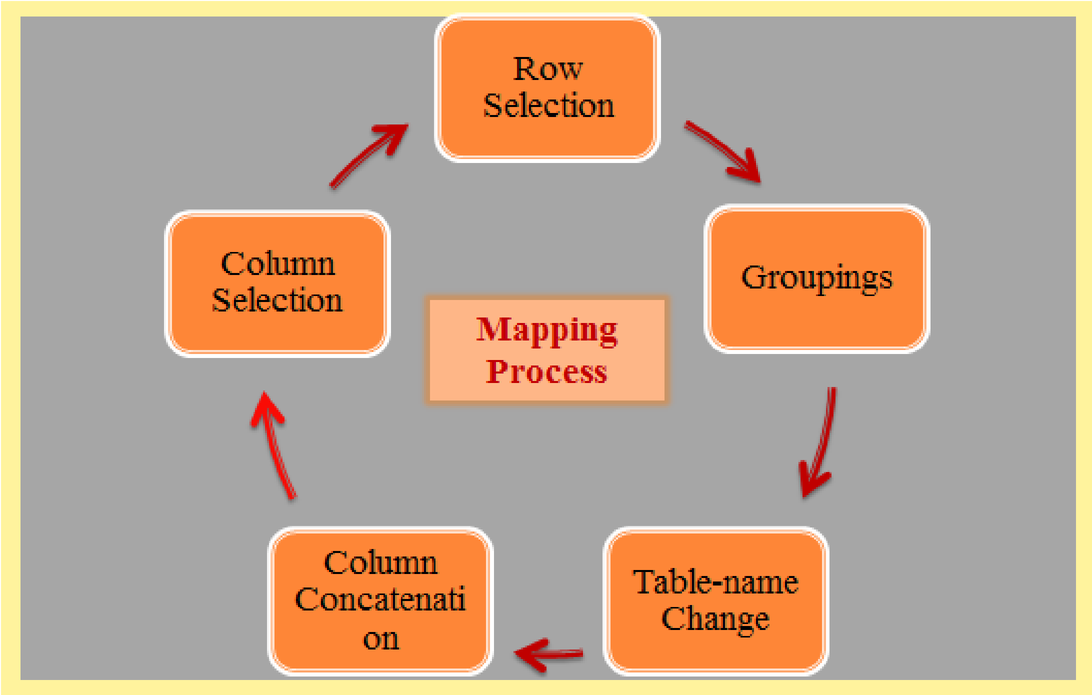


Figure 4.6. Mapping Process Tasks

It can be concluded that without a mapping process, the virtual table cannot be updated or queried. Figure 4.7 shows the relationships between DV components.

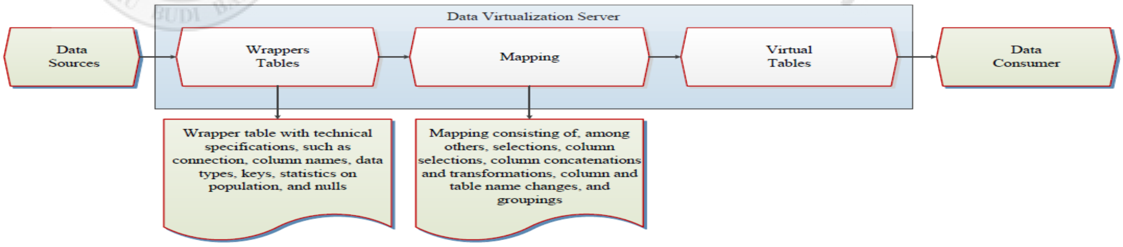


Figure 4.7. The Relationships among Data Virtualization Components

As mentioned earlier, the main outcome from the DVDeM model is to create virtual tables. Hence, the virtual table consists of the huge amount of data, in order to make a virtual table contain only the relevant data to minimize the access time the data as well as reduce the data retrieval time, therefore, the Goal based approach (GODV) for the virtual table has been proposed. The next section details on the GODV proposed approach.

4.3.2.3 Goal based Approach

The main issue behind the proposed approach is typically the vast amount of data involved in a DV application, virtual table, wrappers and imported data sources. BI technical expert may face the difficulties in identifying the necessary information that should be used in BI environment. In order to obtain a virtual table with relevant data, a goal oriented approach based on data virtualization was adopted. GODV is based on the mappings of organization goals into system goals from which requirements are defined. GODV consists of three core steps:

i. Build Goals Structure.
ii. Match these goals with virtual tables.
iii. Deliver virtual table to BI end users

4.3.2.3.1 Step 1: Build Goals Structure

The input for this step is the organization mission. This mission can be broken down into one or more goals, and these goals are further broken down to one or more sub-goals. This process will be performed iteratively until it reaches the leaf goals. The leaf goals will then be broken down into indicators. Essentially, the (And) and (Or) operators can be used for goals at the same level. Additionally, this approach uses the concept goal decomposition tree (GDT) that was proposed by (Simon et al. 2006) in agent modeling and programming. GDT is a mechanism that captures both the declarative and procedural aspects of goals, which offers the ability to reason about goals. GDT provides traceability among the goals using logical inferences through a set of decomposition operators allowing claims on a goal's achievement based on its sub-goals' achievement. On the other hand, the relationship among the components whether between mission and goals, goal and sub-goals, or sub-goals and leaf goal is one to many. Figure 4.8 illustrates the build goal structure.

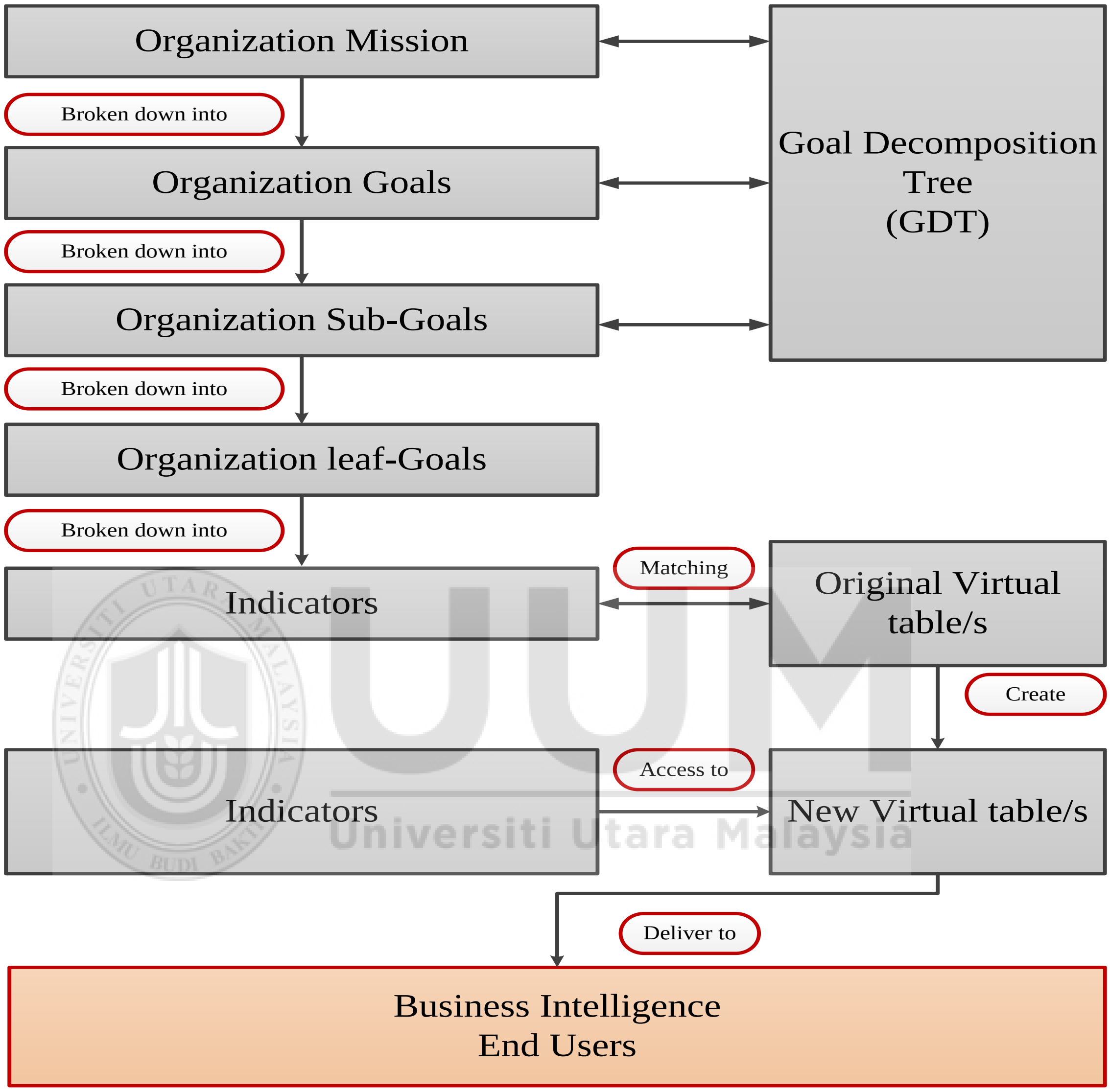


Figure 4.8. Build Goal Structure

4.3.2.3.2 Step 2: Match the Indicators with Virtual Tables

In this step, the indicators that were achieved in step 1 are integrated with a virtual table to later facilitate the acquisition of goal-related information. However, relating records with indicators is based on altering the existing virtual table by adding a new column. The column data type is string and the values for this column is either “related” or “not related” based on the matching process that defined relevant data to the KPIs for intended goal. Therefore, the outcome of this step is the virtual table based on GODV. It can be concluded that the data access and extraction from the virtual table based on GODV is faster than using original virtual table. Figure 4.9 illustrates the proposed GODV approach.

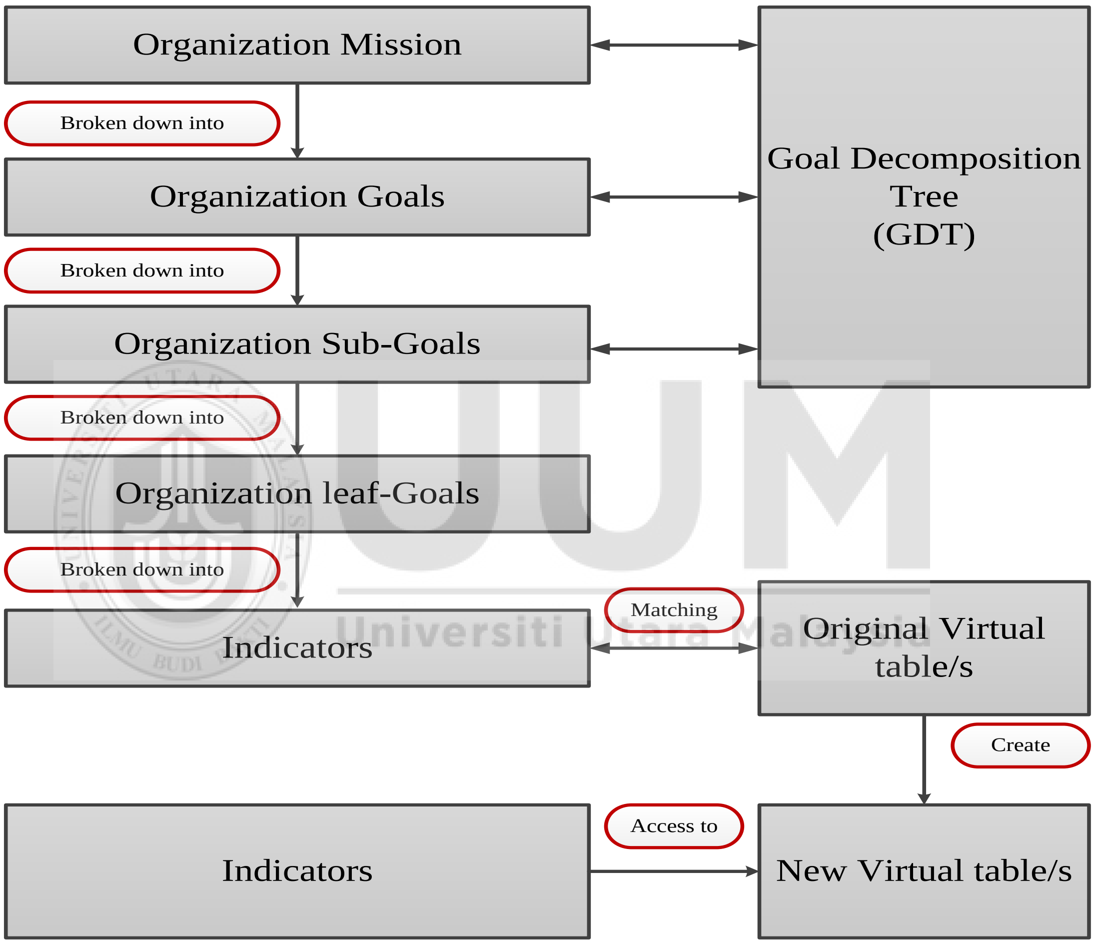


Figure 4.9. The GODV Proposed Approach

4.3.2.3.3 Step3: Deliver Virtual Table to BI End Users

Once the virtual table is created, virtual tables are used by BI end users. In this step, the virtual table may break down into a set of virtual data marts and is usually oriented to a specific business line or team. Now the virtual table or its virtual data marts is ready to use by BI end users. In the context of this study, the virtual data marts can be created either by applying GODV approach or without it. The details about virtual data marts creation is elaborated in the next paragraph.

In summary, the main aim of applying GODV approach is to make the virtual table consist of relevant data only and minimize the virtual table size. However, in some cases if the virtual table already contains the relevant data and it size is not huge; the BI developer can create BI reports without applying GODV.

4.3.2.4 Create Virtual Data Marts

As discussed previously, in order to reach more query performance and the need to reach more agility that includes reducing the access time and search time moreover, to reduce the maintenance (less work and less cost). For example: if T= the time that is needed to perform a query of the virtual table, the time that is needed to query in half of such table is therefore T/2. Thus, DVDeM provides the possibility to create virtual data marts based on data consumers' needs and goals. For example, assuming that the data consumer is data mining tool and the organization's business goal is to "Increase catalog sales to existing customers." The data mining goal is to "Predict how many widgets a customer will buy, given their purchases over the past three years, demographic information (age, salary, and city) and the price of the item". It can be concluded that virtual data mart which serve the data mining, should contain all information related. The data source (input) for these virtual data marts will be virtual table.

It is noted that, there is no difference between the virtual table and the view concept in DBMS (such as SQL view), particularly in terms of the name, the structure, the query definition, and the contents. After the creation of the virtual table, DV presentation is performed next. The following paragraphs explain DV presentation in detail.

4.3.2.5 Near Real Time and Virtual Table

In the context of this study, as mentioned previously the main aim of this study is to provide near real-time data in order to support decision-making process. Therefore, using DVDeM proposed model to delivering virtual tables that mean the data definitely be live data and on demand.

As spelled out in the previous sections, a data that delivering by applying DVDeM proposed model can provide additional data for a complete historic context that complements point-in-time, near real-time data. Even so, if real-time data is useful without contextual data from a DW or other data integration techniques, by applying DVDeM proposed model, the virtual table's service may send transformed and cleansed data directly to a BI platform or other target. Besides, with DVDeM proposed model due to adopt data virtualization as a data processing; it's possible to get current data by going around data integration, these data delivered with fully repurposed, clean, compliant, complete, contextual, or auditable as well as near real-time data.

To support their decision-making, businesses frequently need to integrate data from different source applications in real time or near real time. Traditional technologies can do this in theory, but as spelled out in the previous sections, their long development cycle is often a source of frustration. In addition, the process of collecting data is not without an overhead: by the time incremental changes have been fetched, data quality rules applied, and the results loaded into the traditional data integration techniques, the data may no longer be current. Another option is to build a data store for operational reporting, but even this often cannot give users genuinely real-time results because of data quality and transformation challenges. Therefore, DVDeM proposed model enables near real-time data access for BI end users. The data continues to reside in the source database, so does not become out of date during the access process. Data from other business units, too, can often be accessed in its raw form without getting into a formal process of request and clarification, which means it, too, is available in real time. As a result, every part of the business can now access the up-to-date status of the business as a whole, instantly. With the DVDeM proposed model, it becomes possible to generate quick reports that, while not necessarily complex, provide powerfully support for near real decision-making.

DVDeM proposed model provides a solution for a single version of the truth with respect to multiple databases. DVDeM is a highly efficient way to make data from operational systems available for near real-time needs, and it can also pull in older data from BI databases. That means BI tools can access all the company's business information, both current and historical, in a unified way. In addition, with DVDeM proposed model it becomes possible for multiple BI tools to access the data in the same way, using the same business terminology and the same KPI aggregation rules. The organization can ensure everyone uses the same data.

Besides, in DVDeM by applying GODV approach in order gain the virtual tables with relevant data, as well as near real-time data, by minimizing the time need to access and search for data that used by BI end users to support near real-time decision-making process.

4.3.3 Data Virtualization Presentation

In DV presentation, when the virtual data marts or virtual tables have been created, these data need to be published or exposed. This means that the virtual tables or the data marts become available for data consumers through one or more languages and programming interfaces. Our model provides a web-based application that can used by any data consumer (analytic service, reporting service, data mining tools, and end users) to access to the information they want. Figure 4.10: illustrates the DV presentation.

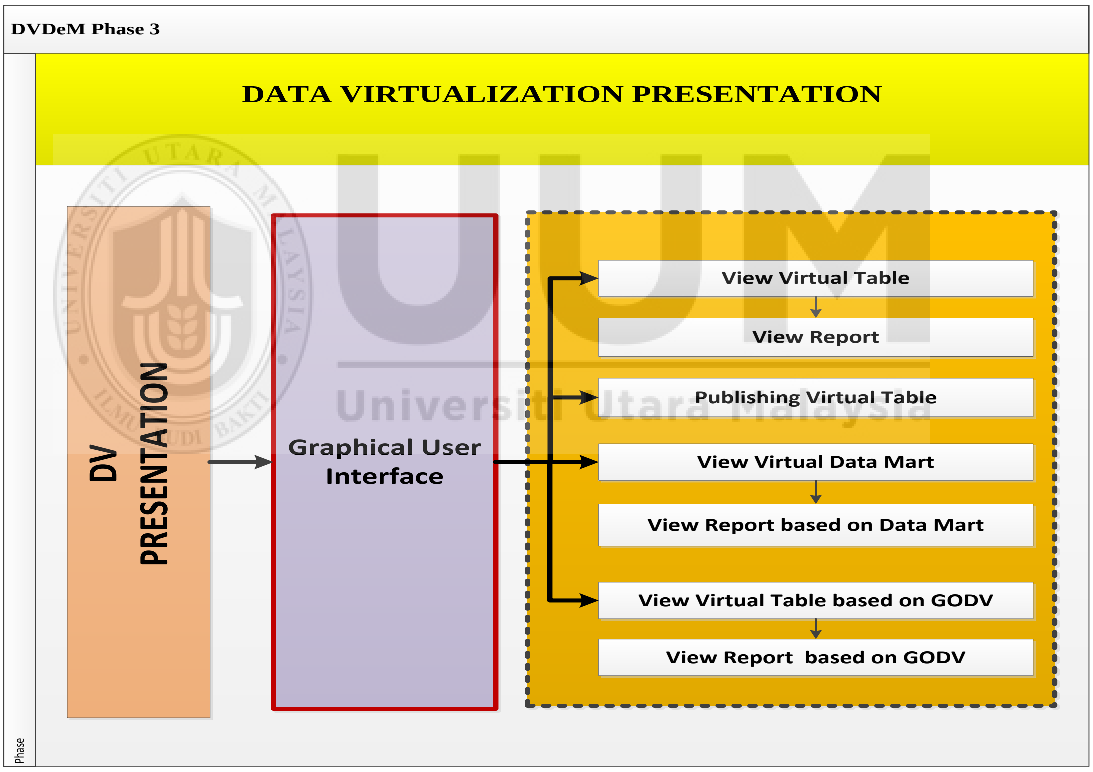
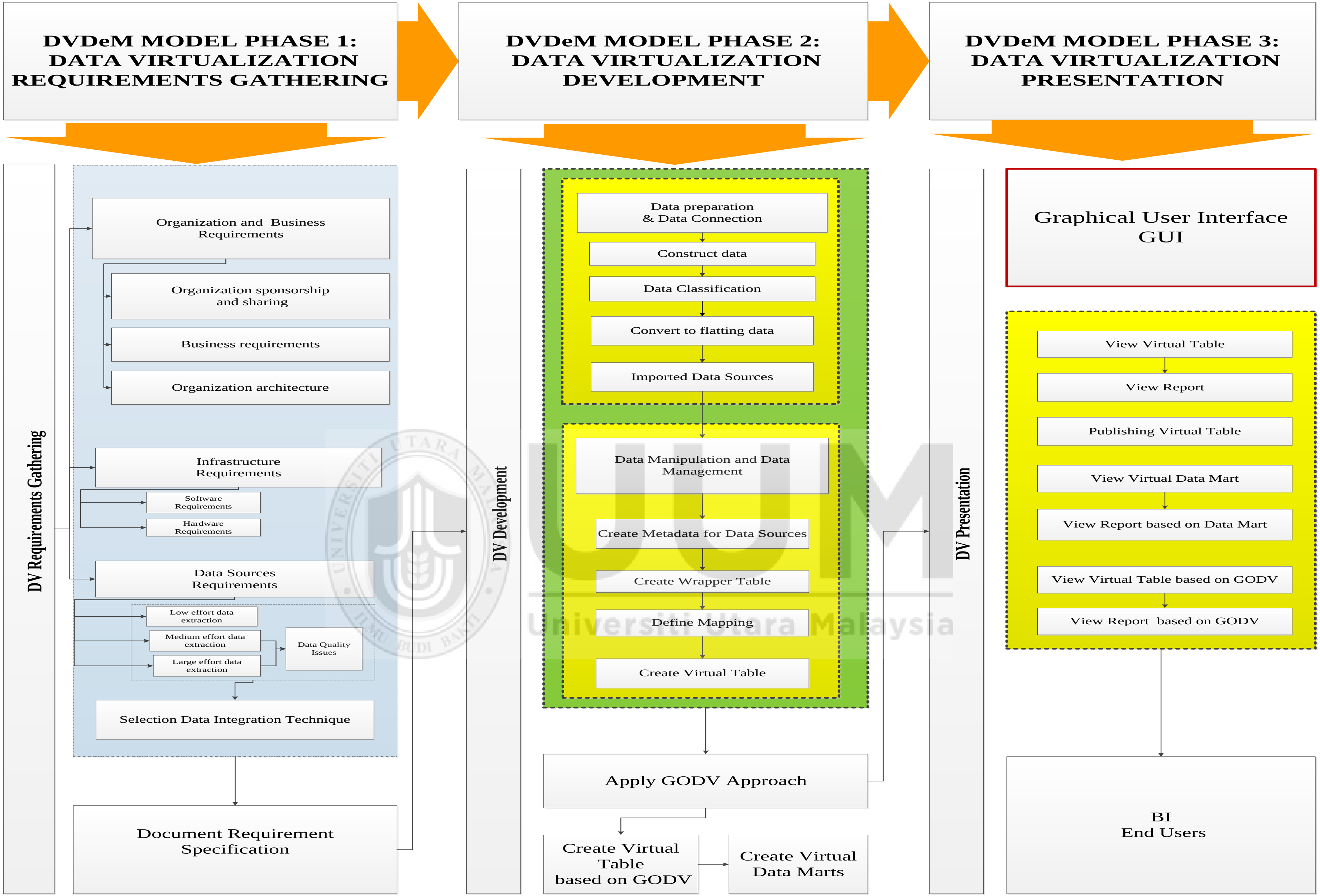


Figure 4.10. Data Virtualization Presentation

4.4 Develop DVDeM (First Version)

In this section, the initial version of the proposed model is discussed. The model was developed based on the combination of all the components that were mentioned in sections 4.3.1 to 4.3.3. Figure 4.11 shows the first version of DVDeM.



Universiti Utara Malaysia

Figure 4.11. DVDeM (First Version)

4.5 Validation and Revision of the Proposed Model

The focus groups in expert review are two of the accepted techniques used to gather qualitative data by means of group interaction on a matter determined by the researcher (Morgan, 1996). Hence, this study uses the expert review technique and focus group discussion to evaluate and verify the proposed DVDeM model, the expert reviews were conducted. This validation is an extensive and iterative process, and is discussed in the next section. Figure 4.12 illustrates the DVDeM validation process.

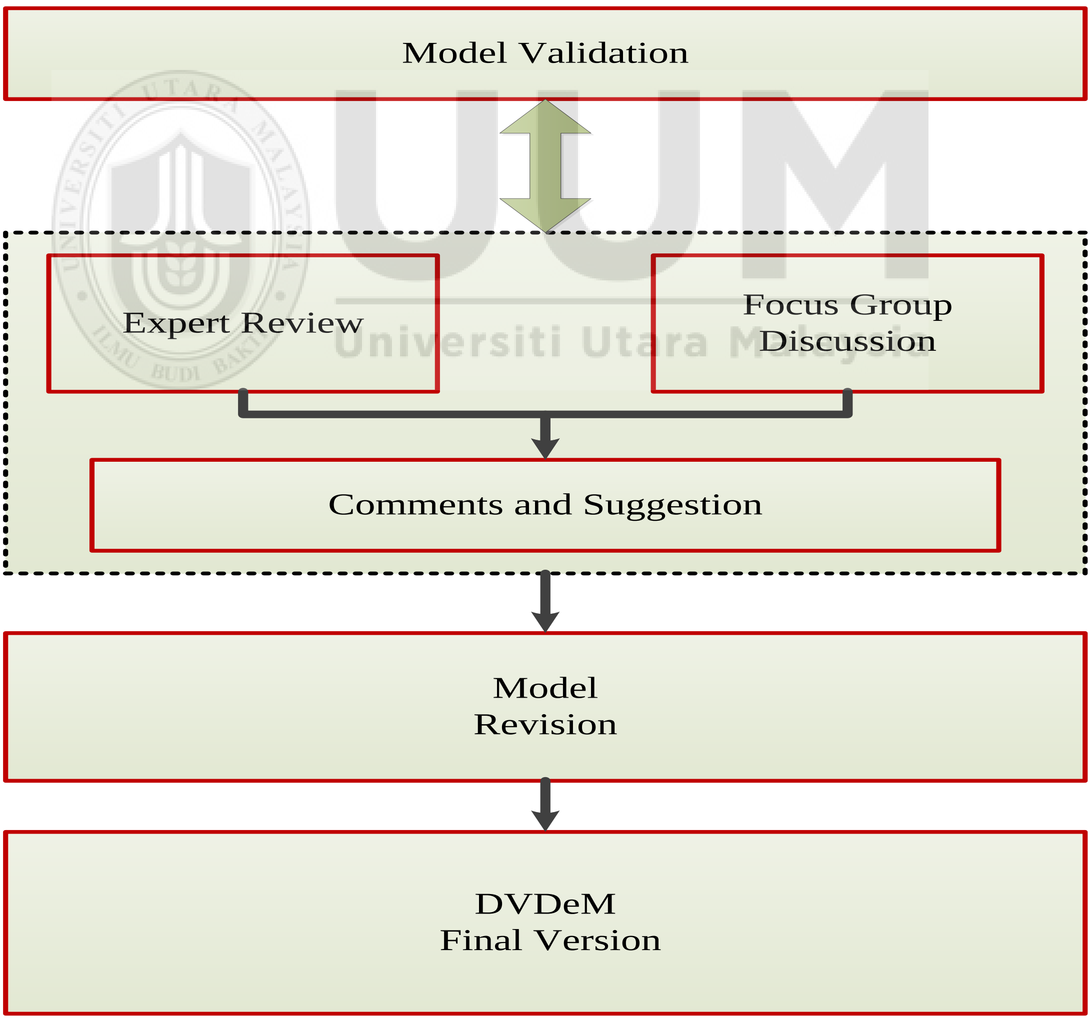


Figure 4.12. DVDeM Model Validation Process

4.5.1 Expert Review

According to Morgan (1996), the expert review is one of the accepted techniques used to gather qualitative data by means of group interaction on a matter determined by the researcher. Hence, this study uses the expert review technique validate and verify the proposed DVDeM model, the expert reviews were conducted. As mentioned in chapter 3, twelve experts are adequate to review the proposed DVDeM model. Therefore, 12 experts come from BI, data integration, and database solution fields were selected. The experts' profiles are listed in Table 4.10. It is worth noting that there are two groups of expert review conducted. Experts involved in this review were lectures and instructors who are represented as the end users that developing and teaching in the area of BI as well as industry experts, and they were classified based on the following criteria:

i. Have PhD or Master in Information Systems (IS) or Computer Science or Business Intelligence, or related areas,
ii. Have at least five years teaching background in IS or CS or BI or related areas, and
iii. Have experiences in designing data integration systems whether the design was in (DV or DV) for at least five years.

The demographic of experts that participated in the model review session are presented in Table 4.10. as clearly shown in Table 4.10, two types of experts are involved; experts from the academic sector and industry sector; in order to obtain the opinions, suggestions, feedback or comments from experts from both sectors.

Universiti Utara Malaysia

Table 4.10
Demographic Profiles of Experts

No	Age	Education	Field of Expertise	Position	Experience (Year)	Sector	Location
1	47	PhD	DI and BI	University lecturer	23	Academic Sector	Malaysia
2	36	PhD	DA/DB	University lecturer	11	Academic Sector	Nigeria
3	34	PhD	CS / DI	BI Developer	10	Industry Sector	Austria
4	54	PhD	CS / Multimedia Communication	University lecturer	30	Academic Sector	Malaysia
5	41	PhD	ICT	BI Developer	16	Industry Sector	Holland
6	39	PhD	CS / Data Retriever	University lecturer	10	Academic Sector	Iraq
7	38	PhD	CS/ and Database	University lecturer	10	Academic Sector	Iraq
8	40	PhD	Computer Science and IT	University lecturer	12	Academic Sector	Iraq
9	39	PhD	CS and IT	University lecturer	17	Academic Sector	Iraq
10	39	PhD	CS and IT	University lecturer	15	Academic Sector	Iraq
11	43	PhD	Information Governance and Architecture	BI Developer	20	Industry Sector	USA
12	44	PhD	CS/IT	University lecturer	14	Academic Sector	Iraq

Universiti Utara Malaysia

4.5.1.1 Procedures and Review Instrument

As clearly indicated in Table4.10, the experts represent different field of expertise from different academic institutions including local and international. This is important to establish reviews and comments. As for the educational background, four of them are professors and two are associate professors. The remaining are PhD holders. This number met with requirement for experts in this study. Their experiences in the respective fields vary from 8 to 35 years. The age of experts varies from 32 to 57 which show the level of maturity in giving opinions and assessments. Meanwhile, the review process took three months from 1th of June until 2th of September 2015.

The objective of the expert review was to conduct a review of the proposed phases and components of DVDeM model, in order to seek the expert view on each item. The main instrument used for this review is a questionnaire which adapted from Siti Mahfuzah (2011), this questionnaire is contains six questions asking about: (i) relevancy of proposed model phases, (ii) relevancy of proposed the model components, (iii) understandability of the terminology used in the proposed model. Besides, four additional questions have asked to the experts regarding the connections and flows of the components are logical, how the phases and their components usable are to guide BI designers in BI environment, and readability and understandability of the proposed items for developers. Along with that, few demographic questions were also asked like age, gender, education and field of expertise. Experts were also encouraged to write their further comments in the provided instrument. All of the experts answered the review questions, and gave feedback for the proposed model in written.

Universiti Utara Malaysia

Furthermore, most of the experts' feedback has been applied and, one of the experts recommends including guidelines to this model to help individual developers to develop their BI systems using data virtualization technology. For the theories, the experts state that most of the theories are relevant and the connection between the theories and components are logical. The proposed model is therefore feasible to the development of BI systems.

4.5.1.2 Findings

Data gathered from the expert review are tabulated in Table 4.11. The data were documented as in frequency of responses of the expert review to the questions asked in the instrument.

Table 4.11

The Frequency Responses of Expert Review

	Items	All are relevant	Some are definitely not relevant	Some may not relevant	Total
Q1	**Relevancy of DVDeM proposed model phases**				
1.1	Model main phases	11	0	1	12
1.2	Linking between Phases	10	0	2	12
1.3	DV Requirements Gathering	12	0	0	12
1.4	DV Model Development	12	0	0	12
1.5	DV Model Presentation	12	0	0	12
1.6	Content Structure	10	0	2	12
1.7	theory	12	0	0	12
Q2	**Relevancy of DVDeM proposed model phases and their components**				
Phase 1. Data Virtualization Requirements Gathering					
2.1	Organization and Business Requirements	10	0	2	12
2.2	Data Sources requirements	11	0	1	12
2.3	Infrastructure requirements	10	0	2	12
2.4	Linking between components	12	0	0	12
Phase 2. Data Virtualization Development					
2.5	Data Preparation and connection	12	0	0	12
2.6	Data manipulation and management	12	0	0	12
2.7	GODV approach	11	1	0	12
2.8	Linking between components	11	0	1	12
Phase 3. Data Virtualization Presentation					
2.9	View data sources	0	5	7	12
2.10	View virtual table	12	0	0	12
2.11	View reports	12	0	0	12
2.12	View virtual data mart	12	0	0	12
2.13	View report based on virtual data marts	11	0	1	12
2.14	Publishing virtual table	2	4	6	12
2.15	Linking between components	11	0	1	12

Universiti Utara Malaysia

Table 4.11 Continued

	Items	Need very details explanation	Needs some explanation	It is easy to understand	Total
Q3	**Understandability of the terminology used in DVDeM proposed model**				
	Clarity of terminology				
3.1	Model main phases	0	3	9	12
3.2	Linking between Phases	0	2	10	12
3.3	DV Requirements Gathering	0	1	11	12
3.4	DV Model Development	0	3	9	12
3.5	DV Model Presentation	0	2	10	12
3.6	Content Structure	0	3	9	12
3.7	Theory	0	2	10	12
Phase 1. Data Virtualization Requirements Gathering					
3.8	Organization and Business Requirements	0	2	10	12
3.9	Data Sources Requirements	1	3	8	12
3,10	Infrastructure Requirements	0	2	10	12
3.11	Requirements Specifications	1	2	9	12
3.12	Linking between components	0	2	10	12
Phase 2. Data Virtualization Development					
3.13	Data Preparation & Connection	1	1	10	12
3.14	Data Manipulation & Management	0	2	10	12
3.15	GODV approach	0	2	10	12
3.16	Linking between components	1	1	10	12
Phase 3. Data Virtualization Presentation					
3.17	View data sources	0	1	11	12
3.18	View Virtual Tables	0	1	11	12
3.19	View Virtual data mart	0	1	11	12
3.20	View report based on virtual table	0	0	12	12
3.21	View report based on virtual data mart	0	0	12	12
3.22	Publishing virtual table	1	3	8	12
3.23	View virtual table based on GODV.	2	3	7	12
3.24	View virtual data marts based on GODV.	1	1	10	12
3.25	Create and view report based on GODV (virtual table)	1	1	10	12
3.26	Create and view report based on GODV (virtual data marts).	2	0	10	12
3.27	Linking between components	0	1	11	12

.

Universiti Utara Malaysia

As clearly indicated in Table 4.11, the majority of the experts agreed that the proposed model contain relevant phases and components, exhibit logical flows, usable to the development of BI environment and the phases and their components are readable. Likewise, the majority of the experts agreed that the proposed model contain relevant phases and components, exhibit logical flows, usable to the development of BI environment and the phases and their components are readable. However, for each examined terminology, at least 17% of the experts think that connections and flows of all phases and used in the proposed model are not clear, therefore, it needs to have better inline. Lastly, they were expected to give overall comments based on their understanding and perception of the clarity, phases, connections, and flow of the proposed DVDeM model as illustrate in Table 4.12.

Table 4.12
Additional Feedback

4 = Strongly Agree, 3 = Agree, 2 = Disagree, and 1 = Strongly Disagree						
	Questions	**4**	**3**	**2**	**1**	**Total**
4	I found that terminologies used are easy to understand.	3	9	0	0	12
5	I found the connections and flows of all components are logically appropriate	10	2	0	0	12
6	I found that proposed phases and components will guide designers to develop of data virtualization system in Business Intelligence.	3	9	0	0	12
7	Generally, I found that the proposed appropriate design data virtualization model is readable and understanding.	10	1	1	0	12

Besides responses in frequency, experts also provide written comments. Majority of the comments are addressed to model components. It is suggested that, since the presentation layer is where the system is exposed to actual users, this is where any logical errors will emerge, or at least the point where users will voice critique that

Universiti Utara Malaysia

should lead to improvement of the system. The current model does not seem to take that into account.

Experts also suggest some elements to be added in content structure components and changed terminologies to give clearer meaning. Besides, the chosen theories and link between the theories need to be revised. The experts noted that the term found in the first part of the proposed model (data virtualization gathering requirement) should be "data integration" rather than "data extraction", since the real problem with loosely typed or schema less data sources is not "extraction" per se but the whole process of cleaning and conforming the data as well as making the ETL resilient to the expected volatility of the data that might be encountered in the source. This review was taken into consideration in revising the proposed model. In addition, the reviews from expert were taken into consideration in revising the proposed model. Table 4.13 lists the comments from all of the experts. Some of the comments were rephrased from the original versions to express clearer meaning.

Table 4.13
Further Comments from the Expert Reviews

	Experts	Comments & Suggestions
1	Expert 1	Data selecting should be one of the data preparation components
2	Expert 2	In development phase, Goal driven approach should be explaining in some details.
3	Expert 3	The feedback component should be inside gathering requirement phase.
4	Expert 4	In gathering requirements phase, the input for GFCBDIT guideline should be documents requirement specification.
5	Expert 5	The end user feedback is already found implicit in organization requirements.
6	Expert 6	Data virtualization presentation phase should have a create report components.
7	Expert 7	The iteration is not clear between the model phases.
8	Expert 8	What's the difference between (infrastructure requirement and (system requirement)? Should choose the right term.
9	Expert 9	Data virtualization presentation should deliver its outputs to data consumers.
10	Expert 10	Remove the component feedback.

Table 4.13 Continued

11	Expert 11	The proposed model design should be user-friendly as much as possible.
12	Expert 12	Overall, the model phases and their components are easily understood but might some of them need to demonstrate in more details.

In line with above comments and suggestions, as well as to make the model clearer in terms of flow and the connection between components and phases, redraw the proposed model in a vertical manner instead of the horizontal manner. Hence, from the comments it can be concluded that majority of the experts inquired for design/development in an appropriate form. Besides that, two of them also gave advice on adding some features to the process of the model design.

4.5.2 Focus Group Discussion

Morgan (1996) describes focus groups discussion as one of the accepted techniques used to gather qualitative data by means of group interaction on a matter determined by the researcher. Hence, this study used focus group discussion to validate the proposed model. Focus group discussion was conducted. Six Selected participants participated in the focus group discussion. Six participants are sufficient for this study as supported by Folch-Lyon and Trost (1981), Kitzinger (1995), Morgan (1996), and Nielsen (1997). Besides, the focus group members came from different backgrounds such as computer science, BI developers, and data integration solutions, with minimum eight (8) years'. The intent for the numbers of the year's experiences is no less eight (8) years is to get more advantages from the participants based on their experience in the intended field as well as to show the level of maturity in giving opinions and assessments, as supported by the previous studies (Ariffin, 2009, Syamsul Bahrin, 2011; Siti Mahfuzah, 2011). The demographic of participants in

Universiti Utara Malaysia

focus group discussion that participated in the model review session are as presented in Table 4.14.

Table 4.14
The Demographic of Members in Focus Group Discussion

No	Gender	Age (Year)	Education	Field of Expertise	Position	Experience (Years)
1	Male	40	PhD	Computer Science / IT	BI developer	10
2	Female	30	PhD	ICT/ BI developer	BI developer	9
3	Male	39	PhD	Computer Science/ BI	BI developer	8
4	Male	35	MSc	Information System	BI developer	11
5	Male	32	MSc	IT/ DI	BI developer	10
6	Male	42	PhD	ICT	BI developer	10

4.5.2.1 Procedures and Review Instrument

The objective that stands behind conduct review in a focus group is to seek the expert view on each item. The main instrument used for this review is a questionnaire which contains four questions asking about: (i) how clear are the terminologies in terms of model main phases, linking between model phases, DV requirements gathering, DV model development, and DV model presentation; (ii) the understandability of the proposed model phases and their components for the BI developers; (ii) how usable are the phases and their components to guide BI designers in the development of BI system in BI environment; (v) the connection and flows of the phases and their components are logical.(refer appendix B). For the first question, verifies the understanding towards the need of more explanation is required on the proposed model and the participant's reviews were required to verify their understandability (i.e. easy to understand or need some explanation or need very detail explanation).

During the session, the researcher briefly explained the background of study and the objectives. Then, the proposed phases and their components were introduced. The review session involved two-way interactions, where experts may ask questions and give their opinions on the focused matter. Afterward, the experts were required to answer all the questions in the validation instrument. Findings of the review are discussed in the next section. Furthermore, experts are also allowed to ask questions related to the proposed model. The focus group discussion lasted for three hours. Along with that, few demographic questions were also asked like age, gender, education and field of expertise. Experts were also encouraged to write their further comments in the provided instrument.

4.5.2.2 Findings

Data gathered from the focus group discussion are explained in paragraph below. The data were documented as in frequency of responses of the participants to the questions asked in the instrument.

The majority of the experts agreed that the proposed model phases and components which contain relevant elements, exhibit logical flows, usable to the development Bi applications and the phases and components are readable. However, for the clarity of the terminologies 85% of the participants think that terminologies that used of the proposed model are easy to understand. However, for each examined terminology, at least 15% of the participants think that the terminologies used in the proposed model need some explanation, therefore, it need to have better inline.

For the phase 1 (DV requirements gathering) of the proposed model, 80% of the participants think the phase 1, its components, and the linking between them, are easy to understand. However, for DV requirements gathering, at least 13% of the participants think that the phase 1used in the proposed model need some explanation, and 7% think that the phase 1 need very details explanation, therefore, it need to have better inline.

For the phase 2 (DV Development) of the proposed model, 85% of the participants think the phase 2, its components, and the linking between them, are easy to understand. However, for DV requirements gathering, at least 8% of the participants think that the phase 1used in the proposed model need some explanation, and 7% think that the phase 2 of the proposed model need very details explanation, therefore, it need to have better inline.

For the phase 3 (DV Presentation) of the proposed model, 75% of the participants think the phase 3, its components, and the linking between them, are easy to understand. However, for DV requirements gathering, at least 15% of the participants think that the phase 1used in the proposed model need some explanation, and 10% think that the phase 3 need very details explanation, therefore, it need to have better inline.

Universiti Utara Malaysia

Besides, the participant 3 asks about if the DW technique would be replaced with DV, the answer was: the proposed model has the ability to support BI application in BI environment with or without DW. Overall, the focus group discussion was helpful for validating the proposed model and the results from discussion have been taken into consideration in revising the proposed model.

In addition, further comments from the participants were also documented during the reviews. Table 4.15 lists the comments from all of the focus group discussion participants. Some of the comments were rephrased from the original versions to express clearer meaning.

Table 4.15

Focus Group Discussion (Comments and Suggestions)

	Participants	Comments & Suggestions
1	Participant 1	✓ The model looks comprehensive but need to modify. ✓ The connection between model's components should be labelled. ✓ Put double arrows for each phase and its components to display the process is iterative or not.
2	Participant 2	✓ The feedback component should be inside gathering requirement phase.
3	Participant 3	✓ The connection between model's components should be improved. ✓ Overall, the components are easily understood but need to more explain.
4	Participant 4	✓ The term (data extraction) should replace by term (data integration).
5	Participant 5	✓ What about the tracing would be nice though to have those incorporate in the model. ✓ The iteration is not clear between the model phases.
6	Participant 6	✓ Some terms are not clear. The term (infrastructure requirement) should replace by term (system requirement). ✓ The connection between the components should be labeled.

From the comments depicted in Table 4.15, it can be concluded that majority of the experts inquired for design/development the proposed model. Besides that, one of the experts remarked on the consistent of some of the terminologies and two of them also gave advice on adding some features to the process in developing the proposed design items. accordingly, these comments and suggestions were taken into consideration to refines the first version of the proposed model and produce the final version of the proposed model.

4.6 The Formal Representation of DVDeM Model

In the context of this study, the proposed model is a representation of a BI system, made of the composition of concepts, components, the relationship between the components and processes which are used to help BI developers know, understand, simulate, or develop a BI system using data virtualization technique. The value of the proposed model is usually directly proportional to how well it corresponds to an existing state, potential, and future state of affairs in supporting decision-making process in BI environment. Therefore, the suitable formal representation of the proposed model is the Conceptual Graphs (CG) is a logical formalism that includes classes, relations, individuals, and quantifiers. This formalism is based on semantic networks, but it has a direct translation to the language of first-order predicate logic, from which it takes its semantics. The main feature is a standardized graphical representation that like in the case of semantic networks allows human to get a quick overview of what the graph means. A conceptual graph is a bipartite orientated graph where instances of concepts are displayed as a rectangle and conceptual relations are displayed as an ellipse. Oriented edges then link these vertices and denote the existence and orientation of relation. A relation can have more than one edges, in which case edges are numbered.

4.7 Refinement of DVDeM Design Model

As regards iteration in the validation process in model development, the majority of the experts (expert review and focus group discussion), agreed that the development process is clear and useful as a guide for BI designers to develop BI applications which embed DV technology components that are perceived could deliver real or near real time data in BI environment. However, there are some comments and suggestions from the experts to improve the sequence in the development process. They gave suggestions and valuable comments that were taken into consideration in the refinement the proposed DVDeM model. Table 4.16 lists the comments and suggestions were obtained from the experts and participants from both methods (expert review and focus group discussion). Besides, some of the comments were rephrased from the original versions to express clearer meaning, some of the experts' comments totally give the same meaning, and therefore, the combination of multi comments was done to produce one meaningful comment, with reference to their experts.

Table 4.16
Experts and Participants Comments and Suggestion

Comments and Suggestions	**Focus Group Participants**	**Experts**
Phase 1. Data Virtualization Requirements Gathering		
✓ The feedback component should build in Organization and business requirements.	__	1
✓ The organization and business requirements were redesigned in more detail to ensure mention the most sub-components that related such as (EBA, EIA, ETA, and ESA).	__	2,10, and 11
✓ The flow between phases and its components should be named.	1	__
✓ In data source requirements, data extraction should convey to data integration.	__	4 and 10
✓ The selection data integration should explain in detail.	2	__
✓ For a better format, the model's phases and components should be drawn vertically.	__	6
✓ In data sources requirements, the term "large" should be converted to “high".	3	__
✓ The infrastructure requirements should build in Organization and business requirements. ✓ The infrastructure requirements are under enterprise technical architecture (ETA).	__	11 and 12
Phase 2. Data Virtualization Development		
✓ The GODV approach should explain in detail as well as shows its components.	__	8,9, and 10
✓ The data input for GODV approach should be via virtual tables, virtual data marts, and imported data sources.	4 and 5	__
✓ Data preparation and data connection look ambiguous	__	12
✓ In general, each phase should be numbered.	6	1,7, and 8
✓ The flow between phases and its components should be named.	__	3,4, and 9
Phase 3. Data Virtualization Presentation		
✓ A data consumer is suitable for data virtualization model instead of end users.	5 and 6	__
✓ Data consumers (BI end user) types should explain in detail in this phase.	__	1,3, and 9
✓ View virtual table/s and publishing virtual tables are similar terms.	__	2,4, and 11
✓ The flow between phases should be named.	__	3,4,
✓ In the era of data virtualization, the term " BI end users" should replace with " Data Consumer "	__	2,4 and 5

Note: the numbers mean the expert number / participant number.

UUM Universiti Utara Malaysia

From the comments depicted in Table 4.16, it can be concluded that majority of the experts inquired for design and develop DVDeM. Besides that, some of the experts remarked on the consistent of some of the terminologies and two of them also gave advice on adding some features to the process in the DVDeM design model. Additionally, some of them affirmed for the importance of numbered the DVDeM model phases and their components as well as the flow among them should be labeled. Accordingly, these suggestions and comments were used iteratively to refine the proposed DVDeM design model as shown in Figures 4.13 to 4.18.

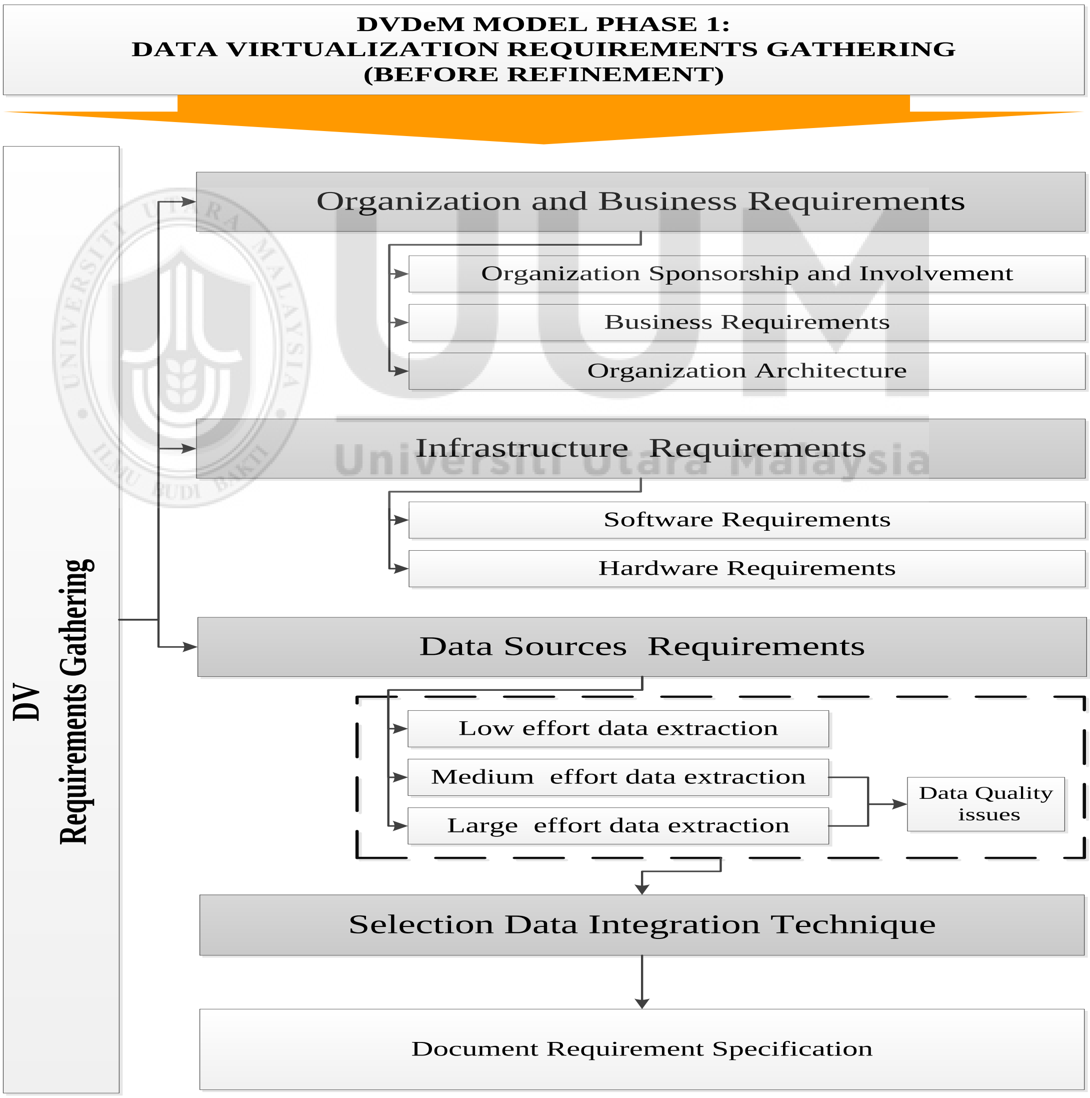


Figure 4.13. DV Requirements Gathering (Before Refinement)

For the first phase of DVDeM, the correction was done according to comments and suggestions regarding flows, connection, and others from experts. (see Figure 4.13 and 4.14).

i. End-user feedback, as well as infrastructure requirements, was built into business requirements component.
ii. The organization and business requirements were redesigned in more detail to ensure mention the most sub- components that related such as (EBA, EIA, ETA, and ESA).
iii. The flow between phases and its components have been named.
iv. The expressions of "data extraction" and " Large" were replaced with "data integration" and "High” respectively.

Universiti Utara Malaysia

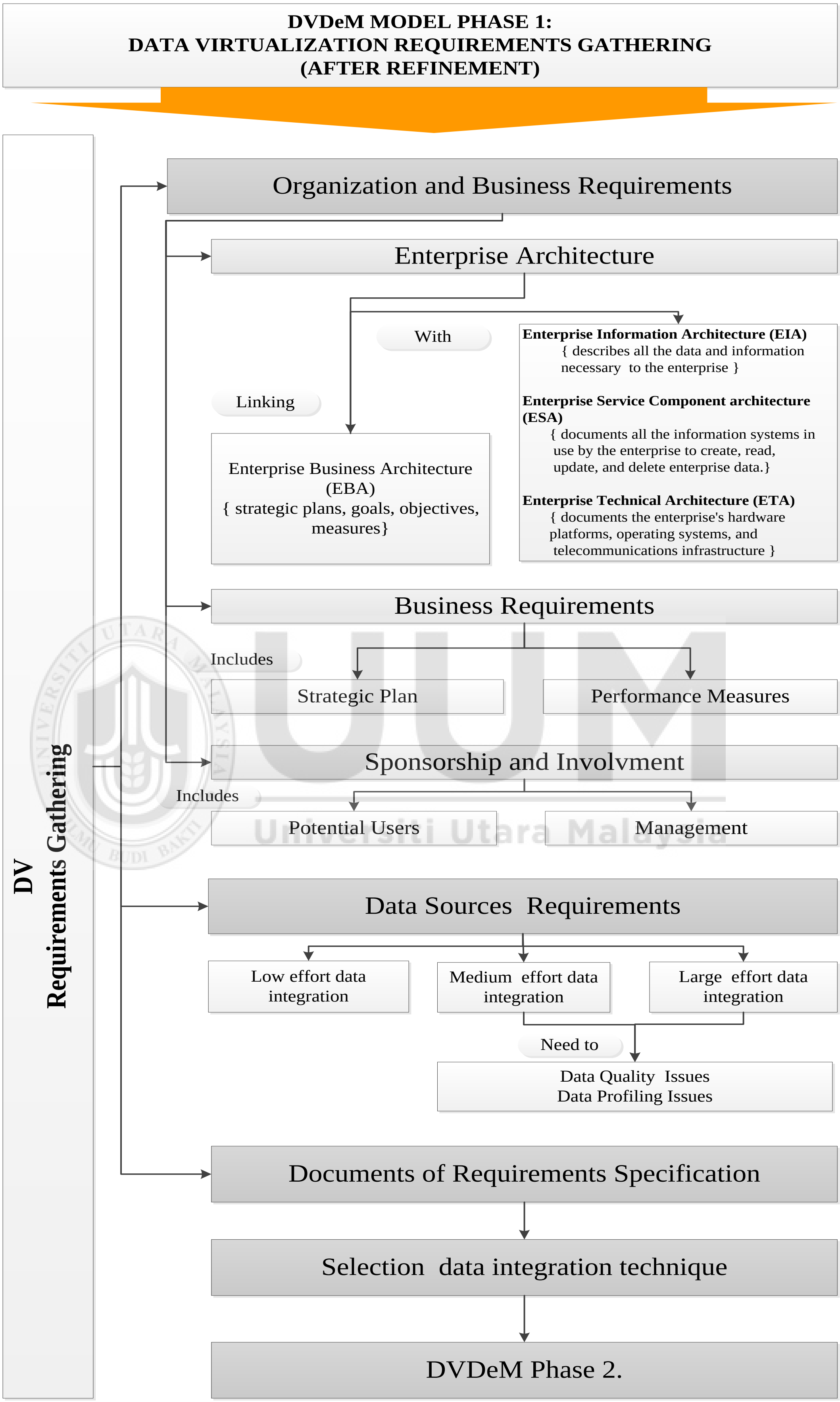


Figure 4.14. DV Requirements Gathering (After Refinement)

For the second phase of DVDeM proposed model, the correction was done according to comments and suggestions regarding flows, connection, and others from experts. (see Figure 4.15 and 4.16).

i. The GODV approach was explained in detail as well as shows its components.
ii. The data input for GODV approach was modified to be virtual tables, virtual data marts, and imported data sources.
iii. The ambiguous was removed regards to data preparation by explained the most sub-components in detail.
iv. In general, all phases were numbered as well as the flow between phases and its components were named.

Universiti Utara Malaysia

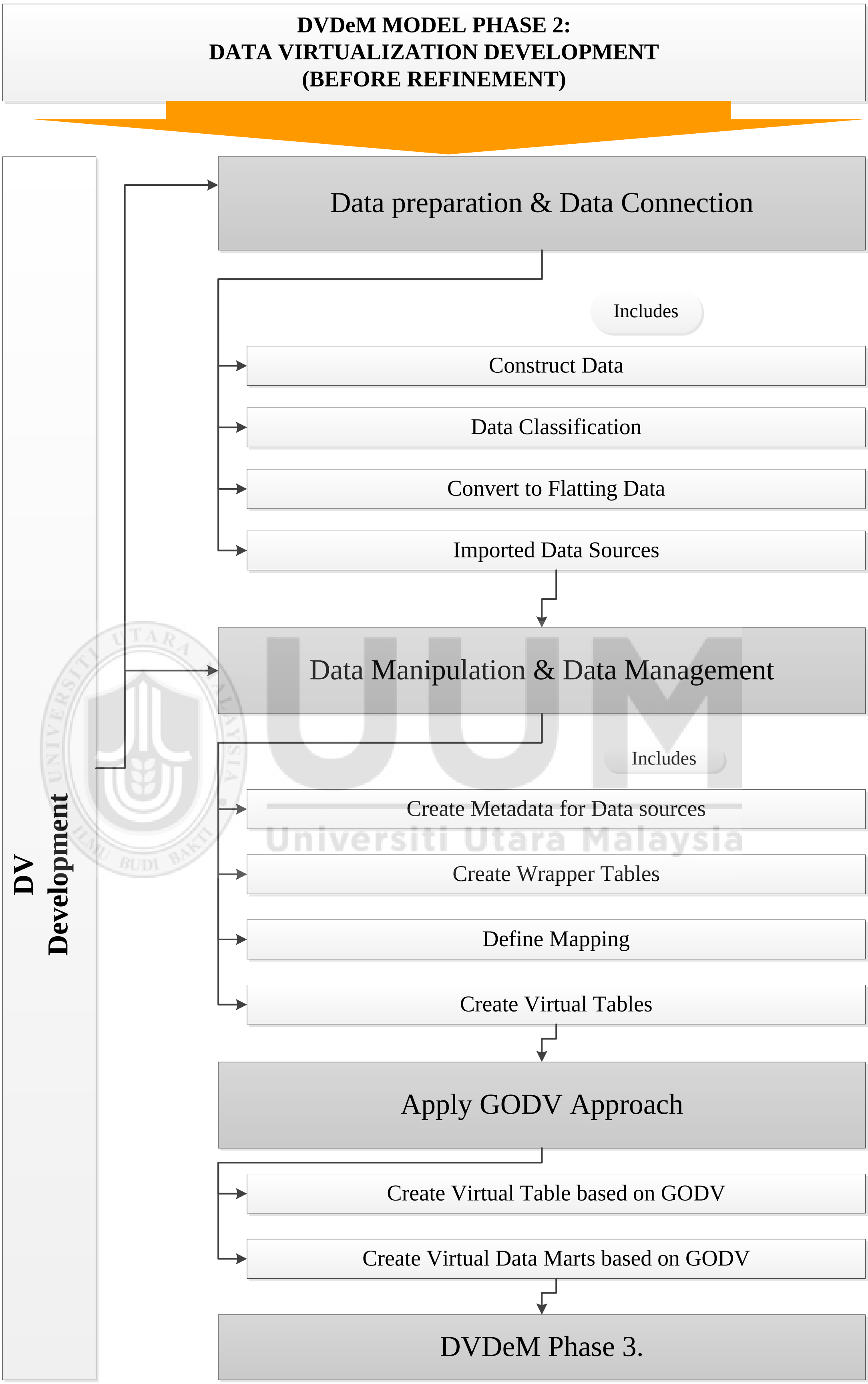


Figure 4.15. DV Development (Before Refinement)

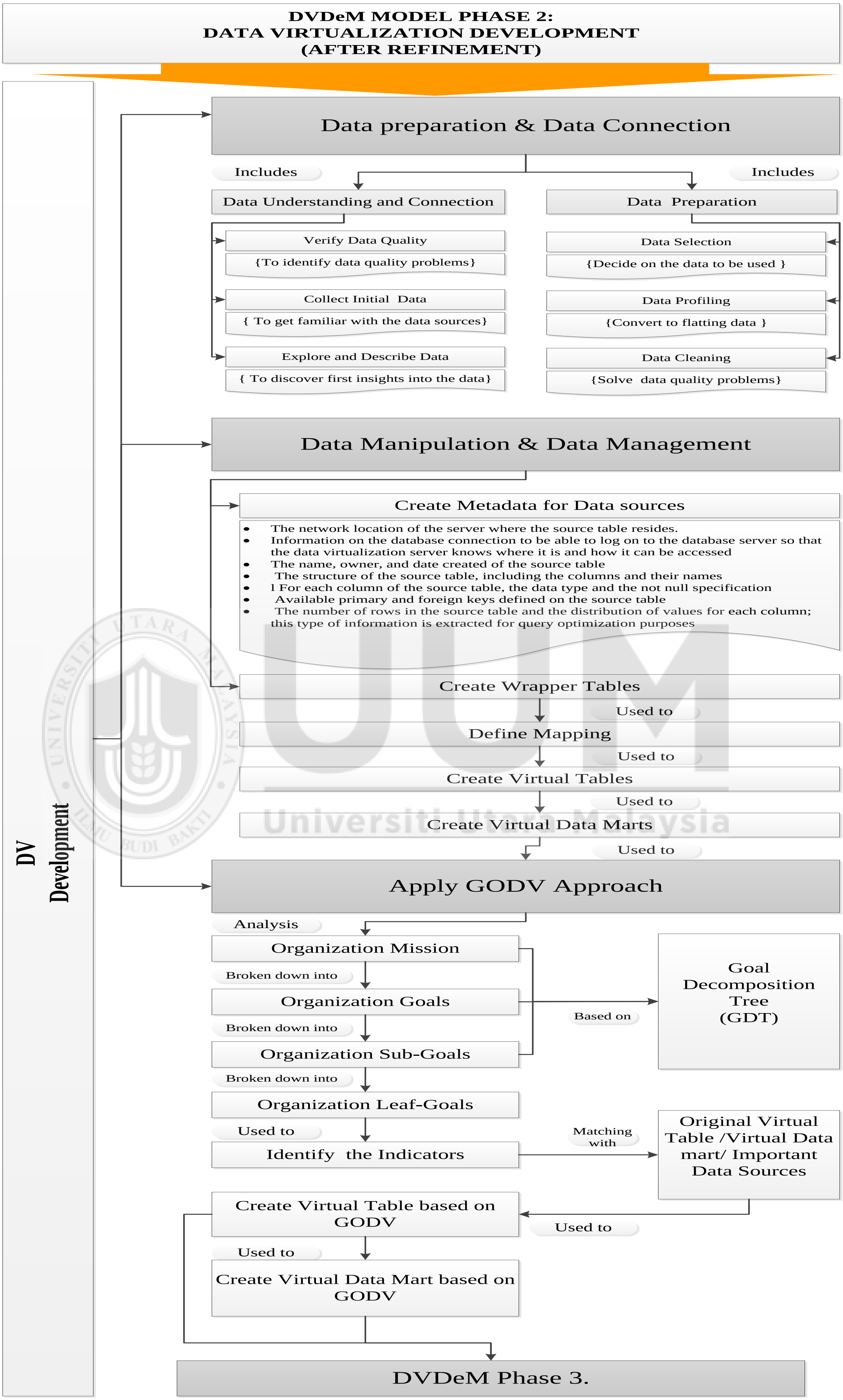


Figure 4.16. Data Virtualization Development (After Refinement)

For the third phase of DVDeM, the correction was done according to comments and suggestions regarding flows, connection, and others from experts. (see Figure 4.17 and 18).

i. The expression “data consumer” was replaced with “End users”.

ii. The data consumers types were explained in detail and its most sub-components were listed.

iii. The expression “publishing virtual table” was removed.

iv. In general, all phases were numbered as well as the flow between phases and its components were named.

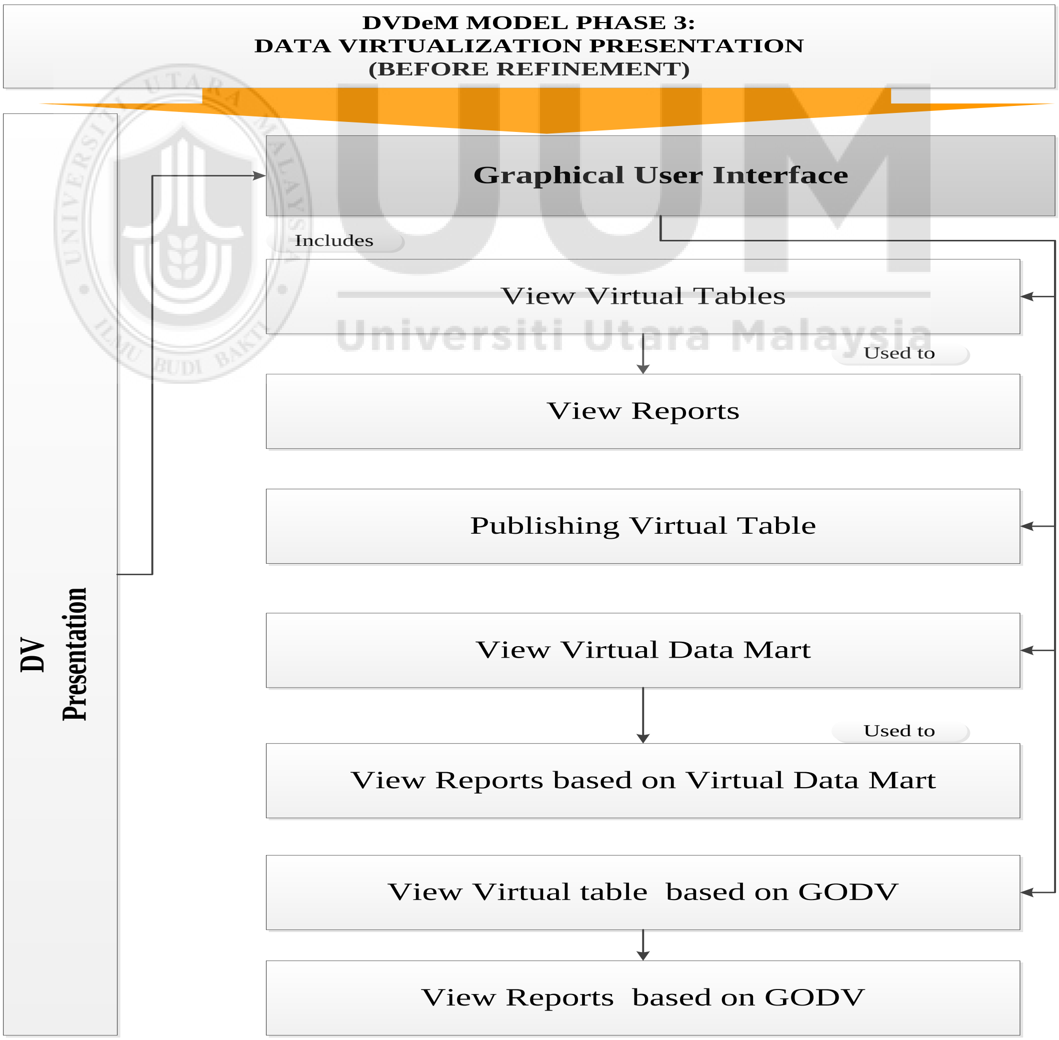


Figure 4.17. DV Presentation (Before Refinement)

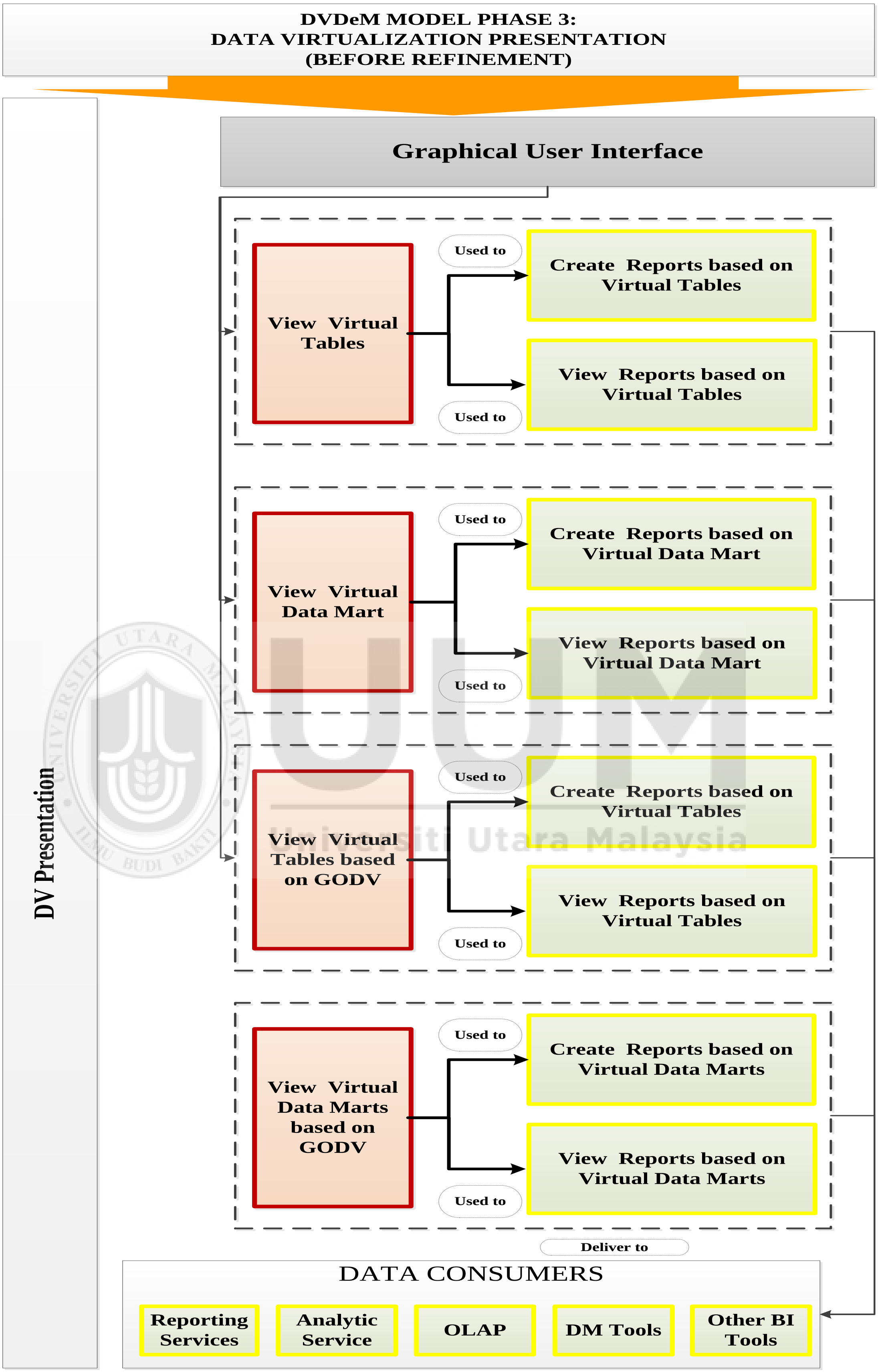


Figure 4.18. DV Presentation (After Refinement)

4.8 Develop the Proposed DVDeM (Final Version)

Eventually, to answer the comments and suggestions regarding connections and flows of all elements and consistency in terminologies, the proposed DVDeM were adapted and redesigned to give a better impression and enhance the readability of the model, which in turn led to the improvement of the final version of the proposed model. The improved model is then used to develop a prototype. Figure 4.19 visualize the main phases of the proposed model, while Figure 4.20 illustrates the final version of the proposed model.

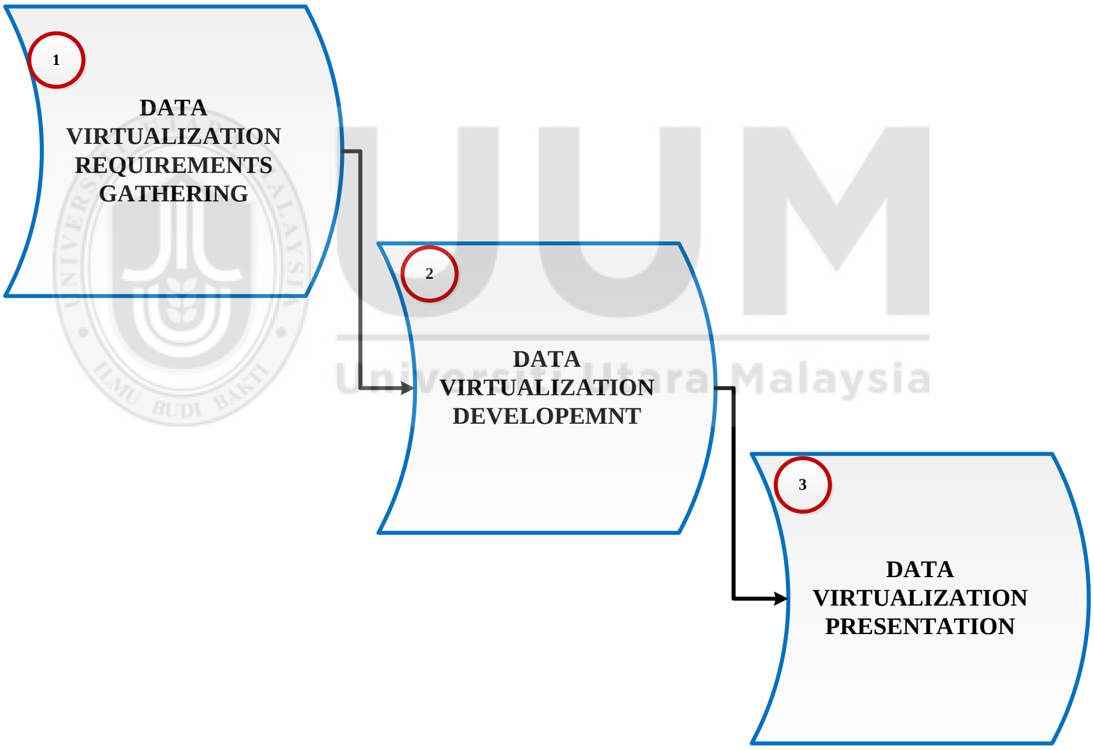


Figure 4.19. The Proposed Model DVDeM (Main Phases)

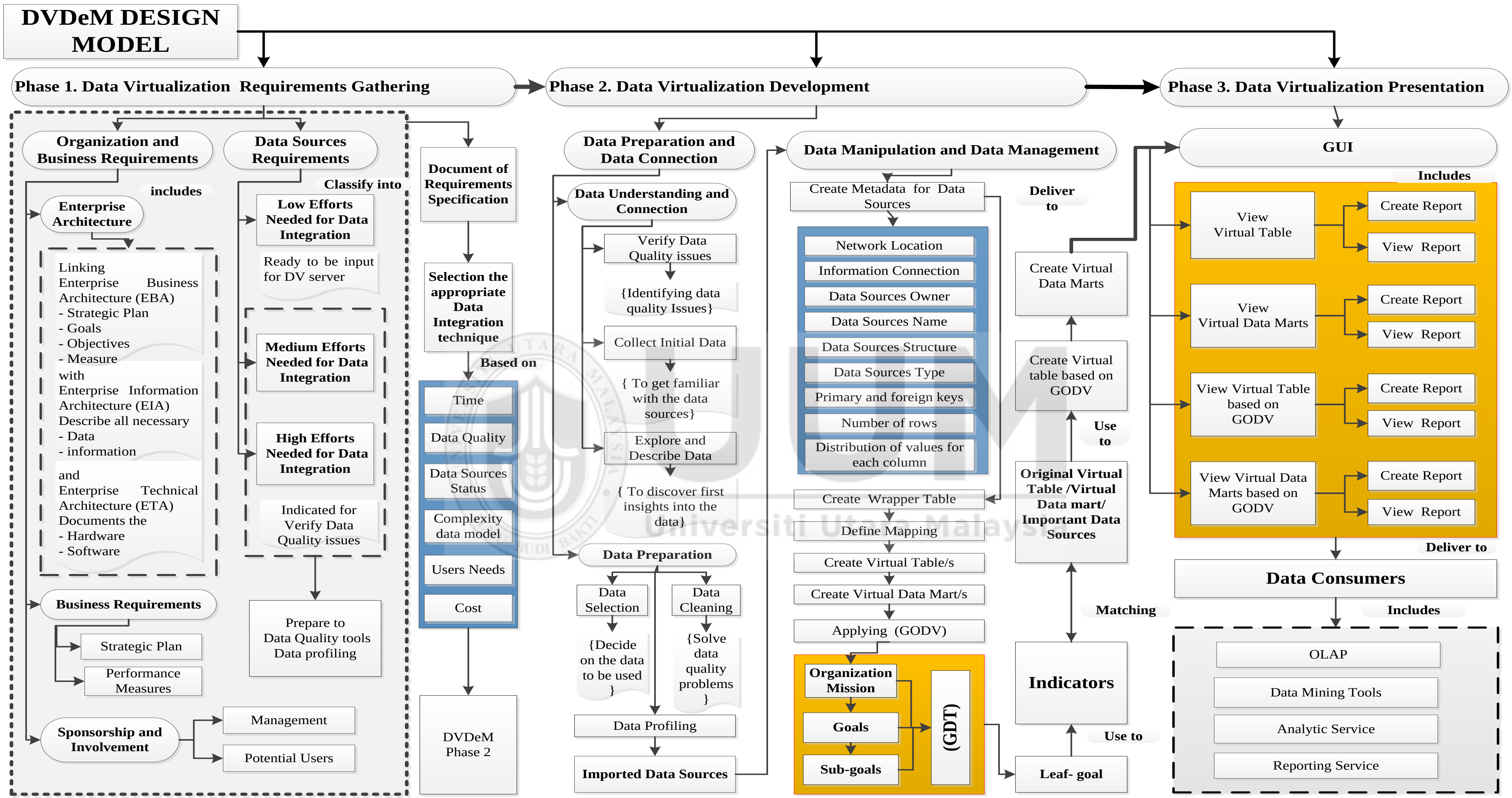


Figure 4.20. The Proposed Model DVDeM (final version)

4.9 Prototype Development

In verifying the proposed model as well as in achieving objective 4, a prototype was developed based on proposed DVDeM model. The prototype is named DV BI APPLICATION. It was developed through three phases based on DVDeM phases: DV requirement gathering, DV development, and DV presentation.

In the context of this study, and as mentioned earlier, the basic building block in the prototype development are data sources. Therefore, the details of the development of DVDeM prototype in practice will be discussed in Chapter 5. Accordingly, the following sections are highlighted on some rules and instructions that guide BI developers to develop BI system Based on DVDeM proposed model.

Universiti Utara Malaysia

4.9.1 DV Requirement Gathering

In line with the development of business intelligence systems where the use of huge data and deal with very complex processes, therefore, in the first phase of the prototype development, this study prepared a set of rules that used as guidance for the developer during the DV development phase. Consequently, in the next paragraphs, each of the components will be highlighted and list the rules related.

4.9.1.1 Organization and Business Requirements

Regarding organization and business requirement, and before starting design and development of virtual data model, BI developers should possess a sufficient knowledge about:

4.9.1.1.1 Enterprise Architecture Rules

i. The developer of BI application should prepare a link the enterprise business architecture (EBA) which includes (strategic plans, goals, objectives, and measures) with its enterprise information architecture (EIA), enterprise service component architecture (ESA) and enterprise technical architecture.

ii. Establishes guidelines, standards, and operational services that define the enterprise's computing technology environment.

iii. Describes all the data and information necessary to the enterprise

iv. Documents all the information systems in use by the enterprise to create, read, update, and delete enterprise data.

v. Before an enterprise can define, design, and implement the architecture for its strategic information management systems, including virtual tables, virtual data marts, decision support systems, and executive information systems, it must first document the environment in which these systems will be implemented.

vi. Developers should also provide at least a minimum of infrastructure that allow the design and implementation of data integration system in virtual manner.(refer to Section 4.3.1.1.3).

4.9.1.1.2 Business Requirements Rules

i. Sufficient knowledge about the specific business problem in the intended organization and should address by the data virtualization solution.

ii. An enterprise should never undertake system development efforts, particularly engineering a BI application, without first determining its strategic business and information requirements.

iii. The best source for the business requirements is the enterprise strategic plan and the performance measures identified in the plan.

4.9.1.1.3 Sponsorship and Involvement Rules

i. The developers of BI systems should ensure there are sufficient resources are available.
ii. DV users will have the most influence on acceptance of the virtualization data as sources for their BI applications; therefore, it is imperative that their needs are addressed. They are also the "owners" and "stewards" of operational data and thus are the best source for subject matter expertise.
iii. All potential stakeholders and users of the data virtualization, even executives, from every organizational unit and level, should be actively involved in data virtualization design, development, and management.

4.9.1.1.4 Data Sources Requirements Rules

In the world of DV, it is necessary to identify and understand the sources of data before the integration process begins. Therefore, some important rules should be undertaken.

i. Identify data source.
ii. Understanding the data source.
iii. Save results for further reuse.
iv. Knowing the organization's ability to deciding which type of data integration technique will choose.

Universiti Utara Malaysia

v. Classify the data sources based on the degree of the integration effort, which are high effort for data integration, medium effort for data integration, and low effort for data integration.
vi. Verifying data quality issues.
vii. Selection data integration technique.

In line with above situations, building a BI system requires effort, time and high cost. In the event of failure or a shortcoming of the use of any part in the construction of these BI systems that will lead to cost, time consuming and effort which it is not easy demolition and reconstruction BI systems. Additionally, after gathering, documenting, and determining the requirements and the data integration technique will select (selection data integration technique), accordingly, BI developer are ready for next phase (DV development).

Consequently, this phase proceeded with designing the interfaces namely the homepage, log-in and registration page, explorer data sources, about and contact us page. The design of these interfaces is intended to allow developers to have a clear and complete picture for the desired data that should be integrated. This also allows developers to understand the roadmap towards the second phase (DV development) in prototype design.

4.9.2 DV Development

In the context of this study, the DV development is considered as a core phase in the prototype development. Consequently, this phase consists of two main tasks: Data Preparation and Data Connection and Data Manipulation and Data Management, each of them have their own sub-components. Besides, this phase involves three main blocks namely, sources tables, virtual tables, and the mapping that location between them. As discussed extensively in the previous chapter, two actors can log on to the prototype concurrently. Furthermore, each of them has certain authorization. In the next paragraphs the highlights of all of them have given.

4.9.2.1 Data Preparation and Data Connection

The DVDeM prototype provides an ability to connect with data sources whether these data located on different servers or on one server. When a connection is established, the prototype can access to all tables that are stored in data sources and select the intended table.

4.9.2.1.1 Data Understanding and Data Connection

As mentioned early, in data understanding and data connection, there is one main task namely data understanding which covers three sub-tasks: i) collect initial data: how to get familiar with the data sources, ii) how to explore and describe data: to discover first insights into the data, and iii) how to verify data quality: identifying data quality problems. In reality, in data connection task, just identifying and explore, describe the data. While data modification will happen in data preparation task which outlines in the next paragraph.

4.9.2.1.2 Data Preparation

As mentioned previously, for the data preparation, the real manipulation of the data would happen. There are three main tasks namely: i) data Selection: decide on the data to be used, ii) data profiling: convert to flatting data, iii) data cleaning: solve data quality issues. In the same aspect, the data preparation phase covers all activities to construct the final dataset (data that will be fed into the virtual table(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include a table, record, and attribute selection, data cleaning, construction of new attributes, and transformation of data for used as an input to create a virtual table(s), this data set named imported data sources. When a source table has been selected for import, the data virtualization server determines whether some values have to be transformed to a more standardized form. The reason is that different database servers store values of particular data types differently. For example, database servers might store floating point values differently. To be able to compare values managed by different database servers correctly, the values have to be transformed to a standard form. In other words, for each column, a data virtualization server has to check whether a transformation is required to transform the values to a more standardized form to make comparisons possible. This means that data virtualization servers have to understand for each different data source how specific data types are handled.

4.9.2.2 Data Manipulation and Data Management

In this part of the DVDeM prototype, there are six main tasks have performed namely: create metadata for data sources, create a wrapper, defined mapping, create a virtual table, create virtual data marts, and apply GODV approach which outlines in the next paragraphs.

4.9.2.2.1 Create Metadata for Data Sources

In this part of DVDeM prototype, the metadata for imported data sources was created, this metadata includes

i. The network location of the server, where the source table resides.

ii. Information on the database connection to be able to log on to the database server so that the data virtualization server knows where it is and how it can be accessed.

iii. The name, owner, and date created of the source table.

iv. The structure of the source table, including the columns and their names.

v. For each column of the source table, the data type and the not null specification.

vi. Available primary and foreign keys defined on the source table.

vii. The number of rows in the source table and the distribution of values for each column; this type of information is extracted for query optimization purposes.

As spelled out in the previous section, by adopting the metadata the wrapper was created, the details described in the next section.

4.9.2.2.2 Create Wrapper

As mentioned above, the DVDeM prototype uses the data sources' metadata to create a wrapper table or wrapper for short. It can be stated here that the wrapper table is exactly source table in a virtual manner with one exception. Source data in the wrapper table have been transformed into standardized manner, as visualized in Figure 21.

Universiti Utara Malaysia

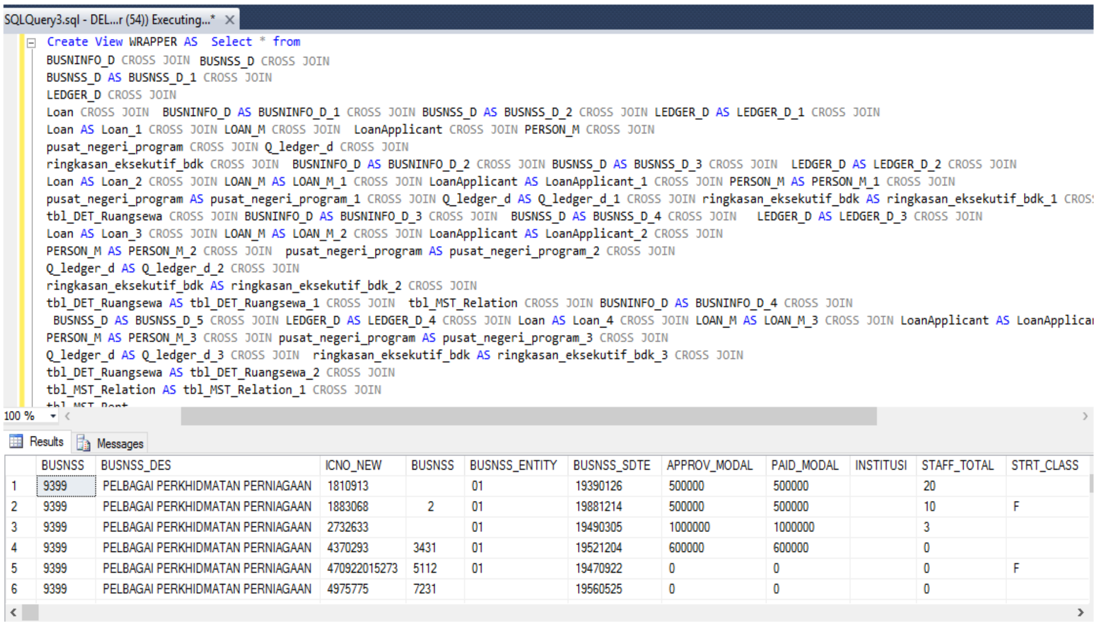


Figure 4.21. The Wrapper Table

As shown in figure 4.21, the wrapper table resembles the source table both in terms of content and structure. Furthermore, the wrapper table used to create the virtual table, thus, the virtual table should define before creating it, defining a virtual table means defining a mapping, the details described in the next section.

Universiti Utara Malaysia

4.9.2.2.3 Define Mapping

As mentioned earlier, the main reason behind the creation of a virtual table is to fulfill data consumers' requirements. Creation of a virtual table should be preceded by defining virtual tables on top of the wrappers. Besides, defining a virtual table refers to the process of mapping a virtual table structure as well as the process of data transformation from wrapper table(s) to virtual table. When the mapping is completed, the prototype provides the ability to create a virtual table/s based on the mapping, the details described in the next section.

4.9.2.2.4 Create Virtual Table

After defining a mapping, the DVDeM prototype has the ability to create virtual table, the pseudo code in Figure 4:22 visualize a mapping in SQL and provides a better understanding of what is possible when defining virtual tables. In this example, a virtual table on the CUSTOMER wrapper that contains the CUSTOMER_ID, the DATE_ OF_BIRTH, the POSTAL_CODE, the EMAIL_ADDRESS for all customers with a postal code that equal to 90017, 19108, or 48075, and customers who registered after 2006. In order to make the virtual table contain the relevant data, the GODV approach was applied, and the virtual tables were created based on such approach, the details about GODV are described in the following paragraphs.

Universiti Utara Malaysia

```
CREATE V_CUSTOMER AS
SELECT      CUSTOMER_ID,
FIRST_NAME || ', ' || MIDDLE_INITIAL || '' LAST_NAME AS FULLNAME,
DATE_OF_BIRTH, POSTAL_CODE, EMAIL_ADDRESS
FROM        CUSTOMER
WHERE       POSTAL_CODE IN ('90017', '19108', '48075')
AND   YEAR (DATE_REGISTERED) > 2006
```

Figure 4.22. Create Virtual Table

The prototype also provides the ability to apply the proposed goal oriented approach based on data virtualization (GODV) as mentioned earlier, based on the first phase of DVDeM model, the GODV goes to analyzing the organization's mission and broken them down into goals, these goals also are broken down into sub-goals. this process continue until reach a leaves goal (goals that cannot be broken); the leaves goals are used to extract the indicators which will use later for matching with existing virtual table or other data sourcing to produce a new virtual table or altering the existing virtual table that definitely contains a relevant data. For example, GODV scans the whole virtual table, refine, and alter it by adding an extra column named Is-relevant (altering existing virtual table). The pseudo code that resulted via applying this approach is shown in Figure 4.23. The same example in Figure 4.22 is used and a new column named Is-relevant is added to the intended virtual table.

```
ALTER V_CUSTOMER ADD Is-relevant string
UPDATE V_CUSTOMER
SET Is-relevant ='relevant'
WHERE POSTAL CODE=90017;
```

Figure 4.23. Adding New Column to Virtual Table

Now, a query posed on the virtual table to extract the intended data using one "Where" statement can be performed. For example, *Select * from V-CUSTOMER where Is-relevant ='relevant'*. In order to orient to a specific business line or team, the virtual data marts were created based on the virtual table, the details described in the next section.

4.9.2.2.5 Create Virtual Data Mart

In addition, the prototype provides the ability to break down the virtual table into a set of virtual data marts based on the privacy settings of the data. For example, human resources data are stored in virtual data mart and sales data in another virtual data mart and so on. The virtual table is divided into many virtual data marts. Finally, the prototype provides data refreshment option in order to get live data (real time data). Besides, as mentioned in the previous paragraph, the main output for this phase is a virtual table / virtual data mart which can be used as an input to the last phase in our prototype (DV Presentation) as described in detail in the next section.

4.9.2.2.6 Apply GODV Approach

In the context of this study, in order to deliver a relevant data and minimize the virtual table size, the GODV approach was proposed. As mentioned earlier, this approach is analysis the organization mission and broken it down into some organization goals, consequently, these goals in turn also break into a set of sub-goals, however, this process continues iteratively until we reach the goals which leaf-goals which are the goals that cannot be divisible. The goal decomposition is based on goal decomposition tree model (GDT). later, use the leaf goals to identify the indicators and accordingly, these indicators will be used in matching with the existing virtual tables and produce new virtual tables or altering the existing, finally, deliver the virtual table based on relevant data to the BI end users (data consumers).

4.9.3 DV Presentation

The results obtained from the above phases are virtual table and/ or set of virtual data marts that contain the required data. This data can now be used to create and display the required reports based on the requirements of the organization concerned. A graphical user interface (GUI) was created using visual studio 2015 associated with SQL server 2015 was developed. This prototype has ability to access and use of data via the GUI, whether these data are stored in the virtual table or in virtual data marts and display the required reports. The DV presentations were extensively discussed in Chapter 5. In the next section, the design of use case diagram to cover all the prototype functions is presented.

Universiti Utara Malaysia

4.9.4 Designing Use Case Diagram

Use case is one of the powerful tools used to represent the prototype functionality. Use case diagram helps to model the structure of the prototype, as well as helping to determine and realize the whole information related to system functionality. Figure 4:24 illustrates the use case diagram for the prototype's core functions.

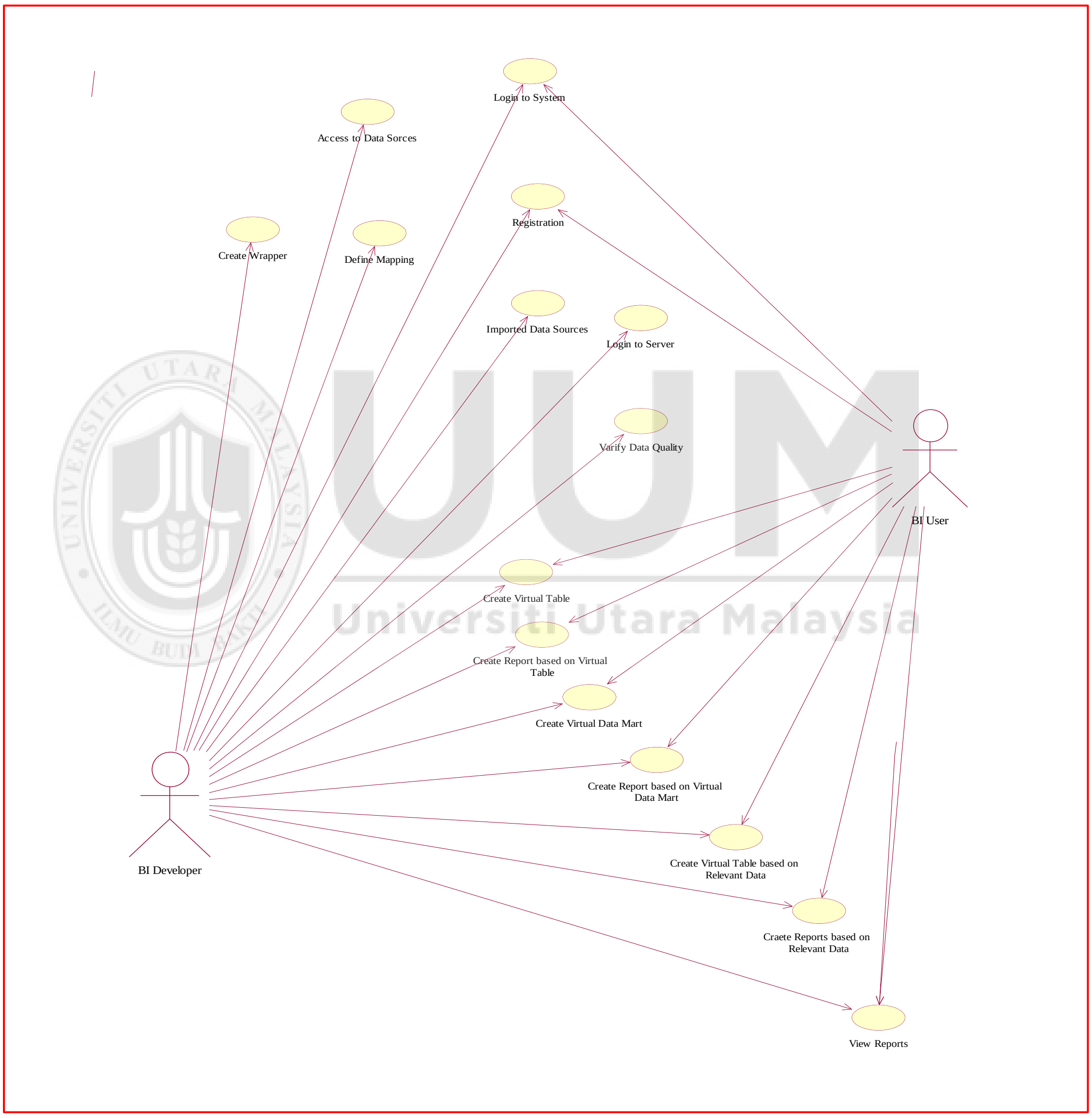


Figure 4.24. Use Case Diagram for Prototype Development

An actor is a behavioral classifier which specifies a role played by an external entity that interacts with the system (e.g., by exchanging signals and data), a human user of the designed system, some other system or hardware using services of the subject.

Accordingly, in the context of this study, there are two types of actors (BI user and BI developer). Nevertheless, each actor has his own privilege to do. In regard to BI developer, this actor that plays important role in DVDeM prototype and he can access to all DVDeM prototype's functions as clearly visualizes in Figure4.26. Besides, the BI user, its role is limited to access to the DVDeM prototype outputs and cannot access the other DVDeM prototype functions. Moreover, both BI developer and BI user can access the system by presenting their IDs and password to access uses the functionalities available, as clearly discussed it the next section.

4.10 The Prototype Authorization

The actors in this prototype are the model developer and the user of the system. Both BI developer and BI user can access the system by presenting their IDs and password in order to access uses the functionalities available. For example, the BI has access to all prototype functions for the purposes of maintenance, update data, and system development. The BI user however, does not have the same permission. Figure 4.25 visualize the prototype authorization window.

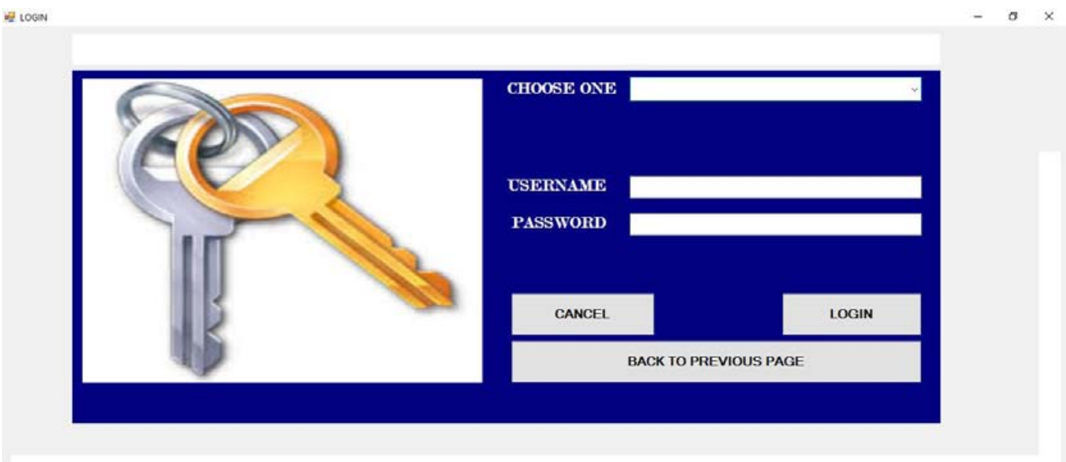


Figure 4.25. The Prototype Authorization Window

4.11 The Applicability and Robustness of DVDeM Prototype

As mentioned early, the prototype was developed based on the DVDeM proposed model, therefore, it does reflect the applicability and robustness of the proposed model in these ways:

i. Supports the integration of multiple data sources.
ii. Supports in verifying data quality issues.
iii. Supports the decision-making process in providing near real-time data.
iv. Is robust in view of changing data sources.
v. Supports flexible transformations.
vi. Can be easily deployed in a suitable implementation environment.
vii. Is robust in error control and help.
viii. Is complete enough to handle the various extraction, transformation, and loading operations.
ix. Is simple in creating and maintaining in comparing with other prototypes.

Universiti Utara Malaysia

It can be concluded that the DVDeM prototype reflects the applicability and robustness of the proposed model, more details regarding DVDeM prototype implementation and prototype evaluation are highlighted in Chapter 5.

4.12 Summary

This chapter starts with a comprehensive definition of DVDeM, the phase of DVDeM divided into three main phases (DV requirement gathering, DV development, and DV presentation). Furthermore, each of these phases divided into elements of the structure components and sub-components, and elements of the navigation components for DVDeM respectively. These components and elements have been derived from a series of comparative study on existing BI model, approaches, methods, and guidelines as well as real-world BI solutions. Also, the proposed processes involved in making DVDeM were outlined.

Next the DVDeM first version was formalized and discussed. In verifying the proposed model, the validation process was conducted. Hence, DVDeM was validated through two validation methods, namely the expert review and focus group discussion. These methods were found to be useful for validating the proposed model as they brought extensive reviews from different perspectives. All the results were examined in order to revise the proposed model. Thus, the final version of the proposed model was formalized. Besides, these outcomes serve to support the objectives one and two of the study.

Universiti Utara Malaysia

To examine whether the expectation in objective three achieved and the need DV design model as stated in Chapter 1 is met, this study has developed a prototype based on DVDeM proposed model. Accordingly, this study was translated the model in the Figure 4.26 into BI prototype.

One of the aims of this chapter is to have a means to validate the proposed DVDeM model as discussed in this chapter through end user testing. Having elaborated the section above, this study concludes that the aims were achieved by having the prototype based on DVDeM proposed model. Besides, this outcome serves to support the objectives three of the study.

Consequently, the prototype based on proposed DVDeM model is ready for use in BI environment. As stated earlier, the prototype should be implemented in real-world case studies, because it is necessary to determine whether the proposed phases and components were mentioned extensively in Chapter 4, be workable in BI environment. In accordance, Chapter 5 details the prototype implementation in case studies.

In line with above situations, the proposed model is intended to help developers to develop BI systems using DV technique. However, the proposed model is not restricted for this purpose only, developers can also apply this model with their existing BI application to enhance and provide real-time data (live data).

Universiti Utara Malaysia

CHAPTER FIVE
PROTOTYPE IMPLEMENTATION IN CASE STUDIES

5.1 Overview

The chapter has identified all the concepts to build the proposed model and its prototype. During the identification, careful consideration was given to derive its components through a cycle of analyses, focus group studies and expert review. As a result, the proposed DVDeM was outlined. Finally, the prototype was developed thus marking the achievement of objective 3. The developed prototype is named DV BI APPLICATION based on the proposed DVDeM model. This chapter also discusses the results of usability of the DVDeM prototype by implementing the DVDeM prototype in two case studies. The purpose of prototype implementation was to measure users' perception towards usability of the proposed DVDeM prototype. The measurements were made through an instrument named Q-U, which comprises of six main dimensions, visibility, flexibility, learnability, application behavior, error control and help, and near real time decision making.

To provide a clearer picture of the case studies, the next paragraph presents the organization background, business problem, existing application, the proposed DV solution, the implementation process, and the benefits that were reaped.

i. The Organization's Background

The organization background is essential as this information will be a high-level of abstraction, which includes its industry, products, markets, revenues and so on.

ii. The Organization's Business Problem

Problem faced by an organization needs be identified by the organization. Then, an organization will review its business needs and seek the most feasible solution. Typically, business needs relate to data integration and the possibility of solving this problem by applying the proposed model.

iii. The Organization's Existing Application

Full description of the environment that is currently used by the organization is discussed under this section. This section also deliberates on the obstacles facing the organization and aspirations of this organization in solving these problems.

iv. The DV Proposed Solution

In this section, the full description of how the proposed model will be applied to meet the business needs. A thorough discussion on the proposed model, particularly the proposed model's components, data sources, architecture, in addition to a comprehensive description of how to source data integration successfully is presented.

v. The Implementation Process

This section provides insights into the process used to implement the solution based on the proposed model. We also highlight implementation advice for others based on lessons learned.

vi. The Benefit

This section describes the enhanced agility and other business benefits as well as IT benefits realized from the adoption the solution based on the proposed model.

Universiti Utara Malaysia

Furthermore, these case studies were performed as proof of concept for any technique or technology that lies in its adoption. In the next paragraphs, the prototyping implementation in case studies is demonstrated.

5.2 Prototype Testing In Business Sector

As mentioned earlier and for the purpose of providing a complete and rich implementation for case studies, the case studies profile should include: organization background, business problem, existing application, the proposed DV solution, the implementation process, and the benefits. In the next sections, the details about the implementation of DVDeM prototype in the business sector are discussed.

5.2.1 The Organization's Background

This organization is an agency under the purview of the Ministry of Rural and Regional Development, was established on 1 March 1966 as a statutory body by an Act of Parliament because of the first Bumiputera Economic Congress resolution in 1965. The Council is responsible for developing, encouraging, facilitating and fostering the economic and social development in the federation, particularly in rural areas. The corporate vision for this organization is to be an outstanding organization of trust, upholding the nation's pride, while the corporate mission of this organization is to spearhead the fields of entrepreneurship, education, and investment to enhance equity holding of Bumiputera. Loan management system is one of this organization systems, additionally, the current system also does not meet the needs of the business community. Accordingly, the DVDeM prototype is implemented on this system.

5.2.2 The Organization's Business Problem

The loan management system was built with physical data integration, which lacks in live data. Therefore, the reports being produced do not show real time or near real time data. This will implicate negatively the decision-making process of the organization. In additions, the current system also does not meet the needs of the organization.

5.2.3 The Organization's Existing Application

The existing loan management is based on physical data integration, accordingly, the data that is used to monitor the loan management system portal is not near real time data due to the difficulties of updating physical data integration techniques as extensively, highlighted in the problem domain of this study. Hence, the existing architecture of the loan management system is depicted in Figure 5.1.

Universiti Utara Malaysia

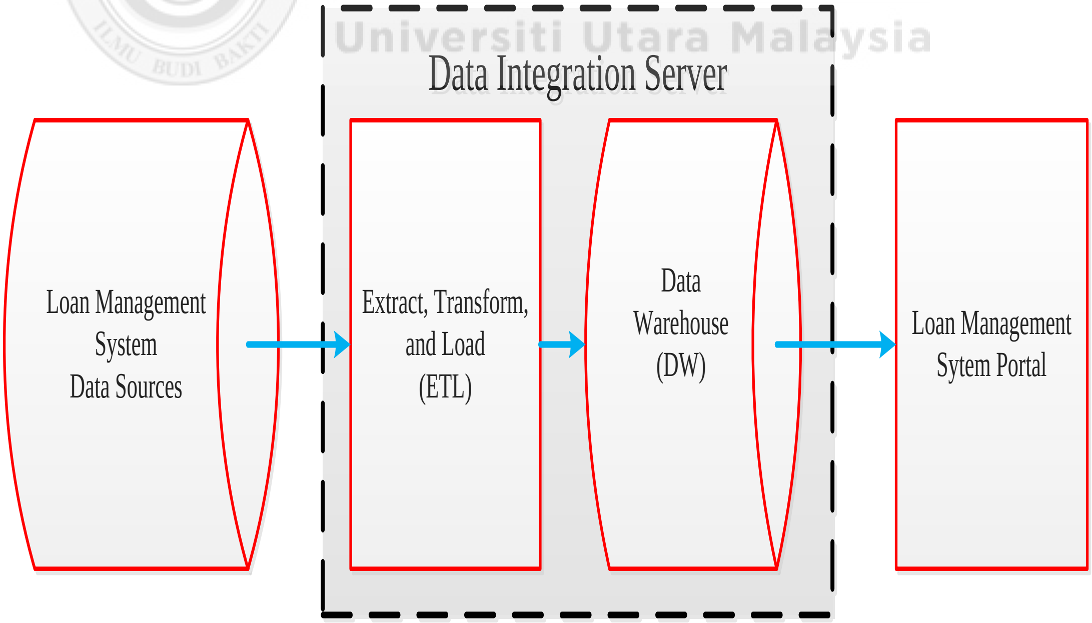


Figure 5.1. Existing Application Architecture

As clearly shown in Figure 5.1, the data were extracted, transformed, and loaded from loan management data sources via ETL to the subjected data (DW) which is located inside the organization BI system. Due to this, data are not regularly updated which could not support decision making process in time.

5.2.4 The DV Proposed Solution

The proposed DV solution of the loan management system was based on DVDeM proposed model, as illustrated in Figure 5.2 and extensively explained in the next sections.

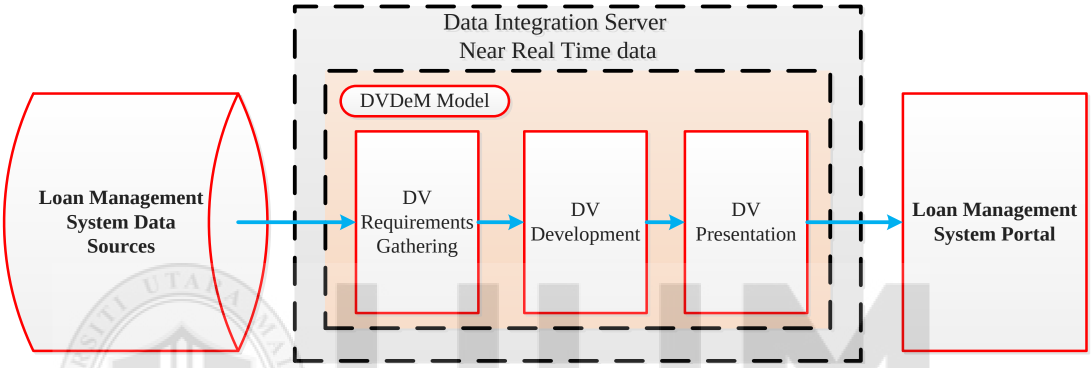


Figure 5.2. The DV Solution Architecture

5.2.5 The Implementation Process

In this section, the DV solution of this organization was implemented based on the proposed DVDeM model. The implementation was done based on all phases involved in the proposed DVDeM model phases starting with DV requirement gathering phase through DV development phase and ends with DV presentation phase. Details of the implementation process of DV solution are highlighted in the next section.

5.2.5.1 DV Requirements Gathering

As mentioned in chapter 4, requirements gathering phase is the foundation of the entire software development project. Hence, the organization requirements must be clear, correct and well-defined. During this phase, a full understanding of the organization’s requirements and business needs is essential as well as understanding the data sources (the location, type, and the verifying data quality issues). In addition to identifying the system requirements, a list of required resources (hardware and software) to implement this for a prototype is prepared. Therefore, in the context of this study, the main concern of this organization is to provide near real-time data for loan management system portal. Furthermore, the main requirements of loan management system are listed.

i. What is the percentage of loan recipients?
ii. What are the performance categories by business sector?
iii. What are the performance space rental premises by category?

In line with above situations, the DV solution based on DVDeM proposed model was developed and implemented for this organization to meet these requirements. The prototype also can deal with the other requirements based on business needs. As mentioned in chapter 4, the DV development of the proposed DVDeM model relies on the outputs of the requirement gathering and analysis.

5.2.5.2 DV Development

Based on DVDeM proposed model, the first task to DV development is to understand the data sources (raw data which is data that has not been processed for use).

The raw data understanding starts with an initial data collection and proceeds with activities to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information. Accordingly, there is a close link between business understanding and data understanding. The formulation of the data integration problem and the project plan require at least some understanding of the available data. Although raw data has the potential to become "information," it requires selective extraction, organization, and sometimes analysis and formatting for presentation.

The raw data for a loan management system consists of the data that collects information about customers who borrow money from a company. Every month the customer must pay at least the capital or the interest rate. Accordingly, the most important raw data are customer information includes contact details and balance fields, which will tell you how much the customer owes, the collector is the person who gives the money to the customers and collects it later point of time, loan given is the loan given by the company to the customers, and the payment received by the company. The database will keep track of payments made by customers, Additionally, payments can be either the interest or/and capital. Accordingly, the raw data for loan management system consists of a lot of attributes spread over so many tables. Table 5.1 list the raw data for the loan management system.

Table 5.1
The Raw Data for Loan Management System

	Table Name	**Description**
1	Customer	✓ Customer-ID. ✓ Customer Name. ✓ Address. ✓ Balance. ✓ Age. ✓ DOB ✓ Status. ✓ Order-timestamp ✓ Customer-Address
2	Collector	✓ Collector-ID. ✓ Collector Name. ✓ Address. ✓ Age.

		✓ DOB. ✓ Status.	
3	Loan-Given	✓ Transaction-ID ✓ Customer-No ✓ Date-Loan-Given ✓ Amount-Given ✓ Interest-Rate ✓ Collector-ID (foreign key - from collector table). ✓ Customer-ID (foreign key - from customer table).	
4	Payment	✓ Return-ID ✓ Customer-ID ✓ Payment-Date ✓ Payment-Month ✓ Capital-Paid ✓ Interest-Paid ✓ Remarks	Near real-time data
5	Business-info	✓ Icno_New ✓ Busnss ✓ Busnss_Entity ✓ Busnss_Sdte ✓ Approv_Modal ✓ Paid_Modal ✓ Institusi ✓ Staff_Total ✓ Strt_Class ✓ Strt_Dte ✓ Strt_Dept ✓ Current_Class ✓ Current_Dte ✓ Current_Dept ✓ Untung_Bersih ✓ Amt_Jualan	
6	Mst-relation	✓ Nama ✓ Acc-No. ✓ Sis-Id ✓ Sis-Desc ✓ Status ✓ Status-Id ✓ Sp. ✓ Ups-Date	

As clearly indicated in Table 5.1 and in the context of the loan management system, the related data sources are always changing, hence, any delay of data updating will give a negative effect on the decision-making process. Accordingly, the near real time data used in the loan management system is payment data which includes Payment-Date (the last payment record - if any), Payment-Month (the last payment record - if any), Capital-Paid (the last payment record - if any), and Interest-Paid (the

Universiti Utara Malaysia

last payment record - if any). Furthermore, the raw data of the loan management system is needed in many activities to construct the final dataset (data that will be fed into the virtual table(s)).

As mentioned earlier, the data sources that are to be used to create virtual tables are not always correct. For instance, names are spelled incorrectly, numeric values are outside realistic boundaries, values in two fields have accidentally been switched, stored values do not represent reality, and values or rows are completely missing. If no actions are taken, this incorrect data is presented by the data virtualization server to the data consumers. The consequence is that incorrect data is used for decision-making. The quality of business decisions is for a large part dependent on the quality of the data. Therefore, data quality is an important consideration when designing business intelligence systems.

Generally, data integration is based on the mapping of the relationship between the old and new system database, and data analysis is to establish mapping relations, which also includes the code data analysis. Data integration steps generally, includes the process called data cleaning, data cleaning is mainly directed against the source database for the occurrence of ambiguity, duplication, incomplete, in violation of business rules or logic corresponding data. Data quality analysis is required before cleaning operation, it will identify problems in the raw data. With respect to the loan management system, the following integrity rule holds for the CUSTOMER table: date values in the ORDER_TIMESTAMP column should all be greater than *31 December 1999.* The reason is that the company did not exist before that date. In the other aspects, the misspelled value should have been corrected. The SQL statements which used for verifying data quality issues are listed below.

```
CREATE VIEW NEW-CUSTOMER AS SELECT * FROM CUSTOMER
```

```
WHERE ORDER_TIMESTAMP > DATE ('1999-12-19')
UPDATE NEW-CUSTOMER
SET Status = ‘Discontinued’ WHERE Status=’Discontineud’;
```

As clearly indicated from the SQL code, all the rows with a timestamp that is too old are removed from the result as well as all values in the column STATUS contains one misspelled value ‘Discontineud’, which is transformed to Discontinued. Accordingly, the New-Customer table will be one of the virtual table sources. Hence, data consumers retrieving data from this virtual table will not see those rows (incorrect data).

Based on the DVDeM proposed model, after verifying data quality issues the raw data is ready to be imported. Importing a source tables means that it is made known to the data virtualization server. Accordingly, the imported data sources for the loan management system consists of six tables (Customer, Collector, Loan-Given, , Payment, Mst-relation, and Business-info) as well as their related attributes, to be used later to create wrapper table as highlighted in the next paragraph. Accordingly, the wrapper table is defined on a source table, and it consists of all imported data sources attributed as shown in Table 5.2.

Table 5.2
The Loan Management Wrapper Table

	Table Name	Attributes	Description
1	CUSTOMER	✓ CUSTOMER-ID. ✓ CUSTOMER NAME. ✓ ADDRESS. ✓ BALANCE. ✓ AGE. ✓ DOB ✓ STATUS. ✓ ORDER-TIMESTAMP ✓ CUSTOMER-ADDRESS	
2	COLLECTOR	✓ COLLECTOR-ID. ✓ COLLECTOR NAME. ✓ ADDRESS. ✓ AGE.	

		✓ DOB. ✓ STATUS.	
3	LOAN-GIVEN	✓ TRANSACTION-ID ✓ CUSTOMER-NO ✓ DATE-LOAN-GIVEN ✓ AMOUNT-GIVEN ✓ INTEREST-RATE ✓ COLLECTOR-ID (FOREIGN KEY - FROM COLLECTOR TABLE). ✓ CUSTOMER-ID (FOREIGN KEY - FROM CUSTOMER TABLE).	
4	PAYMENT	✓ RETURN-ID ✓ CUSTOMER-ID ✓ PAYMENT-DATE ✓ PAYMENT-MONTH ✓ CAPITAL-PAID ✓ INTEREST-PAID ✓ REMARKS	Near real-time data
5	MST-RELATION	✓ NAMA ✓ ACC-NO. ✓ SIS-ID ✓ SIS-DESC ✓ STATUS ✓ STATUS-ID ✓ SP. ✓ UPS-DATE	
6	BUSINESS-INFO	✓ ICNO_NEW ✓ BUSNSS ✓ BUSNSS_ENTITY ✓ BUSNSS_SDTE ✓ APPROV_MODAL ✓ PAID_MODAL ✓ INSTITUSI ✓ STAFF_TOTAL ✓ STRT_CLASS ✓ STRT_DTE ✓ STRT_DEPT ✓ CURRENT_CLASS] ✓ CURRENT_DTE] ✓ CURRENT_DEPT] ✓ UNTUNG_BERSIH] ✓ AMT_JUALAN]	

Consequently, The SQL statements which is used for creating loan management wrapper table are listed below.

```
CREATE VIEW [dbo].[Loan-Management-Wrapper]
AS   SELECT          dbo.Payment.*,   dbo.Loan-Given.*,   dbo.Collector.*,
dbo.Customer.*,dbo.Business-Info.*,dbo.Mst-Relation>*
FROM             dbo. Payment CROSS JOIN dbo. Loan-Given CROSS JOIN dbo.
Collector CROSS JOIN dbo. Customer
```

UUM Universiti Utara Malaysia

```
GO
```

As indicated in Table5.2, the wrapper table shows the full contents of a source table. Also, a wrapper has the same structure as the source table it is bound to. in the context of the loan management system, not all the data to be used via data consumers, therefore, the virtual table should contain the only related data. Furthermore, wrapper table also manipulates the data sources to transform the data values to standard forms to be an input for the virtual table. The SQL statements which are used for creating the loan management virtual table are detailed below.

```
CREATE VIEW [dbo].[loan management Virtual Table]
AS SELECT * from  [dbo].[loan management Wrapper] where Condition1
Condition2,etc.
GO
```

Universiti Utara Malaysia

Based on the DVDeM proposed model, and to reduce the size of the virtual table and to make it consists of relevant data only, a goal-oriented approach based on data virtualization (GODV) was applied. The details regarding GODV are highlighted in the next paragraphs.

Applying GODV starts with the main goal which was obtained from the analysis of the mission of the organization “Spearheading the fields of entrepreneurship, education, and investment to enhance equity holding of Bumiputera”, Accordingly, one of the main goals obtained is “Developing successful and innovative entrepreneurs and producing a globalized human capital with integrity”. Based on GODV, the main goal can be further divided into three sub-goals which are “to provide a system that enables and promotes best practices for the loans business processes”, “to streamline the guaranteed loans business processes consistently across all regions and sectors”, and “to provide a loan management system that is supportable and maintainable”, as shown in figure 5.3.

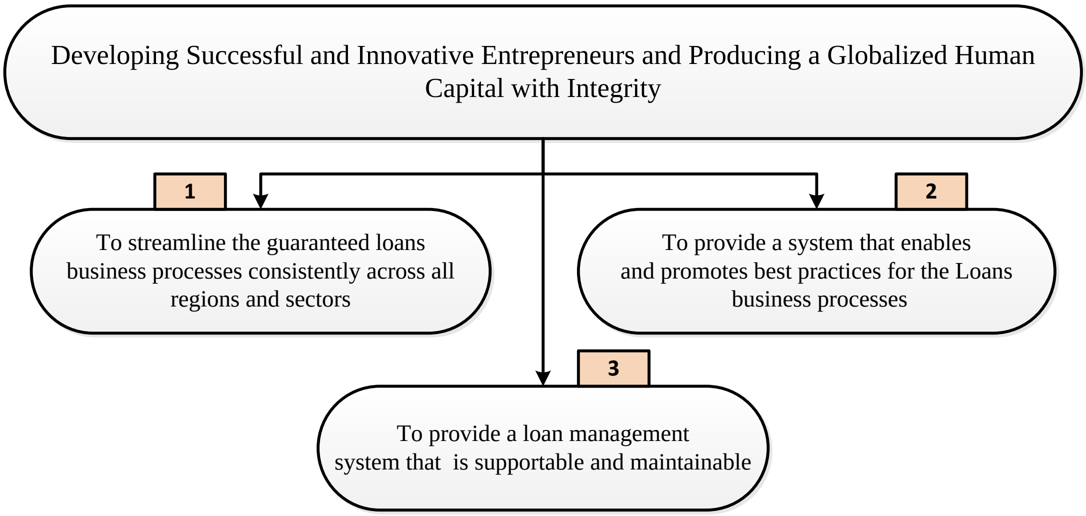


Figure 5.3. The Organization Main Goals Diagram

As clearly indicated in the Figure 5.3, three main goals are tackled in this organization, consequently, and in the context of the loan management system, hence, all these goals are related and should be analysed. Next paragraph list the goal analysis of loan management system.

A Developing successful and innovative entrepreneurs and producing globalized human capital with integrity.

A.1 To provide a system that enables and promotes best practices for the Loans business processes.

A.2 To streamline the guaranteed loans business processes consistently across all regions and sectors.

A.3 To provide a loan management system is dependable with an architecture that is supportable and maintainable.

A1.1 To provide a tool that assists both operational and management level users to apply best business practices.

A.1.2 To provide for accurate and efficient data collection and management.

A.2.1 To provide a tool that supports consistent and streamlined business processes across the department for loan management system.

A.2.2 To provide an assistance with the business transformation needed to streamline and gain efficiencies by improving the automation of the loan management business process and procedures.

A.2.3 To improve management capability for oversight, monitoring and control of the department's authority.

A.3.1 To provide a fully bilingual, modern and secure web enabled system.

A.3.2 To provide an automated ‘Real Time' Interface to the corporate financial system.

A.3.3 To provide an integrated and seamless interface with the department's document management suite.

In line with above situations, Figure 5.4 shows the goal decomposition tree for loan management system, which includes some of the related goals that are broken down into some sub-goals for loan management system that is managed by the intended organization.

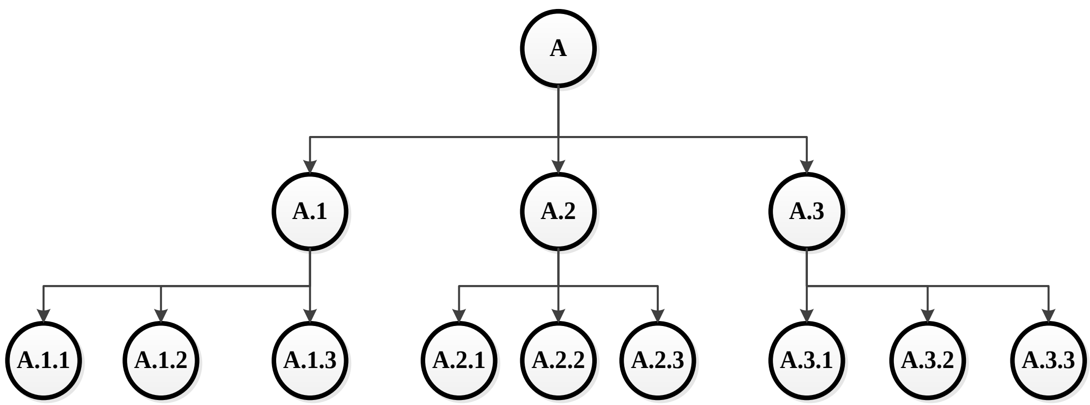


Figure 5.4. The goal Decomposition Tree for UTLC Goals (sub-goals)

In the same aspect, as mentioned earlier and based on GODV approach, the process of breaking down the goals into sub-goals will continue iteratively (sometimes two or more goals are combined using AND / OR operators) until it reaches the leaf-goals. Figure 5.5 shows the goal decomposition tree with leaf-goals.

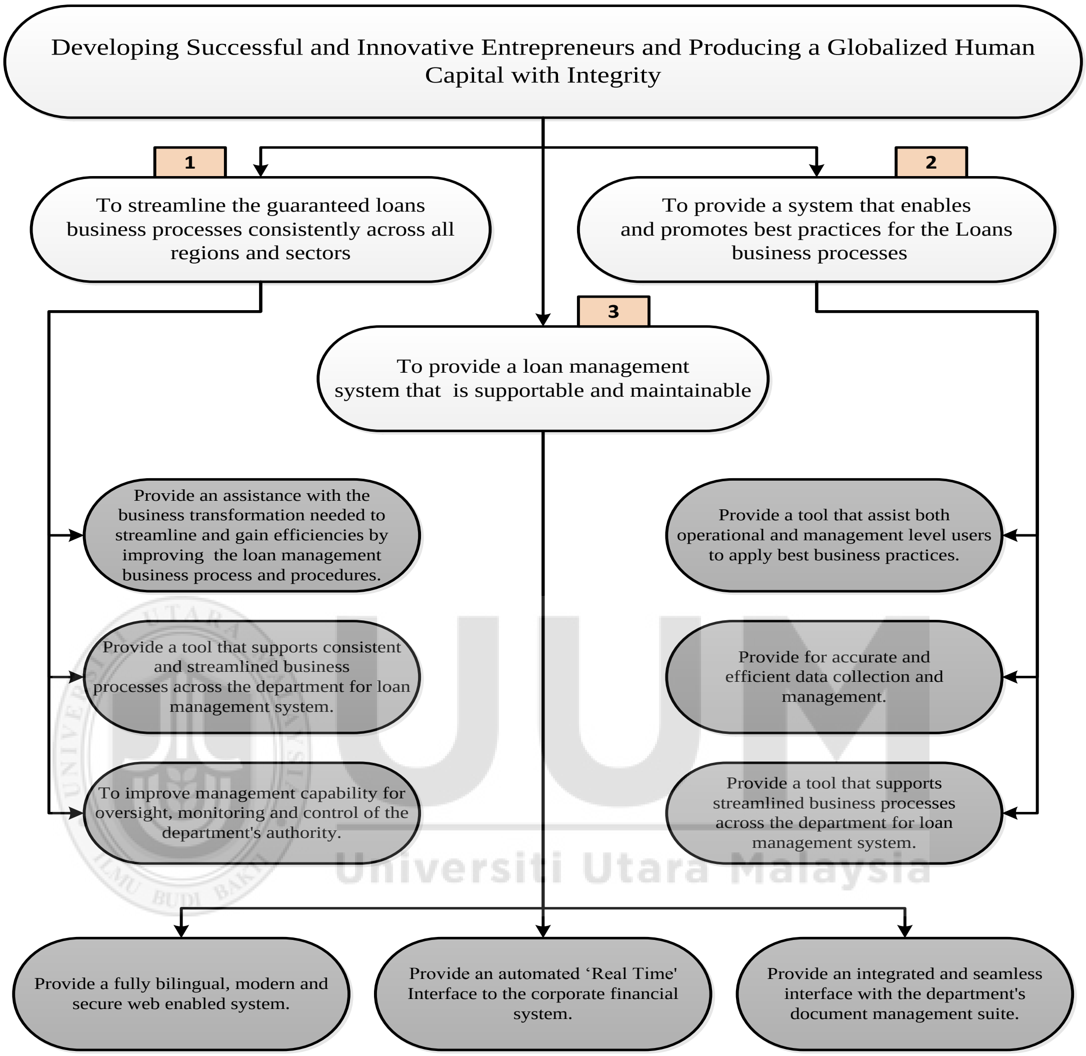


Figure 5.5. Goal Decomposition Tree with Leaf-goals

As clearly indication from the Figure 5.5, the leaf-goals were represented by the shaded area.

As mentioned in chapter 4, the core purpose behind the proposed goal oriented approach based on data virtualization is to create a virtual table which consists of relevant data only. Therefore, the attributes analysis will be used to describe the relationship between the measured attributes and the leaf goals. The attributes abstracted in this case study and their respective descriptions are listed in Table 5.3.

Table 5.3
The Relationship between Leaf goals and Measured Attributes

	The Leaf Goal	The Measured Attribute	Attribute Description
1	To provide an assistance with the business transformation needed to streamline and gain efficiencies by improving the loan management business process and procedures.	Loan management	The assistance that focuses on the transformation needed to streamline
2	To provide a tool that supports consistent and streamlined business processes across the department for loan management system.	Business processes across loan management system.	The focuses on the method of business process across loan management.
3	To improve management capability for oversight, monitoring and control of the department's authority.	System monitoring	The monitoring of department's authority for the organizations.
4	Provide a tool that assist both operational and management level users to apply best business practices.	Tools assistance	Tool assistance.
5	To provide for accurate and efficient data collection and management.	Loan management data collection	The accurate and efficient data collection and management.
6	To provide a tool that supports streamlined business processes across the department for loan management system.	Loan management	The use of both online and traditional business processes.
7	To provide a fully bilingual, modern and secure web enabled system.	Loan management security	A fully bilingual, modern and secure web enabled system
8	To provide an automated 'Real Time' Interface to the corporate financial system.	Real-time User interface for loan management	Automated 'Real Time' Interface to the corporate financial system.
9	To provide an integrated and seamless interface with the department's document management suite.	Documents loan management	integrated and seamless interface with the department's document management suite.

As clearly indicated from the Table 5.3, relevant attributes were determined and extracted; and accordingly, these extracted attributes are used to match with the data sources (wrappers/ original virtual table) to create the virtual table based on relevant data. The main difference between the virtual table based on GODV and existing virtual table is the resulting table (the virtual table based on GODV) contains the relevant data only. In a nutshell, the obtained relevant data is based on the analysis of the organization goals and broken down into sub goals in iteratively manner. This process will continue until it reaches to the leaf goals which are used to match the original data sources to extract the relevant data only. Figure 5.6 illustrates GODV approach at all levels (analysis of one goal only).

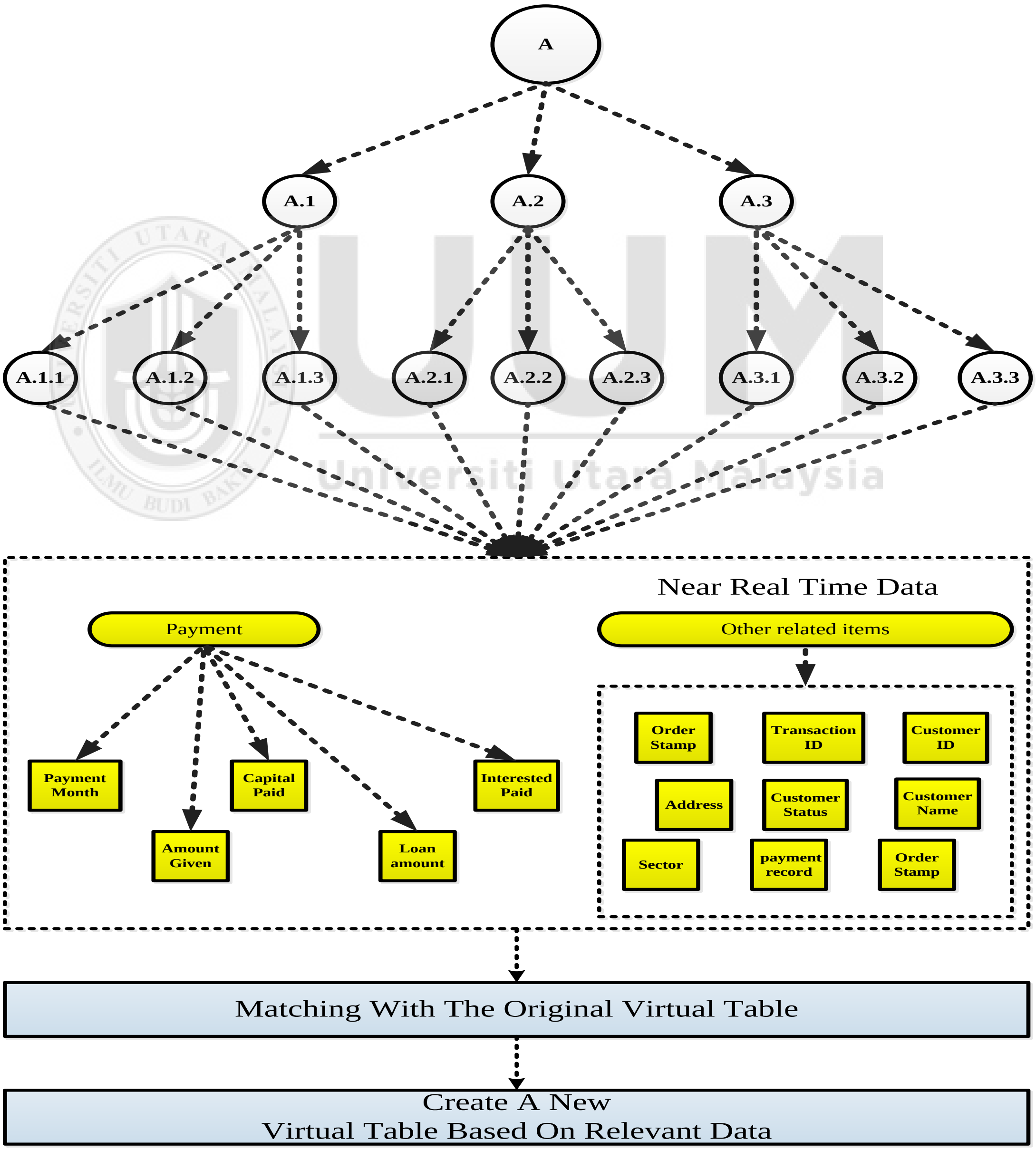


Figure 5.6. Applying GODV Approach (all levels)

As clearly indicated in the Figure 5.6, the findings from the goal analysis using GODV are many indicators which are used later to match with the original virtual table to create a new virtual table based on relevant data.

Furthermore, the procedure to create and present the required reports require creating multi-level virtual tables to obtain data; usually, some near real-time data such as (Loan-Given, Capital-paid, Customer, Payment-amount, and Collector) are resulted from conducting mathematic operations and aggregation on other data. The SQL statements which are used for preparing near real time data for the loan management system are detailed below.

```
CREATE VIEW [dbo].[v_LoanRecipients] AS SELECT        COUNT(*) AS count,
SIS_Desc FROM              dbo.tbl_MST_Relation WHERE          (SIS_Desc IN
('Pembiayaan Perniagaan', 'Sewa', 'Bumiloan')) GROUP BY SIS_Desc
GO
```

```
CREATE VIEW [dbo].[v_PercentageOfAllSector] AS SELECT         COUNT(*) AS
Expr1, SIS_Desc FROM          dbo.tbl_MST_Relation GROUP BY SIS_Desc
GO
```

```
CREATE VIEW [dbo].[v_PerformanceCategoriesbyBusinessSector] AS SELECT
COUNT(*) AS qty, LEFT(BUSNSS, 1) AS bussid, CASE WHEN LEFT(BUSNSS, 1) =
'1' THEN 'Argriculture' WHEN LEFT(BUSNSS, 1) = '2' THEN 'Manufacturing'
WHEN LEFT(BUSNSS, 1) = '3' THEN 'Wholesale' WHEN LEFT(BUSNSS, 1) = '4'
THEN 'Commerce' WHEN LEFT(BUSNSS, 1) = '5' THEN 'Construction' WHEN
LEFT(BUSNSS, 1) = '6' THEN 'Services' WHEN LEFT(BUSNSS, 1)= '7' THEN
'Others' WHEN LEFT(BUSNSS, 1) = '8' THEN 'Transportation' WHEN LEFT
(BUSNSS, 1) = '9' THEN 'Others Services' END AS CATEGORIES FROM
dbo.BUSNINFO_D WHERE(LEFT(BUSNSS, 1) NOT IN ('A', ' ')) GROUP BY
LEFT(BUSNSS, 1)
GO
```

Universiti Utara Malaysia

These data will be used to create the intended virtual tables and then use to create the intended BI reports for loan management system as clearly indicated in the next section.

5.2.5.3 DV Presentation

As mentioned in chapter 4, the presentation phase can be considered as an end user layer and should contain tools that have the ability to display meaningful information in different formats based on user needs. In this regard, a graphical user interface was created to allow users to access the virtual tables. Users can then view the virtual tables and generate reports from it. This GUI consists of eight main components namely, view virtual table, view virtual data marts, create reports based on virtual table, create reports based on virtual data marts, view virtual table based on GODV, create reports based on GODV and view reports as well as other general components.

As mentioned previously, the main aim of this study is to support decision making process by delivering near real time data that can be used by BI stakeholders in supporting the decision-making process. Therefore, the output for the phase 3 in this model (data virtualization presentation) is to present the virtual tables as well as the organization's reports based on organizations requirements. Hence, Figures 5.7, 5.8, and 5.9 depict the reports of business sector organization which are created based on the organization requirements.

Universiti Utara Malaysia

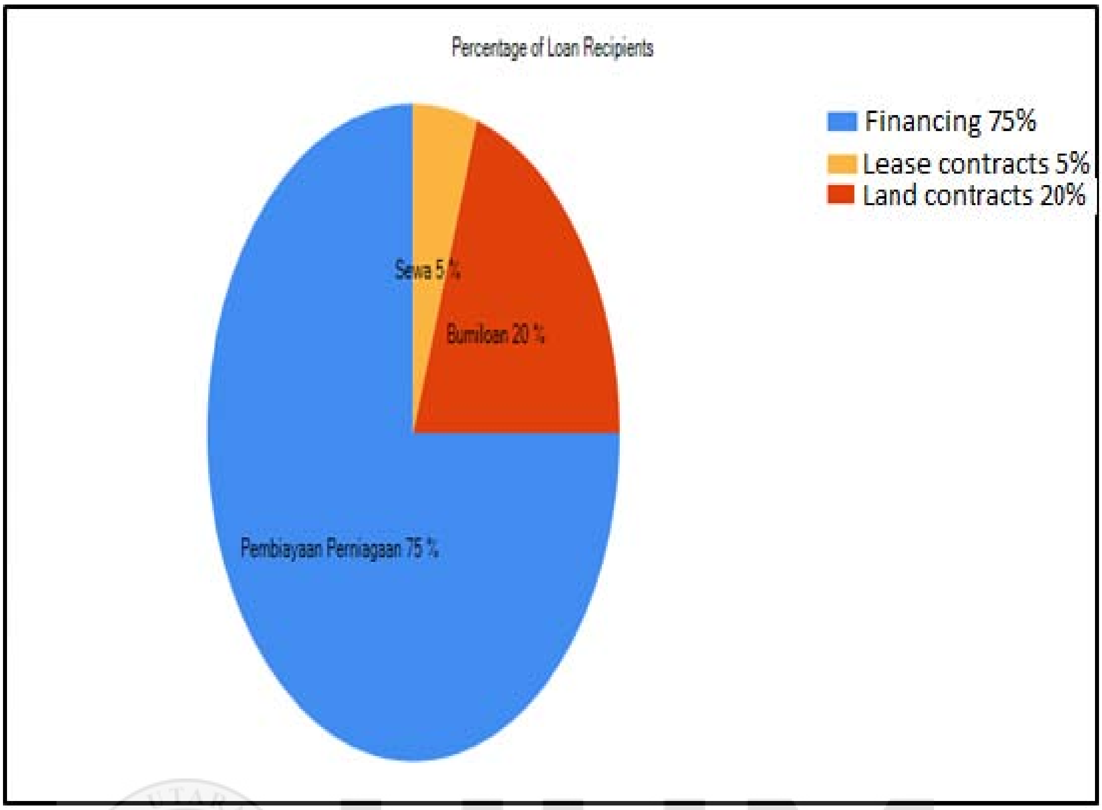


Figure 5.7. The percentage of loan recipients

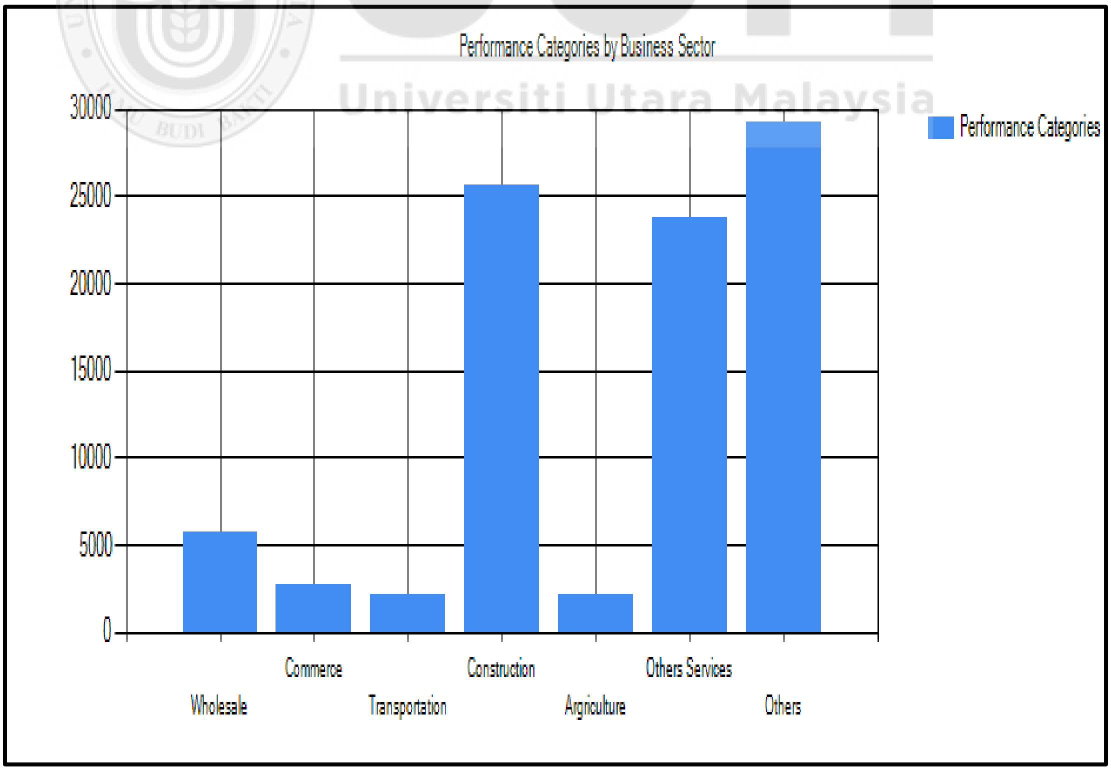


Figure 5.8. The performance categories by business sector

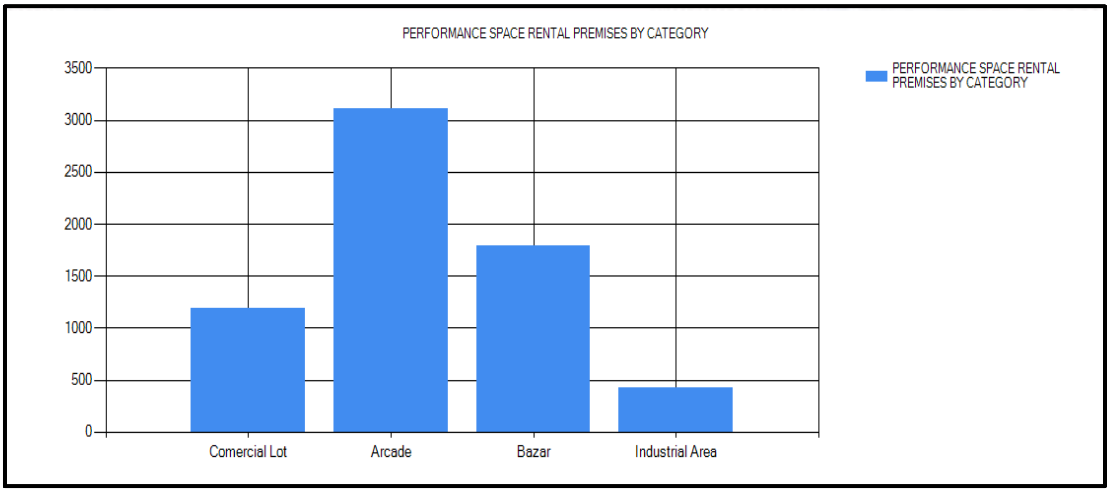


Figure 5.9. The performance space rental premises by category

Consequently, the procedure of creating and publishing of the organization reports will iteratively be applied in order to create and view the rest of BI reports based on the organization requirements.

As clearly shown in figures 5.7, 5.8 and 5.9, the organization reports are presented. These reports are reflecting the organization requirements. Moreover, the respondents generally expressed positive opinions about their experiences with these reports in terms of these reports are up-to-date and contain live data to support the Organization's decision-making process in near real-time, as extensively discussed in section 5.3.7. Moreover, this procedure was iteratively applied to create and view the rest of the reports based on the organization requirements.

Technically, it is difficult to join more than three database tables at the same time as this will result in very complex query statements. In order to solve this problem, the multi-virtual tables should be created to obtain the final virtual table that can be used to create the required BI reports. Refer to appendix G for related pseudocode.

Universiti Utara Malaysia

5.2.6 The Benefit

There are many benefits of applying the proposed DVDeM model in this organization. Mainly, the proposed model and its prototype can support decision making process in near real time data to data consumers, as well as the capability to add and delete data sources based on business needs without the needs to change the end user application. Furthermore, the model reduces the infrastructure and implementation cost.

5.2.7 The Prototype Evaluation

As discussed in Chapter 3, the usability testing for the prototype was conducted. The measurements were made through an instrument named Q-U, which comprises of six main dimensions, visibility, flexibility, learnability, application behavior, error control and help, and near real-time decision making as a usability attributes. This instrument is composed of 36 items spread over the six attributes (refer to Chapter 3 section 3.4.3.1.1). During the usability test, 30 respondents with computer science, software engineering, and IT background participated. The overall Q-U instrument is available in Appendix C. The respondents were required to answer questions after trying the DVDeM prototype.

Descriptive statistics was used. in order to describe the basic features of the data in this study, to provide summaries about the sample and the measures, together with graphic analysis, and to form the basis of virtually every quantitative analysis of data. The justification for using descriptive statistics is because it is useful to summarize the group of data using a combination of tabulated description (i.e., tables), graphical description (i.e., graphs and charts) and statistical commentary (i.e., a discussion of the results). Refer to appendix G for all descriptive statistics. Moreover, the findings are adequately discussed in the next sections.

5.2.7.1 The Visibility of DVDeM Prototype

As clearly indicated by the Figure 5.10 and Table 5.4 that all respondents agreed that the DVDeM prototype is able to extract useful information. Particularly, the majority of respondents (16.66 % strongly agree and 76.67% agree) confirm that DVDeM prototype can display the information (virtual tables, virtual data mart, and BI reports) in an uncluttered and well-structured manner; while 3.34% of them took the side of neutrality. Besides, all the respondents (6.66% and 93.34%) strongly agree and agree respectively that all DVDeM prototype instructions are visible and self-explanatory. Regarding the navigation options in the DVDeM prototype, the finding shows (13.33% strongly agree and 80% agree) of the respondents think that the navigation options such as (Links, shortcuts, home, back, forward, etc.) are displayed in a visible manner, while 6.67% of them took the side of neutrality. Likewise, all respondents (20% strongly agree and 80 % agree) found that DVDeM prototype has ability to communicate the status at all time (whether resting, processing etc.). Furthermore, the majority of respondents (13.33% strongly agree and 80% agree) affirmed that the data in DVDeM prototype is concisely presented, while 6.67% of them took the side of neutrality. Noteworthy, all respondents (23.33% strongly agree and 76.67% agree) state that the DVDeM prototype meets the user expectation in terms of functions and capabilities. Besides, (13.33 strongly agree and 70 % agree) that the DVDeM prototype has an accepted interface; while 16.67% of them took the side of neutrality. On the same aspect, (26.66% strongly agree and 60% agree) stated that the DVDeM prototype has a pleasant interface, while the remaining 13.34% took

Universiti Utara Malaysia

the side of neutrality. The majority of respondents (36.66% strongly agree and 53.34 % agree) found that the organization information in DVDeM prototype visualizes in a clear manner; while 10% of them took the side of neutrality. All the findings above are supported by the bar chart and statistics in Figure 5.10 and Table 5.4.

Table 5.4
The Visibility of DVDeM Prototype

Visibility Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Std. Deviation	Total
Item 1	N= 05 16.66%	N= 23 76.67%	N= 02 6.67%	N= 00 00.00%	N= 00 00.00%	3.100	0.481	**30**
Item 2	N= 02 06.66%	N= 28 93.34%	N= 00 00.00%	N= 00 00.00%	N= 00 00.00%	4.070	0.254	**30**
Item 3	N= 04 13.33%	N= 24 80.00%	N= 02 06.67%	N= 00 00.00%	N= 00 00.00%	4.070	0.450	**30**
Item 4	N= 06 20.00%	N= 24 80.00%	N= 00 00.00%	N= 00 00.00%	N= 00 00.00%	4.200	0.407	**30**
Item 5	N= 04 13.33%	N= 24 80.00%	N= 02 06.67%	N= 00 00.00%	N= 00 00.00%	4.070	0.450	**30**
Item 6	N= 07 23.33%	N= 23 76.67%	N= 00 00.00%	N= 00 00.00%	N= 00 00.00%	4.230	0.430	30
Item 7	N= 04 13.33%	N= 21 70.00%	N= 05 16.67%	N= 00 00.00%	N= 00 00.00%	3.970	0.565	**30**
Item 8	N= 08 26.66%	N= 18 60.00%	N= 04 13.34%	N= 00 00.00%	N= 00 00.00%	4.130	0.629	**30**
Item 9	N= 11 36.66%	N= 16 53.34%	N= 03 10.00%	N= 00 00.00%	N= 00 00.00%	4.270	0.640	**30**

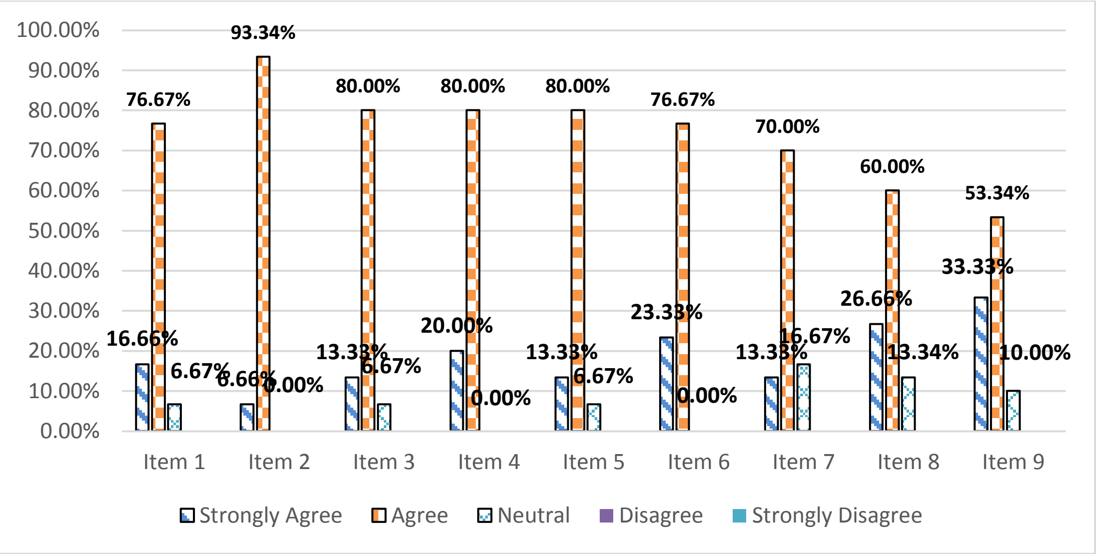
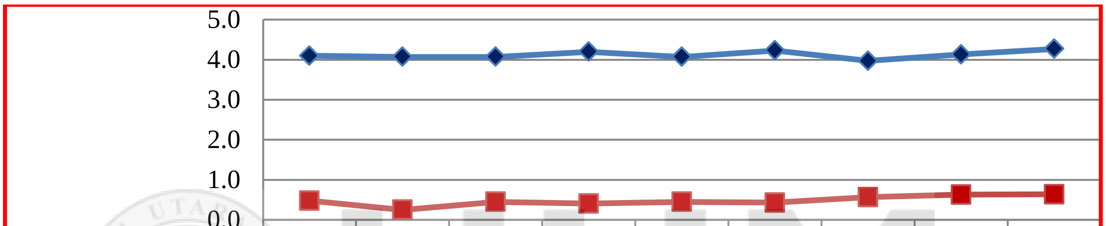


Figure 5.10. The Visibility of DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.11 shows the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have positive reaction about the DVDeM prototype usability in terms of visibility; and thereby strengthens the conclusion. The implication of this finding is that DVDeM and its prototype are visible. Besides, the standard deviation values for each measurement item are less than 0.7, indicating how close to the average the data is clustered.



	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9
Mean	4.100	4.070	4.070	4.200	4.070	4.230	3.970	4.130	4.270
Std. Deviation	0.481	0.254	0.450	0.407	0.450	0.430	0.565	0.629	0.640

Figure 5.11. The Descriptive Statistic of DVDeM Prototype Visibility

Universiti Utara Malaysia

5.2.7.2 The Flexibility of DVDeM Prototype

The results reveal that the DVDEM prototype is flexible. Particularly, the majority of the respondents (36.66% strongly agree and 46.67% agree) with feeling in control during usage of DVDeM prototype; while 16.67% tend to neutrality. Besides, the majority of the respondents (50.00% and 36.66%) strongly agree and agree respectively that the DVDeM prototype has customizable feature; while 13.34% of them preferred to be in neutral. Regarding the data entry in DVDeM prototype, the finding shows (23.33% strongly agree and 60% agree) of the respondents think that the design data entry is flexible; while 16.67% of them believe in stay in the neutral side. Likewise, the majority of the respondents (43.33% strongly agree and 43.33 %

agree) found that by using DVDeM prototype the data can be used, manipulated, and processed in an easy manner; while 13.34% of them believe in the neutral side. Eventually, most of the respondents (33.33% strongly agree and 53.33% agree) found that the DVDeM prototype can handle user-specified windows; while the remaining 13.34% took the side of neutrality. All aforementioned findings are supported by the bar chart and statistics in Figure 5.12 and Table 5.5. Respondents have affirmed that the DVDeM prototype is flexible.

Table 5.5
The Flexibility of DVDeM Prototype

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Std. Deviation	Total
Item 1	N= 11 36.66%	N=14 46.67%	N= 5 16.67%	N= 0 00.00%	N= 0 00.00%	3.230	0.728	**30**
Item 2	N= 15 50.00%	N= 11 36.66%	N= 4 13.34%	N= 0 00.00%	N= 0 00.00%	4.370	0.781	**30**
Item 3	N= 7 23.33%	N= 18 60%	N= 5 16.67%	N= 0 00.00%	N= 0 00.00%	4.070	0.640	**30**
Item 4	N= 13 43.33%	N= 13 43.33%	N= 4 13.34%	N= 0 00.00%	N= 0 00.00%	4.300	0.702	**30**
Item 5	N= 10 33.33%	N= 16 53.33%	N= 4 13.34%	N= 0 00.00%	N= 0 00.00%	4.200	0.664	**30**

Universiti Utara Malaysia

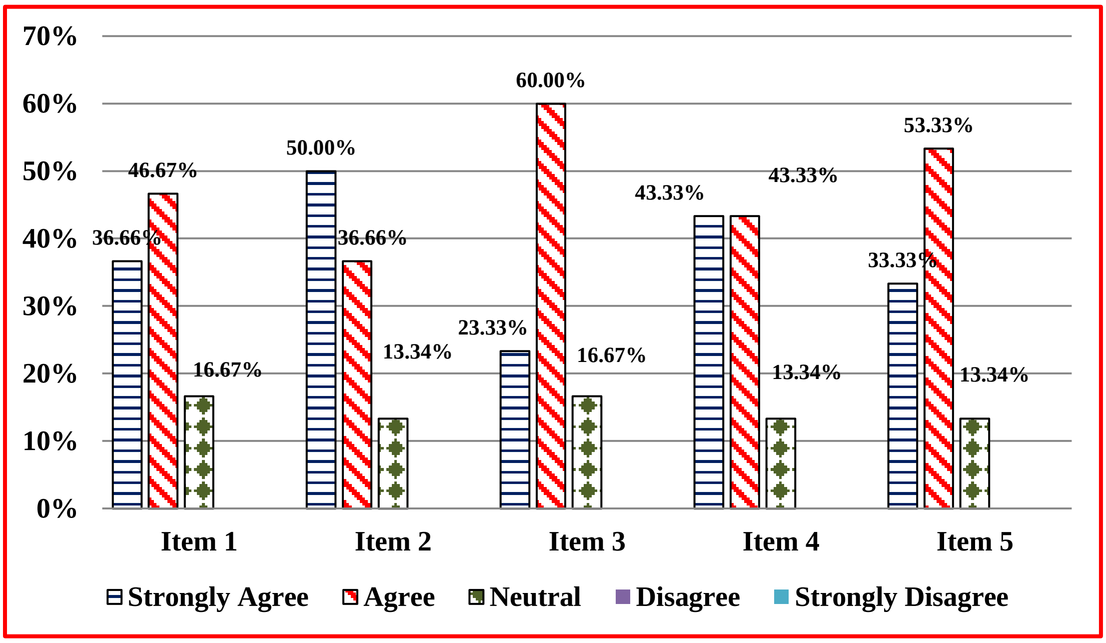
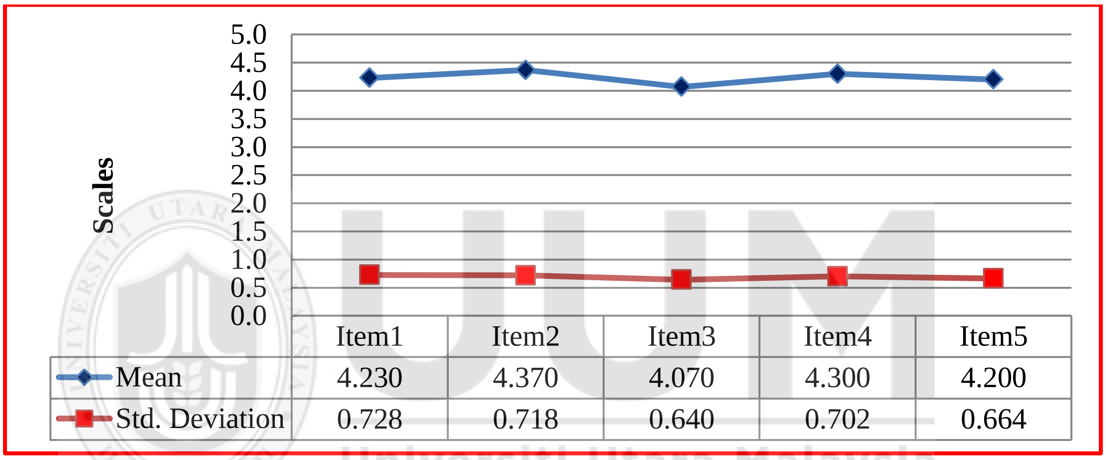


Figure 5.12. The Flexibility of DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.13 showed the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype usability in terms of flexibility; and thereby strengthens the conclusion. The implication of this finding is that DVDeM and its prototype are workable in practice in term of flexibility. Besides, the standard deviation values for each measurement item are less than 0.8, this indicates how close to the average the data is clustered.



Universiti Utara Malaysia

Figure 5.13. The Descriptive Statistic of DVDeM Prototype Flexibility

5.2.7.3 The Learnability of DVDeM Prototype

As clearly indicated by the Figure 5.14 and Table 5.6, the results reveal that the learnability is high. The majority of the respondents (16.66% strongly agree and 70% agree) found the data grouping of DVDeM prototype is reasonable for easy learning; while 13.33% tend to neutrality. Besides, 23.66% and 70% strongly agree and agree respectively, agree that the DVDeM prototype promotes learnability to make it accessible for infrequent usage; while 6.66% of them preferred to be in neutral. Regarding the grouping of menu options in DVDeM prototype, the finding shows most of the respondents (20% strongly agree and 73.33% agree) think that it is logical; while 6.66% tend to neutrality. Eventually, majority of respondents (20% strongly agree and 73.33 % agree) found that by using DVDeM prototype could become productive quickly; while 6.66% of them preferred to be in neutral. All aforementioned findings above are supported as shown in the bar chart and statistics in Figure 5.14 and Table 5.6.

Table 5.6
The Learnability of DVDeM Prototype

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Std. Deviation	Total
Item 1	N= 5 16.66%	N=21 70%	N= 4 13.34%	N= 0 00.00%	N= 0 00.00%	4.030	0.556	**30**
Item 2	N= 7 23.33%	N= 21 70%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.170	0.531	**30**
Item 3	N= 6 20.00%	N= 22 73.33%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.130	0.507	**30**
Item 4	N= 6 20.00%	N= 22 73.33%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.130	0.501	**30**

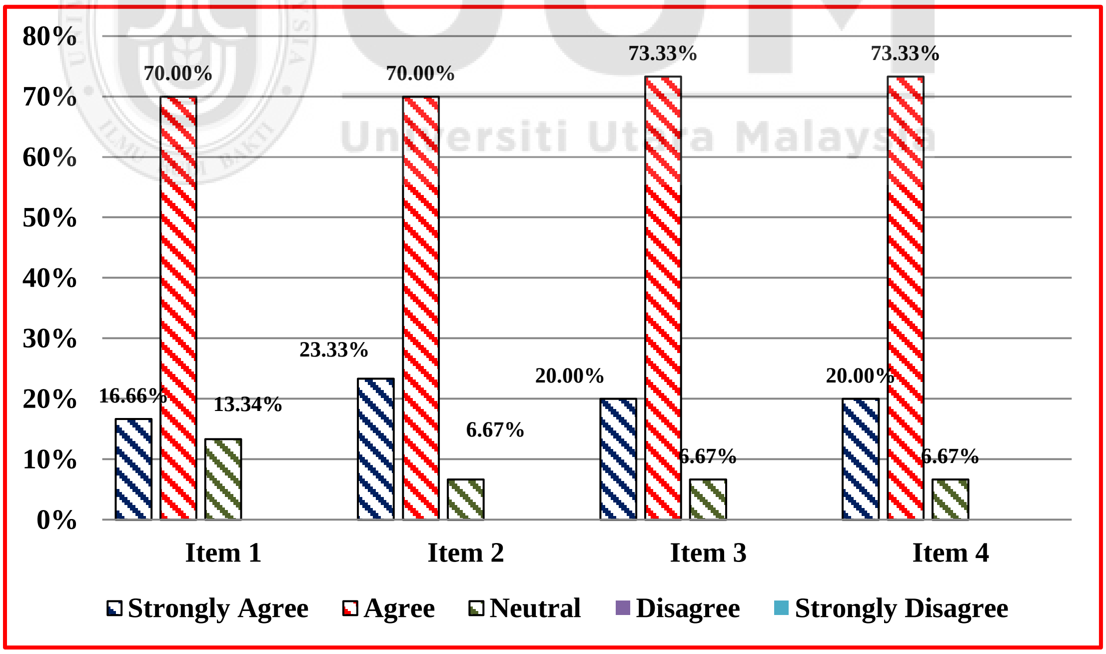


Figure 5.14. The Learnability of DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.15 showed the mean values for each measurement item are greater than 4 this indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype usability in terms of learnability; and thereby strengthens the conclusion. The implication of this finding is that DVDeM and its prototype are highly learnable. Besides, the standard deviation values for each measurement item are less than 0.6. This indicates how close to the average the data is clustered.

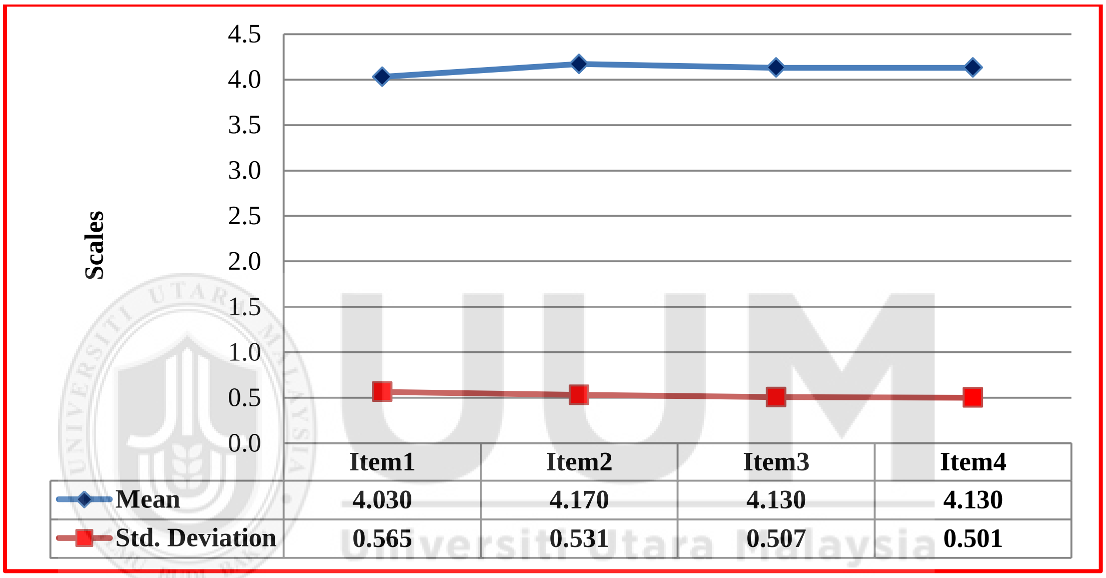


Figure 5.15. The Descriptive Statistic of DVDeM Prototype Learnability

5.2.7.4 The Application Behavior of DVDeM Prototype

The findings indicate that the respondents (13.33% strongly agree and 86.67% agree) that by using DVDeM prototype will enhance their efficiency through a consistently rapid response rate. Likewise, all respondents (26.66% and 70%) either strongly agree and agree respectively, believes that the DVDeM prototype is consistent, but 3.34% of them tend to neutrality. Besides, the majority of respondents (16.66% strongly agree and 80% agree) think that the information provides by DVDeM prototype was effective in helping them complete the tasks and scenarios; but 3.34%

of them preferred to be in neutral. In regard to the availability of information, the majority of respondents (26.66% strongly agree and 66.66% agree) think that by using DVDeM prototype it was easy to find, however, 6.67% of them preferred to be in neutral. When any mistake occurred, the majority of the respondents (23.33% strongly agree and 70% agree) believe that it can recover easily and quickly; further, 6.67% of them prefer to be on the neutral side. Also, the majority of the respondents (23.33% strongly agree and 70% agree) believe that they are able to complete the tasks and scenarios quickly via using DVDeM prototype, further, 6.67% of them prefer to be on the neutral side. Refer to Figure 5.16 and Table 5.7.

Table 5.7
The Application Behaviour of DVDeM Prototype

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Std. Deviation	Total
Item 1	N=5 16.66%	N=25 83.34%	N= 0 00.00%	N=0 00.00%	N=0 00.00%	4.170	0.379	**30**
Item 2	N= 8 26.66%	N= 21 70.00%	N= 1 03.34%	N= 0 00.00%	N= 0 00.00%	4.230	0.504	**30**
Item 3	N= 5 16.66%	N= 24 80.00%	N= 1 03.34%	N= 0 00.00%	N= 0 00.00%	4.130	0.434	**30**
Item 4	N= 8 26.66%	N= 20 66.66%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.200	0.551	**30**
Item 5	N= 7 23.33%	N= 21 70.00%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.170	0.531	**30**
Item 6	N= 7 23.33%	N= 21 70.00%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.170	0.531	**30**

Universiti Utara Malaysia

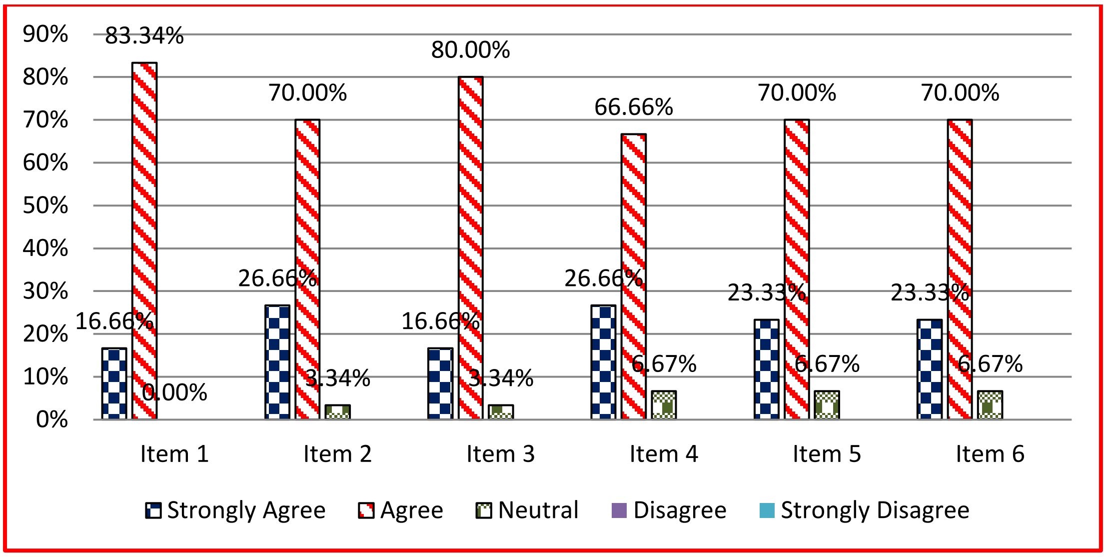


Figure 5.16. The Application Behaviour for DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.17 showed the mean values for each measurement item are greater than 4 this indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype usability in terms of application behavior; and thereby strengthens the conclusion. Besides, the standard deviation values for each measurement item are less than 0.6, which indicates how close to the average the data is clustered.

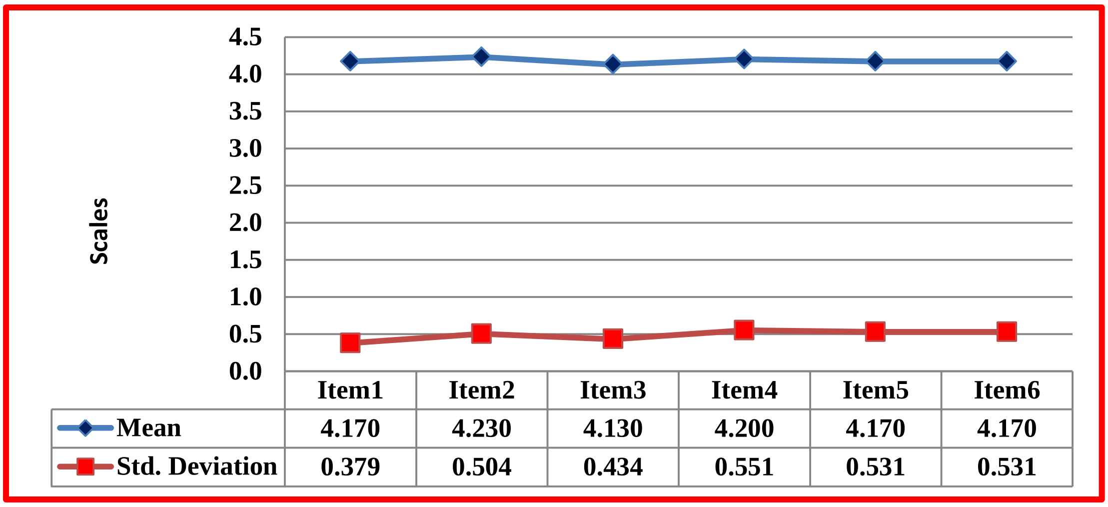


Figure 5.17. The Descriptive Statistics of DVDeM Prototype Application Behaviour

Universiti Utara Malaysia

5.2.7.5 The Error Control and Help of DVDeM Prototype

The majority of the respondents (30% strongly agree and 60% agree) think that DVDeM prototype has the ability for error prevention and error recovery; while 10% tend to neutrality. Besides, (40.00% and 50%) of them strongly agree and agree respectively think that DVDeM prototype provides online help and onscreen messages in a clear manner. Likewise, the majority of the respondents (40% strongly agree and 50 agree) agree that DVDeM prototype can give error messages that clearly inform what is the problem and how to fix. In terms of on-demand help, the majority of the respondents (26.66% strongly agree and 60% agree) found that DVDeM prototype provides such function, but 13.34% tend to neutrality. Noteworthy, the majority of the respondents (43.33 % for both strongly agree and agree) feel satisfied with DVDeM Prototype. The majority also (36.66% strongly agree and 50% agree), feel satisfied with how easy to use DVDeM prototype; but 13.34% tend to neutrality. Figure 5.18 and Table 5.8, has affirmed that DVDeM prototype provides error control and help.

Universiti Utara Malaysia

Table 5.8
The Error Control and Help for DVDeM Prototype

Items	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Mean**	**Std. Deviation**	**Total**
Item 1	N= 09 30.00%	N=18 60.00%	N= 03 10.00%	N=00 00.00%	N= 00 00.00%	4.200	0.610	**30**
Item 2	N= 12 40.00%	N= 16 50.00%	N= 02 10.00%	N=00 00.00%	N= 00 00.00%	4.300	0.651	**30**
Item 3	N= 12 40.00%	N= 15 50.00%	N= 03 10.00%	N= 00 00.00%	N= 00 00.00%	4.330	0.606	**30**
Item 4	N= 08 26.66%	N= 18 60.00%	N= 04 13.34%	N= 00 00.00%	N= 00 00.00%	4.130	0.629	**30**
Item 5	N= 13 43.33%	N= 13 43.33%	N= 04 13.34%	N= 00 00.00%	N= 00 00.00%	4.300	0.702	**30**
Item 6	N= 11 36.66%	N= 15 50.00%	N= 04 13.34%	N= 00 00.00%	N= 00 00.00%	4.230	0.679	**30**

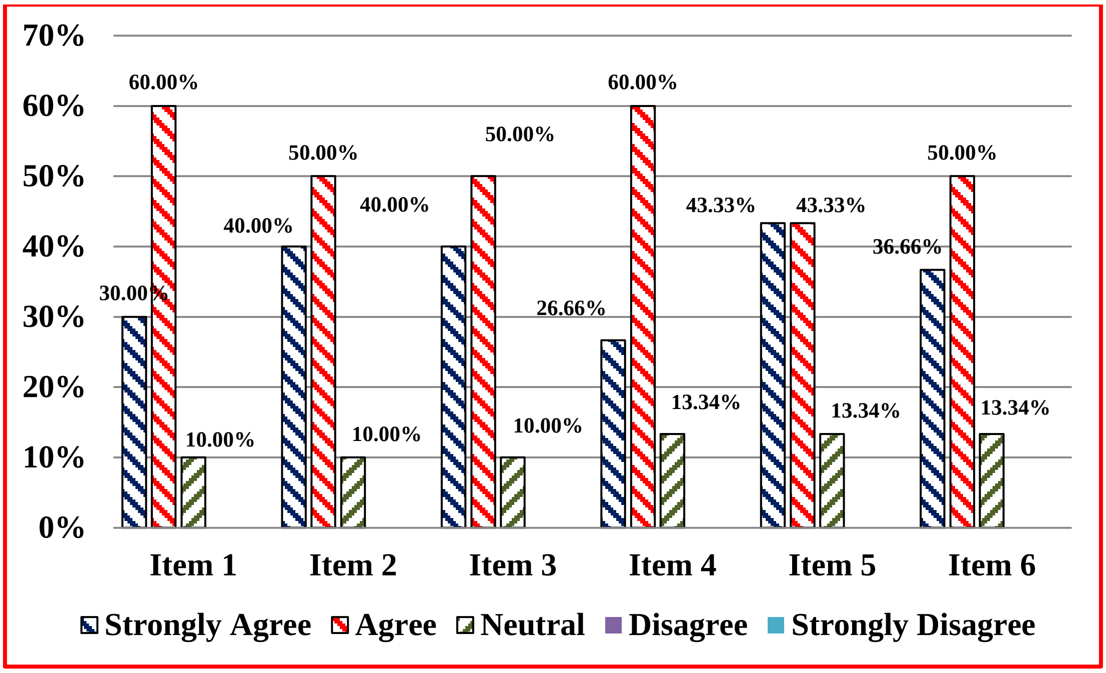
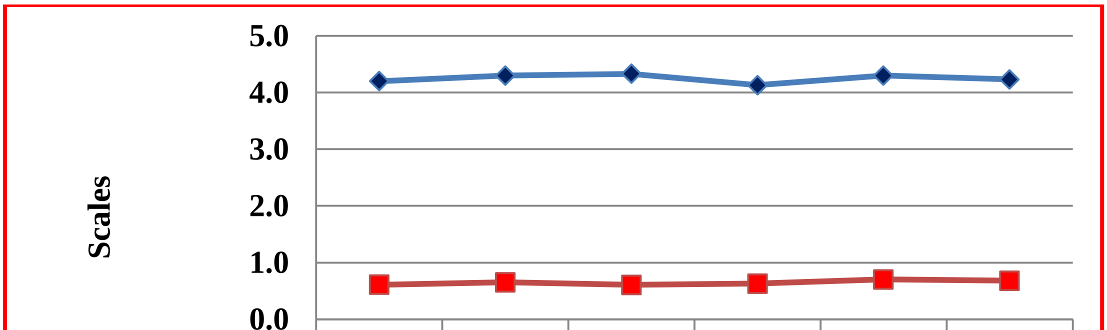


Figure 5.18. The Error Control and Help for DVDeM Prototype

UUM
Universiti Utara Malaysia

On the same aspects, the descriptive analysis presented in Figure 5.19 show the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype in terms of error control and help; and thereby strengthens the conclusion. Besides, the standard deviation values for each measurement item are less than 0.8, which indicates how close to the average the data is clustered.



	Item1	Item2	Item3	Item4	Item5	Item6
Mean	4.200	4.300	4.330	4.130	4.300	4.230
Std. Deviation	0.61	0.651	0.606	0.629	0.702	0.679

Figure 5.19. The Descriptive Statistic for Error Control and Help

5.2.7.6 Near Real-Time Decision Making of DVDeM Prototype

As clearly indicated by the results visualize in Table 5.9, that the DVDeM prototype is able to provide near real-time data for decision-making. Particularly, the majority of the respondents (40% strongly agree and 53.33% agree) agree that the knowledge sharing is allowed in DVDeM prototype; while 6.67% tend to neutrality. Likewise, most of the respondents (26.66% and 63.33%) strongly agree and agree respectively, that DVDeM prototype assists the decision-making process by providing the information visualization functionality (comparison charts, graphs in revealing trends etc.); but 6.67% tend to neutrality. Besides, the majority of the respondents (43.33% strongly agree and 50% agree) believe that DVDeM prototype provides sufficient coverage for all data resources; while 6.67% tend to neutrality. Furthermore, all the respondents (20% strongly agree and 80% agree) affirm that DVDeM prototype provides data on time to take suitable actions and decisions. likewise, all the respondents (36.66% strongly agree and 63.34% agree) affirm that the data provided by DVDeM prototype are always live data and up to date, which means DVDeM proposed model and its prototype provides fully supported for near real-time decision-making. The majority of them (26.66 % strongly agree and 63.34% agree) found that DVDeM prototype provides high-level data quality; however, 10% tend to neutrality. Refer to Figure 5.20 and Table 5.9, which affirmed that the DVDeM prototype is believe to provide near real-time data for decision making.

Universiti Utara Malaysia

Table 5.9
The Near Real-Time Decision Making for DVDeM Prototype

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Std. Deviation	Total
Item 1	N= 12 40.00%	N= 16 53.33%	N= 02 06.67%	N= 00 00.00%	N= 00 00.00%	4.330	0.606	**30**
Item 2	N= 08 26.66%	N=20 63.34%	N= 03 10.00%	N= 00 00.00%	N= 00 00.00%	4.200	0.551	**30**
Item 3	N= 13 43.33%	N=15 50.00%	N= 02 06.66%	N= 00 00.00%	N= 00 00.00%	4.370	0.615	**30**
Item 4	N= 06 20.00%	N=24 80.00%	N= 00 00.00%	N= 00 00.00%	N= 00 00.00%	4.200	0.407	**30**
Item 5	N= 11 36.66%	N=19 63.34%	N= 00 00.00%	N= 00 00.00%	N= 00 00.00%	4.370	0.490	**30**
Item 6	N= 08 26.66%	N=19 63.34%	N= 03 10.00%	N= 00 00.00%	N= 00 00.00%	4.170	0.592	**30**

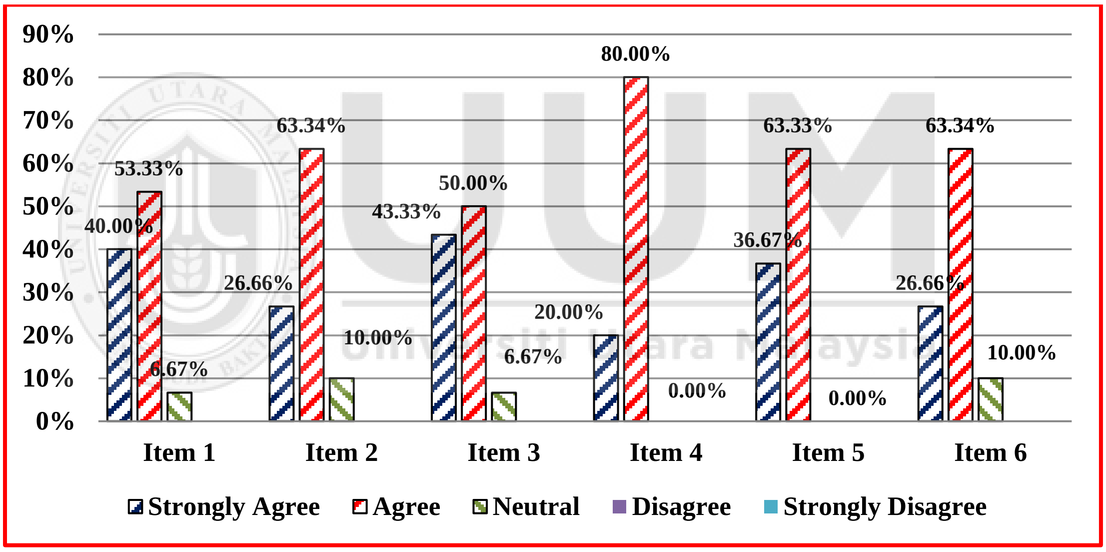
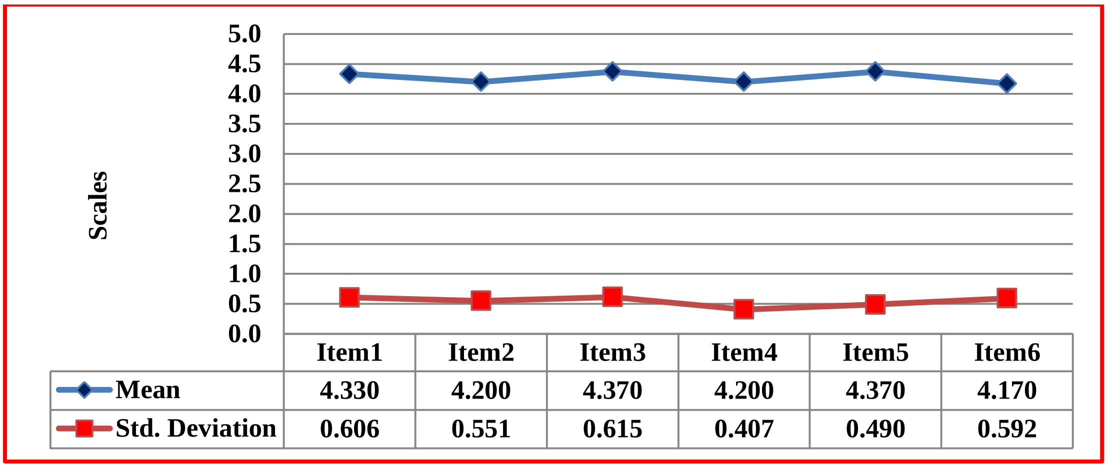


Figure 5.20. Near Real-Time Decision Making for DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.21 showed the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype usability in terms of near real-time decision making, and thereby strengthens the conclusion. Besides, the standard deviation values for each measurement item are less than 0.7, which indicates how close to the average the data is clustered.



	Item1	Item2	Item3	Item4	Item5	Item6
Mean	4.330	4.200	4.370	4.200	4.370	4.170
Std. Deviation	0.606	0.551	0.615	0.407	0.490	0.592

Figure 5.21. The Descriptive Statistic for Near Real-Time Decision Making

5.4.7.7 Overall Usability Finding

In order to obtain the overall finding, the average of findings for each usability attribute was calculated. The overall usability is made of six dimensions as depicted in Table 5.10 and Figure 5.22. Calculating the average agreement value for each dimension, the figure show that for each dimension, the majority of respondents agree that the DVDeM prototype provide data visibility, is flexible, easy to learn, behave as expected, provide necessary error control and help, and data to allow near real time decision making. Overall, on total average 91.42 %of the respondents agree to this.

Table 5.10
The Overall Usability Finding

	Usability Dimension	Strongly Agree + Agree
1	Visibility	92.98%
2	Flexibility	85.83%
3	Learnability	91.96%
4	Application Behavior	95.55%
5	Error Control & Help	88.33%
6	Near Real-Time Decision Making	93.88%
	The Average	**91.42%**

Universiti Utara Malaysia

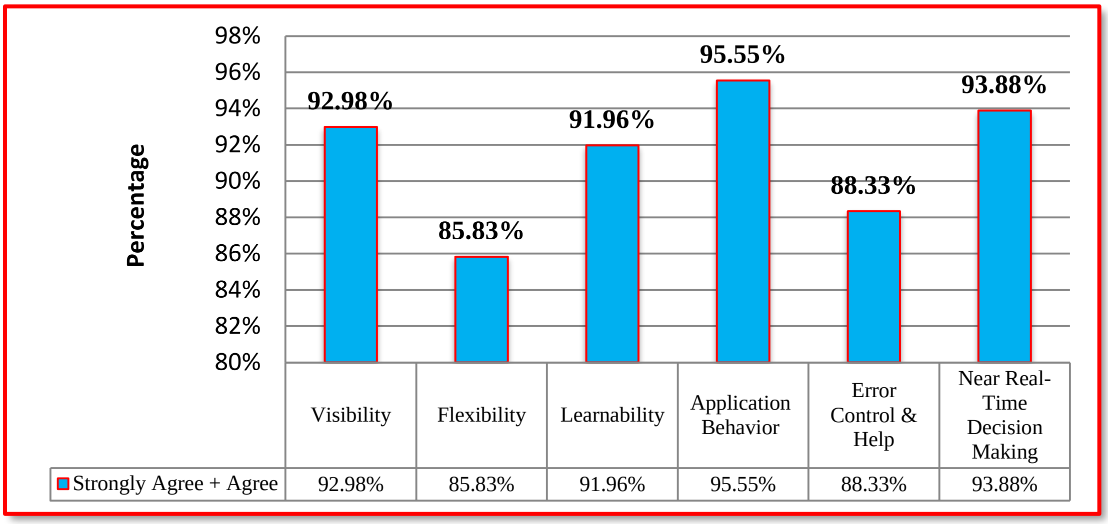


Figure 5.22. The Overall Usability Finding

5.3 Prototype Testing in Education Sector

As mentioned earlier and for the purpose of providing a complete and rich implementation for case studies, the case studies profile should include: organization background, business problem, existing application, the proposed DV solution, the implementation process, and the benefits. In the next sections, the details about the implementation of DVDeM prototype in the education sector were discussed.

5.3.1 The Organization's Background

Institutions of higher education have started developing web portal for student's use. Almost every institution had its own Learning Management System (LMS) for their students. LMS is a vital tool that can be used to support teaching and learning process. To date many institutions of higher learning have employed LMS as a platform to manage their e-learning program. Typically, various tasks can be performed through LMS such as delivering and managing the course materials as well as manage the administration of online teaching and learning activities. Managing online courses, registering students, tracking student progress and assessments are all parts of managing learning. Additionally, LMS provides a complete e-Learning infrastructure for creating, managing, tracking, delivering, and assessing eLearning, whether in the classroom or online. The successful of LMS is not only in having great functions of the system, but it also needs to be well designed and has usable user interface. User interface acts like a medium when user interacts with the system. A good design of user interface is required to make user feel comfortable and enjoyable while using the LMS.

5.3.2 The Organization Business Problem

Blended learning is one of LMS parts, it is an education program (formal or informal) that combines online digital media with traditional classroom methods. It requires the physical presence of both teacher and student, with some element of student control over time, place, path, or pace. However, the blended learning system was built with physical data integration, which lacks in live data, therefore, the reports being produced do not show real time or near real time data. This could implicate negatively the decision-making process of the organization. In addition, the current system also does not meet the needs of the organization.

5.3.3 The Organization Existing Business Intelligence Application

In this organization, the existing blended learning system is based on physical data integration, accordingly, the data which is delivered to the LMS portal is lacking in updating and not near real-time data; due to the difficulties of physical data integration technique as extensively mentioned in the problem domain of this study. The LMS existing architecture of this organization is depicted in Figure 5.23.

Universiti Utara Malaysia

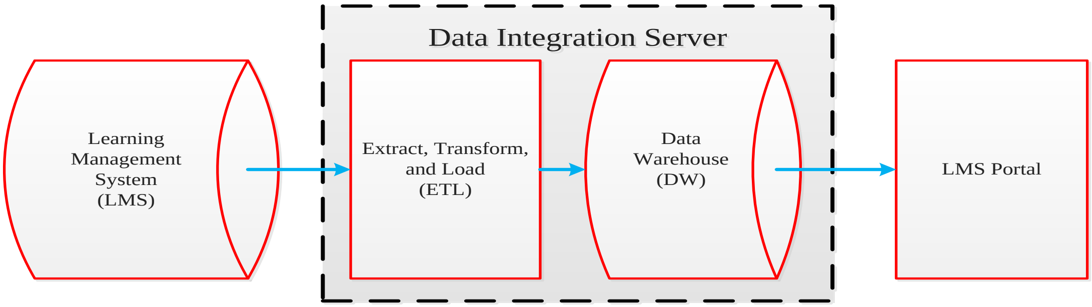


Figure 5.23. The Organization Existing Business Intelligence Application

As clearly shown in Figure 5.23, the data were extracted, transformed, and loaded from the learning management system via ETL to the subjected data (DW) which is located inside the organization BI system. Due to this, data are nor regularly updated which could not support decision making process in time.

5.3.4 The Data Virtualization Proposed Solution

The proposed DV solution of this organization was based on DVDeM proposed model, as illustrated in Figure 5.24 and extensively explained in the next sections.

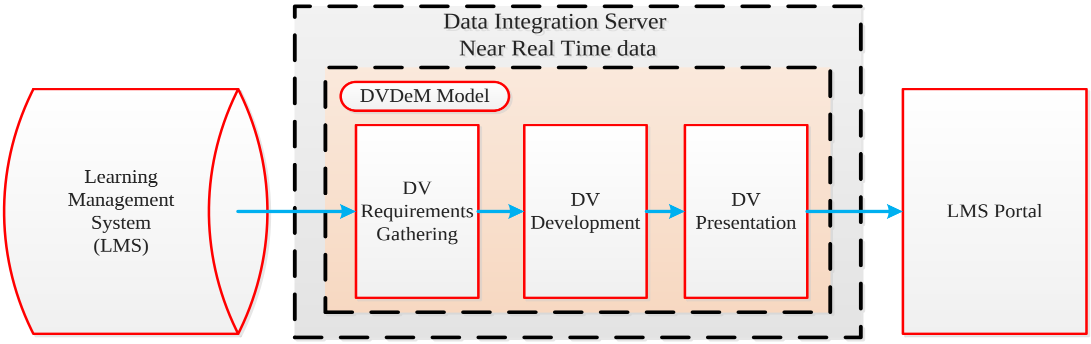


Figure 5.24. DV Proposed Solution

5.3.5 The Implementation Process

In this section, the DV solution of this organization was implemented based on the proposed DVDeM model. The implementation was done based on all phases involved in the proposed DVDeM model phases starting with DV requirement gathering phase through DV development phase and ends with DV presentation phase. Details of the implementation process of DV solution are highlighted in the next section.

5.3.5.1 DV Requirements Gathering

As mentioned in chapter 4, requirements gathering phase is the foundation of the entire software development project. Hence, the organization requirements must be clear, correct and well-defined. During this phase, a full understanding of the organization's requirements and business needs is essential as well as understanding the data sources (the location, type, and the verifying data quality issues). In addition to identifying the system requirements, a list of the required resources (hardware and software) to implement this for a prototype is prepared. Therefore, in the context of this study, the main concern of this organization is to provide near real-time LMS portal. Furthermore, the main requirements of LMS are listed.

i. Which academic staff is well-trained to use technology for teaching?
ii. Which academic staff is well-trained in pedagogical teaching?
iii. Are the courses taught by academic staff obtained more than 75% of marks for individually evaluated courses per session?
iv. How many academic staff can develop and conduct the students' courses through the online learning platform?
v. Which academic staff can develop and conduct the students' courses through the online learning platform?

In line with above situations, the DV solution based on DVDeM proposed model was developed and implemented for this organization to meet these requirements. The prototype also can deal with the other requirements based on business needs. As

Universiti Utara Malaysia

mentioned in chapter 4, the DV development of the proposed DVDeM model relies on the outputs of the requirement gathering and analysis.

5.3.5.2 DV Development

Based on the DVDeM proposed model, the first task to DV development is to understand the data sources (raw data which is data that has not been processed for use.

The raw data understanding starts with an initial data collection and proceeds with activities to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information. Accordingly, there is a close link between business understanding and data understanding. The formulation of the data integration problem and the project plan require at least some understanding of the available data. In the context of this case study, the raw data for blended learning system consists of a lot of attributes spread over so many tables. Table 5.11 lists the raw data for the blended learning system.

Table 5.11
The Raw Data for Blended Learning System

	Table Name	Attributes	Description
1	Staff	✓ Staff-ID. ✓ Staff-Name. ✓ Staff-Age. ✓ Position. ✓ First appointment Data as an Academic Staff. ✓ Service end date. ✓ Date of birth. ✓ Cohort. ✓ Academic Qualification. ✓ Professional qualification. ✓ Faculty. ✓ Field.	

		✓ Status ✓ Date start. ✓ Date end. ✓ Citizenship.	
2	Course	✓ JumMod. ✓ Module. ✓ Course-ID. ✓ Course-short. ✓ Course-full. ✓ Staff-Name ✓ Staff-ID. ✓ Lecturer-School-ID ✓ Lecturer-School	Near real time data (Module) ✓ Information (Label and Syllabus) ✓ Resources (Book, Folder, Page, Resource, and URL). ✓ Activities (Chat, Choice, feedback, Forum, Glossary, and Lessen) ✓ Assessment (Assignment and quiz).
3	School	✓ School-ID. ✓ School-Name ✓ College-Name ✓ Course-ID. ✓ Staff-ID ✓ Staff-Name ✓ Lecturer-School-ID ✓ Lecturer-School ✓ Sim-ID ✓ Sim Duration	

Universiti Utara Malaysia

As clearly indicated in Table 5.11 and in the context of the blended learning system, the related data are changing dynamically, hence, any delay of data updating will give a negative effect on the decision-making process. Accordingly, the near time data used in the blended learning system is (i) Information which includes (Label and Syllabus), (ii) Resources which includes (Book, Folder, Page, Resource, and URL), (iii) Activities which includes (Chat, Choice, Feedback, Forum, Glossary, and Lesson), and (iv) Assessment which includes (Assignment and quiz) as well as other relevant data. Furthermore, the raw data of the blended learning is needed many activities to construct the final dataset (data that will be fed into the virtual table(s)).

As mentioned earlier, the data sources to be used to create virtual tables is not always correct. For instance, names are spelled incorrectly, numeric values are outside realistic boundaries, values in two fields have accidentally been switched, stored values don't represent reality, and particular values or rows are completely missing. If no actions are taken, this incorrect data is presented by the data virtualization server to the data consumers. The consequence is that incorrect data is used for decision-making. The quality of business decisions is for a large part dependent on the quality of the data. Therefore, data quality is an important consideration when designing business intelligence systems.

With respect to the blended learning system, there are many incorrect data such as some tables mentioned the long school name, while the other tables put an abbreviation name; in this case, the school's names must be united. Another correction regarding removed null data. The SQL statements which used for verifying data quality issues are listed below.

i. Remove Null values

```
DELETE FROM Table-Name
WHERE Column-Name IS NULL;
```

ii. Column value united

```
UPDATE Table-Name
SET column1 = value1 WHERE condition;
```

Based on the DVDeM proposed model, after verifying data quality the raw data is ready to be imported. Importing a source tables means that it is made known to the data virtualization server, accordingly, the imported data sources for the blended learning system consists of three tables (Staff, School, and Courses) as well as their related attributes, to be used later to create wrapper table as highlighted in the next paragraph.

A wrapper table for blended learning system has been defined on the source table, and it consists of all imported data sources attributed as shown in Table 5.12.

Universiti Utara Malaysia

Table 5.12
The Wrapper Table for Blended Learning System

Table Name	Attributes	Description
Blended Learning Wrapper	✓ Staff-ID. ✓ Staff-Name. ✓ Staff-Age. ✓ Position. ✓ First appointment Data as an Academic Staff. ✓ Service end date. ✓ Date of birth. ✓ Cohort. ✓ Academic Qualification. ✓ Professional qualification. ✓ Faculty. ✓ Field. ✓ Status ✓ Date start. ✓ Date end. ✓ Citizenship. ✓ JumMod. ✓ Module. ✓ Course-ID. ✓ Course-short. ✓ Course-full. ✓ Staff-Name ✓ Lecturer-School-ID ✓ Lecturer-School ✓ School-ID. ✓ School-Name ✓ College-Name ✓ Course-ID. ✓ Lecturer-School-ID ✓ Lecturer-School ✓ Sim-ID Sim Duration	Near real time data (Module) ✓ Information {Label and Syllabus} ✓ Resources {Book, Folder, Page, Resource, and URL}. ✓ Activities {Chat, Choice, feedback, Forum, Glossary, and Lessen} Assessment {Assignment and quiz}.

Universiti Utara Malaysia

Consequently, the SQL statements which is used for creating wrapper table are listed below.

```
CREATE VIEW [dbo].[Blended Learning Wrapper]
AS SELECT   dbo.School.*, dbo.STAFF$.*, dbo.Student.*
FROM        dbo.School CROSS JOIN dbo.STAFF CROSS JOIN dbo.Course
GO
```

As indicated from Table 5.12, the wrapper table shows the full contents of a source table. Also, a wrapper has the same structure as the source table it is bound to. in the context of the blended learning system, not all the data to be used via data consumers, therefore, the virtual table should contain the only related data. Furthermore, Wrapper table also manipulates the data sources to transform the data values to standard forms to be an input for virtual table. The SQL statements which is used for creating the blended learning virtual table are detailed below.

```
CREATE VIEW [dbo].[Blended Learning Virtual Table]
AS  SELECT  *  from   [dbo].[Blended  Learning  Wrapper]  where  Condition1
  ondition2,etc.
GO
```

Universiti Utara Malaysia

Based on the DVDeM proposed model and in order to reduce the size of the virtual table and to make it consists of relevant data only, a goal-oriented approach based on data virtualization (GODV) was applied. the details regarding GODV are highlighted in the next paragraphs.

Applying GODV starts with the main goal of this obtained from the analysis the mission of the organization, hence, one of the main goals obtained is “to become a leading management organization”. Based on GODV, the main goal can be further divided into four sub-goals which are “to be the center of excellence in research”, “to be the center of excellence in publishing”, “to be the center of excellence in consulting", and "to be the center of excellence in teaching and learning" as shown in Figure 5.25.

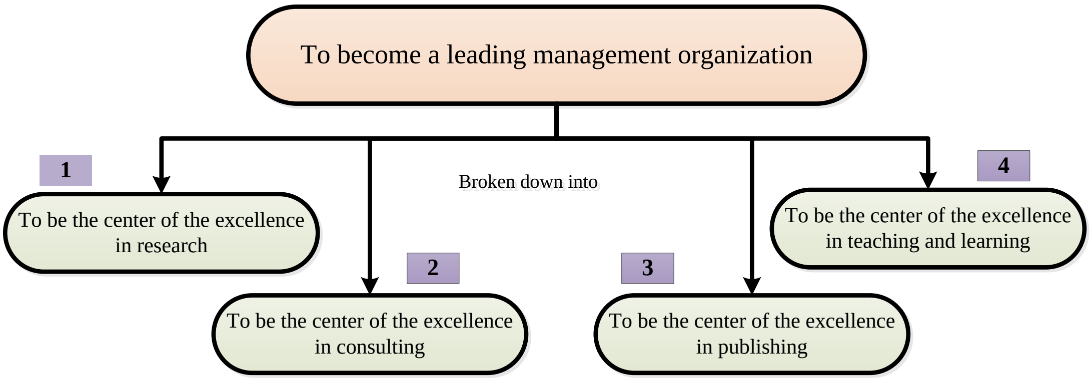


Figure 5.25. The Organization Main Goals Diagram

Based on the scope of this study, which is data warehouse schema for monitoring university teaching and learning KPIs, the researcher's investigation identified that one of the departments given the responsibility of meeting the needs of technological innovation in teaching and learning is the University Teaching and Learning Centre (UTLC; with the motto: "Transforming Teaching, Advancing Learning"). The main goal of UTLC is "To transform teaching and advance learning." This goal can further be decomposed into sub-goals as listed below:

A. To become a leading management organization.

A.1. To be the center of excellence in publishing.

A.2. To be the center of excellence in consulting.

A.3. To be the center of excellence in research.

A.4 To be the center of excellence in teaching and learning.

A.4.1 To deliver excellent learning experience through innovative teaching.

A.4.2 To ensure that RM100k of grant per year is given out to the selected of SOTL grants.

Universiti Utara Malaysia

A.4.3 To ensure that all staff can develop and conduct the students' courses through the online learning platform.

A.4.4 To ensure that the academic staff are well-trained to use technology for teaching.

A.4.5 To ensure that the academic staff are well-trained in pedagogical teaching.

A.4.5.1 To ensure availability of more than 25 training using Web 2.0 tool.

A.4.5.2 To ensure all academic staff acquires training using online training platform.

A.4.5.3 To ensure availability of more than 25 training program on technology every year .

A.4.6 To ensure that all the courses taught shall obtain more than 75% of marks for individually evaluated courses per session.

A.4.7 To select at least one eligible academic staff to receive award.

A.4.7.1 To ensure at least that one academic staff receives DTA award.

A.4.7.2 To ensure at least that all academic staff acquires training using web 2.0.

A.4.7.3 To ensure there are one academic staff who wins an AAN award.

In line with above situations, Figure 5.26 shows the goal decomposition tree for LMS goals and broken down into sub-goals that are related to teaching and learning that managed by the intended organization.

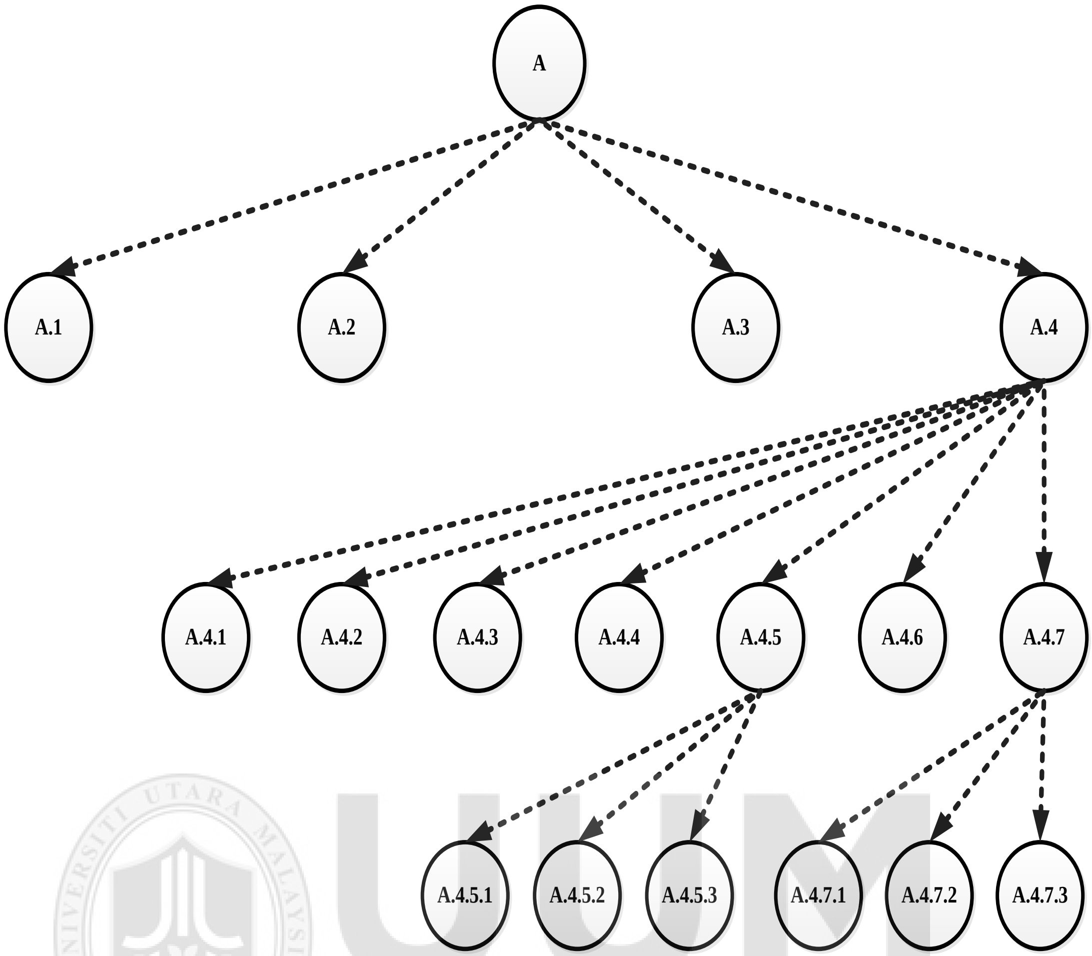


Figure 5.26. The goal Decomposition Tree for LMS Goals (sub-goals)

Universiti Utara Malaysia

As mentioned in chapter 4 and based on GODV approach, the process of breaking down the goals into sub-goals will continue iteratively (sometimes two or more goals are combined using AND / OR operators) until it reaches the leaf- goal. Figure 5.27 shows the goal decomposition tree with leaf-goals.

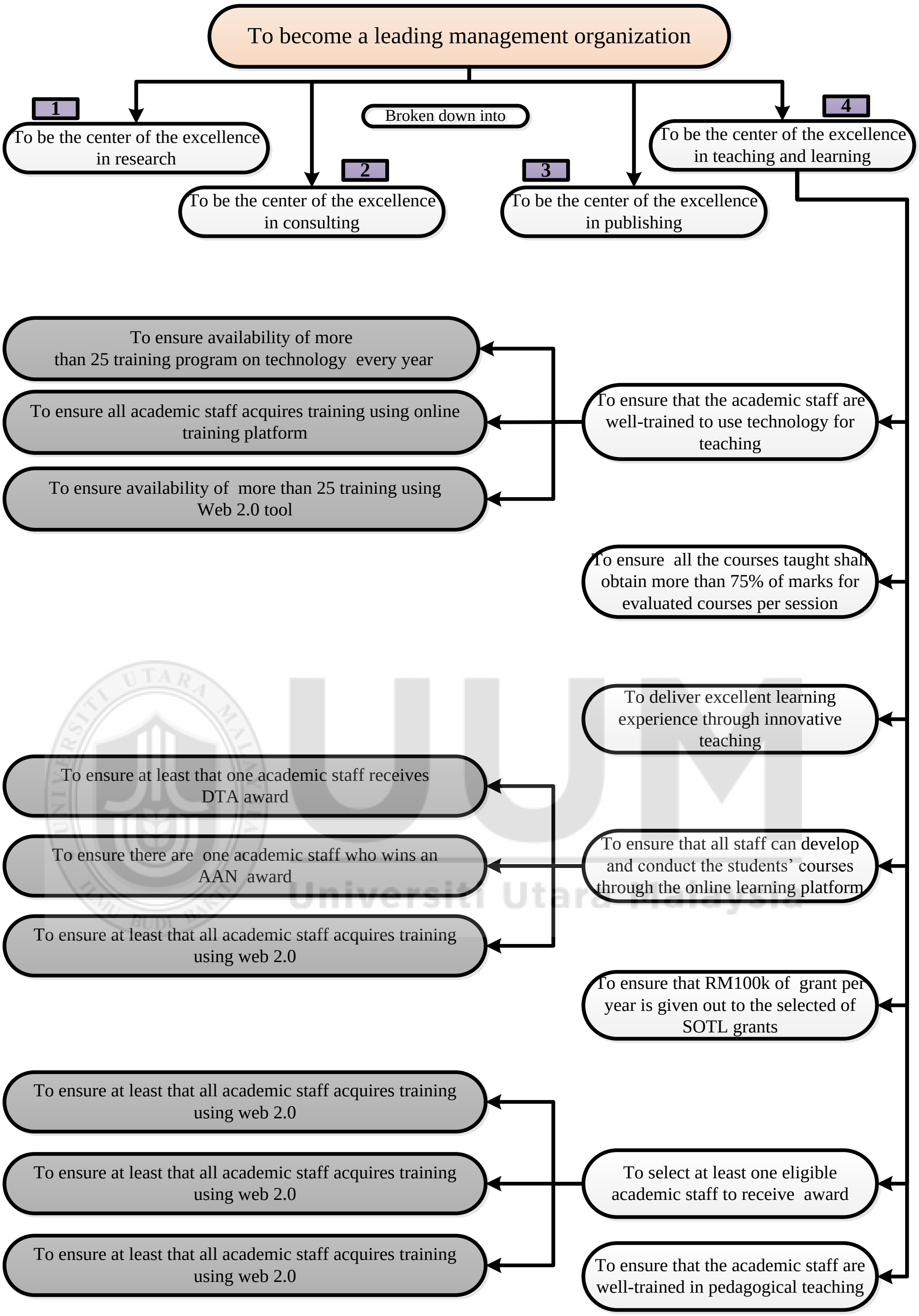


Figure 5.27. Goal Decomposition Tree with Leaf-goals

As clearly indicated in the Figure 5.27, the leaf-goals are represented by the shaded area.

As mentioned in chapter 4, the core purpose behind the proposed goal oriented approach based on data virtualization is to create a virtual table which consists of relevant data only. Therefore, the attributes analysis will be used to describe the relationship between the measured attributes and the leaf goals. The attributes abstracted in this case study and their respective descriptions are listed in Table 5.13.

Table 5.13
The Relationship between Leaf goals and Measured Attributes

	The Leaf Goal	**The Measured Attribute**	**Attribute Description**
1	To ensure that the academic staff are well-trained to use technology for teaching.	Technology Training	The training that focuses on the use of learning and teaching technologies like E-cEvas, Web 2.0 etc.
2	To ensure that the academic staff are well-trained in pedagogical teaching.	Pedagogy Training.	The training that focuses on the method of teaching and instructional deliveries.
3	To ensure that all the courses taught by academic staff shall obtain more than 75% of marks for individually evaluated courses per session.	Course Evaluation	The evaluation of courses taken by the organizations.
4	To select at least one eligible academic staff to receive awards.	Awardees Selection	The selection of the eligible persons to be awarded.
5	To ensure that RM100k of grant per year is given out to the selected of SOTL grants.	Grant Allocation	The allocation of grants to selected faculty members.
6	To ensure that the academic staffs are able to develop and conduct the students' courses through the online learning platform.	Blended Learning	The use of both online and traditional face-to-face teaching as learning processes.

As clearly indicated from the Table 5.13, relevant attributes were determined and extracted; and accordingly, these extracted attributes are used to match with the data sources (wrappers/ original virtual table) to create the virtual table based on relevant data. The main difference between the virtual table based on GODV and existing virtual table is the resulting table (the virtual table based on GODV) contains the relevant data only. In a nutshell, the obtained relevant data is based on the analysis of the organization goals and broken down into sub goals in iteratively manner, this process will continue until reaching to the leaf goals which are used to match the original data sources to extract the relevant data only. Figure 5.28 illustrates GODV approach for all levels (analysis one goal only).

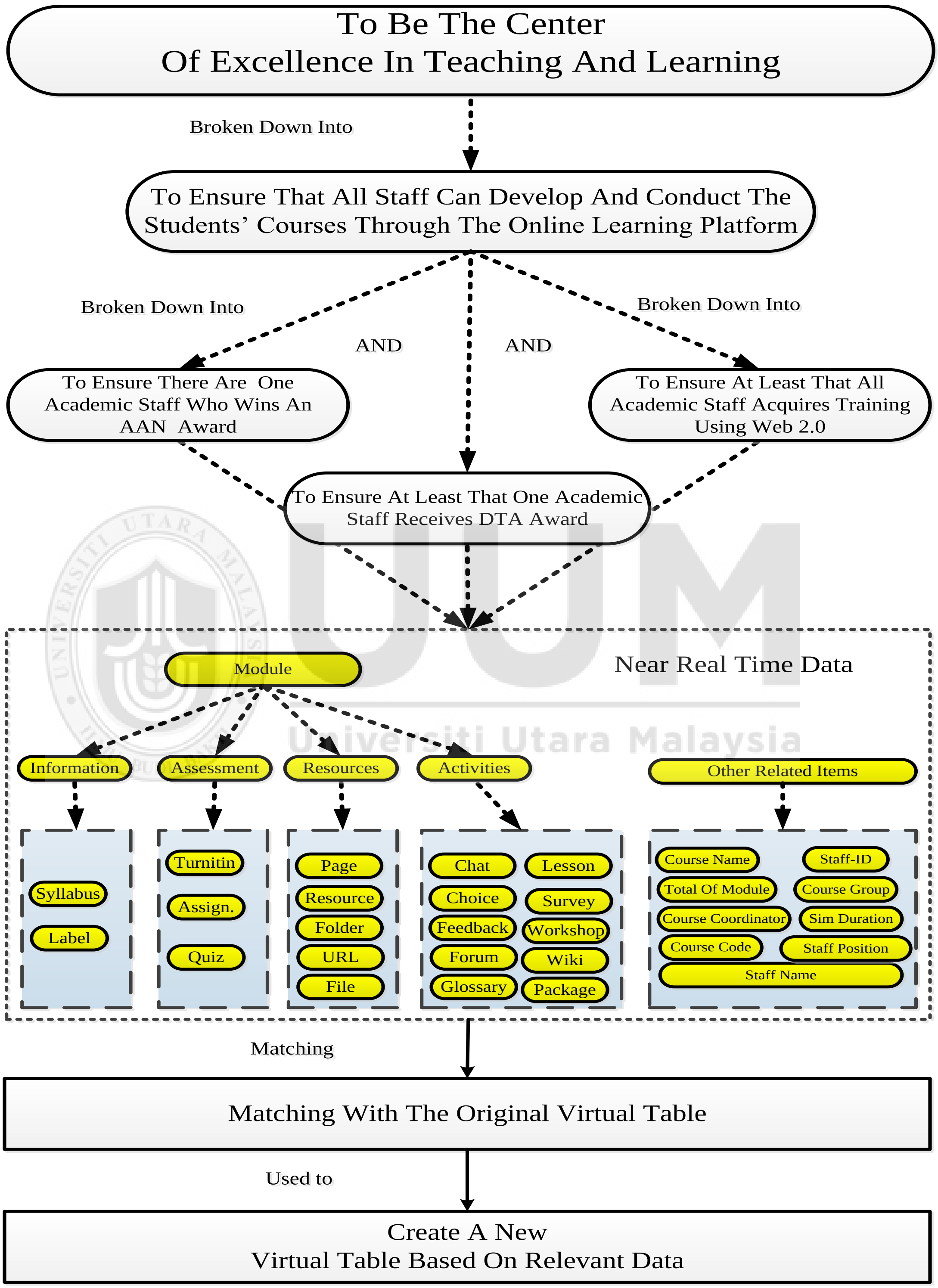


Figure 5.28. Applying GODV Approach (all levels)

As clearly indicated in the Figure 5.28, the findings from the goal analysis using GODV are many indicators which are used later to match with the original virtual table to create a new virtual table based on relevant data.

Furthermore, the procedure to create and present the required reports require creating multi-level virtual tables to obtain data; usually, some near real-time data such as (Information, Activities, Resources, and Assessment) are resulted from conducting mathematic operations and aggregation on other data. The SQL statements which used for preparing near real time data for blended learning are detailed below.

Preparing near real time data (Information):

```
SELECT ISNULL (SUM([JumMod]),0) AS NoOfModules from ['RAW DATA$'] where
(([Module]='Syllabus' or [module]='Label'or [module]='')and( [StaffNo]='"
+ textBox1.Text + "' ))");
```

Preparing near real time data (Resources):

```
SELECT ISNULL(SUM([JumMod]),0) AS NoOfModules from ['RAW DATA$'] where
(([StaffNo]='" + textBox1.Text + "') and (([Module]='File') or
([Module]='Folder') or ([Module]='Page') or ([Module]='URL')or
([Module]='resource'))) ");
```

Preparing near real time data (Activities):

```
SELECT ISNULL(SUM([JumMod]),0) AS NoOfModules from ['RAW DATA$'] where
(( [StaffNo]='" + textBox1.Text + "') and (([Module]='chat') or
([Module]='choice') or ([Module]='feedback')or ([Module]='Forum')or
([Module]='Glossary')or ([Module]='Lesson')or ([Module]='Survey')or
([Module]='Scorn Package')or ([Module]='Workshop')or ([Module]='Wiki')or
([Module]='Survey'))) ");
```

Preparing near real time data (Assessment):

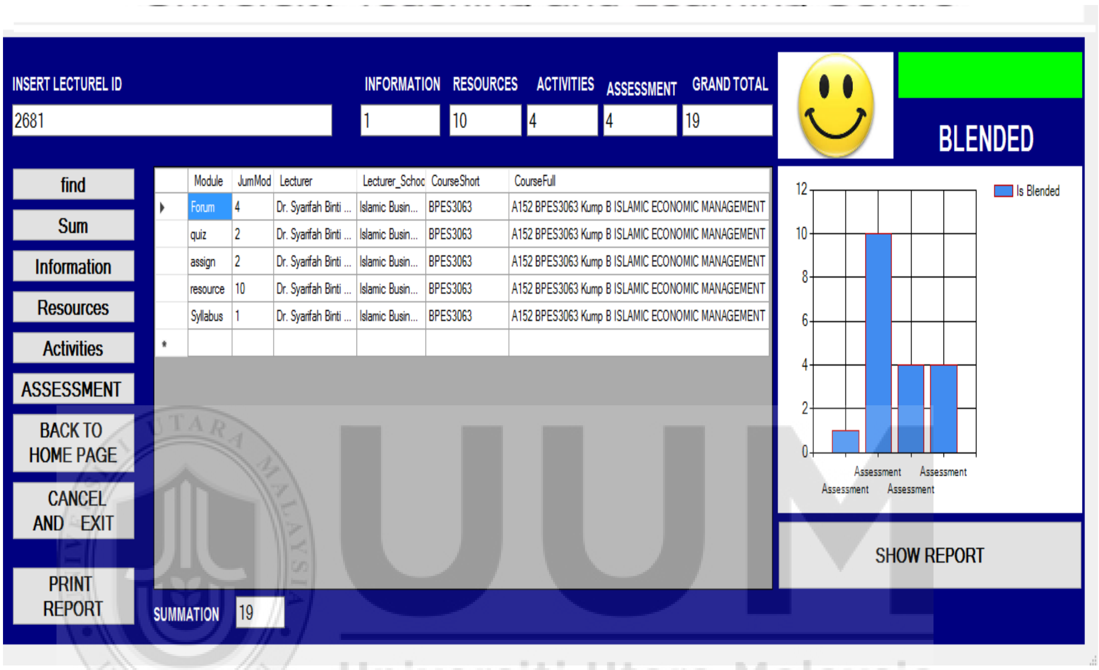
```
SELECT [StaffNo]='" + textBox1.Text + "') and (([Module]='assign') or
([Module]='quiz') or ([Module]='turnitintooltwo'))) ");
```

These data will be used in creating the intended virtual tables to used them in creation the intended BI reports for blended learning system as clearly indicated in the next section.

Universiti Utara Malaysia

5.3.5.3 DV Presentation

After applying the proposed GODV approach and obtaining all the relevant attributes which is used to create the virtual table, the reports were created based on the requirements of the organization. Figures 5.29, 5.30, 5.31 and 5.32 show some of the organization's BI reports.



Universiti Utara Malaysia

Figure 5.29. The Organization Business Intelligence Report1

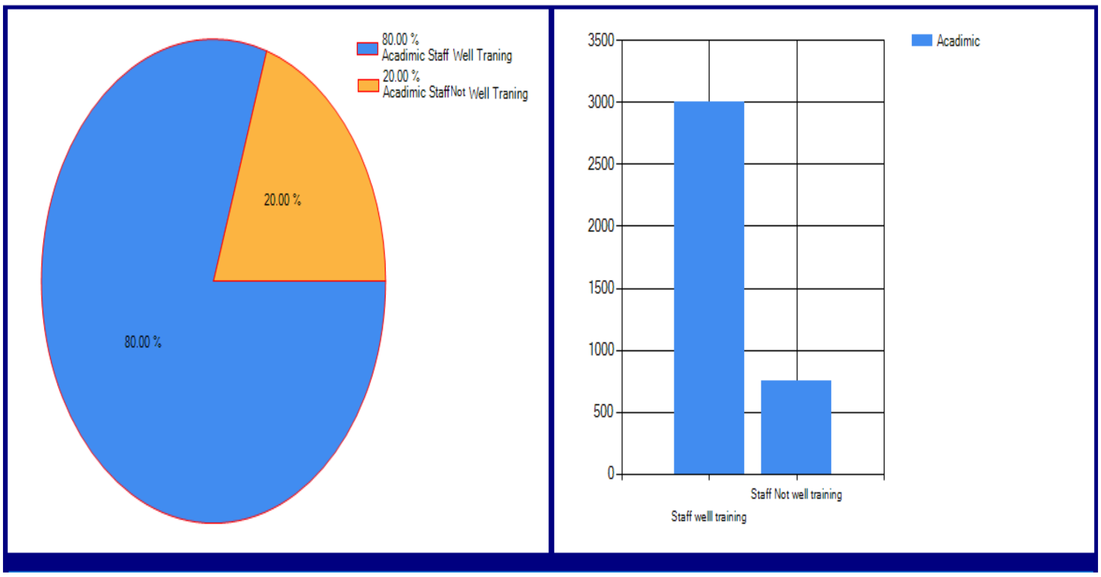


Figure 5.30. The Organization Business Intelligence Report2

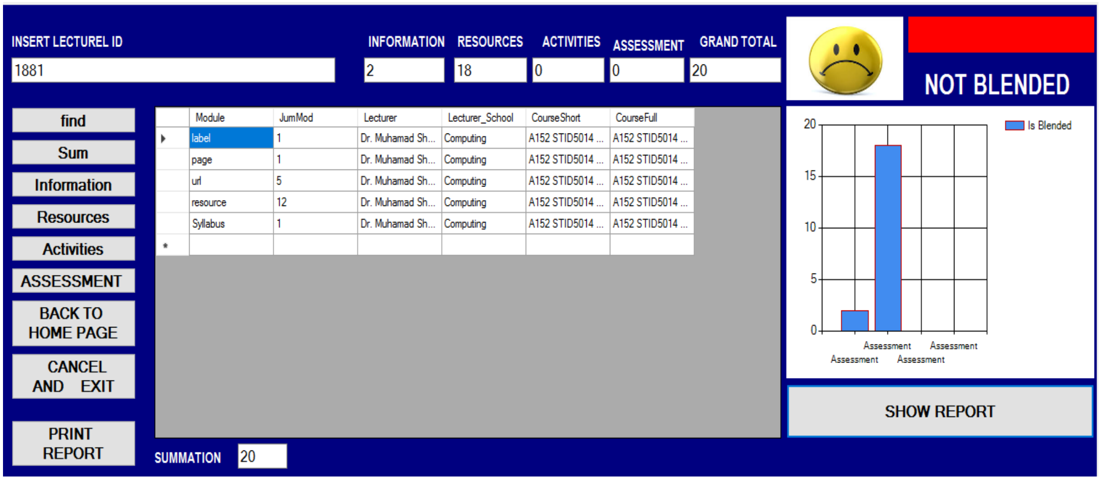


Figure 5.31. The Organization Business Intelligence Report3

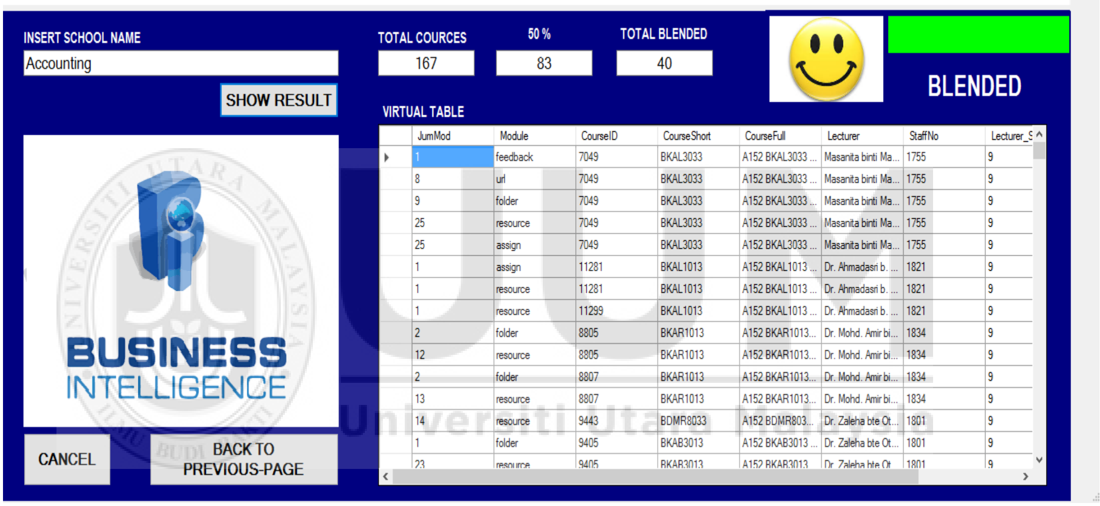


Figure 5.32. The Organization Business Intelligence Report4

Consequently, the procedure of creation and publishing of the organization reports will iteratively apply in order to create and view the rest of BI reports based on the organization requirements.

As clearly shown in figures 5.29, 5.30, 5.31, and 5.32, the organization's reports are presented. These reports are reflect the organization requirements. Moreover, the respondents generally expressed positive opinions about their experiences with these reports in terms of these reports are up-to-date and contain live data to support the

Organization's decision-making process in near real-time, as extensively discussed in section 5.4.7.

Technically, it is difficult to join more than three database tables at the same time as this will result in very complex query statements. In order to solve this problem, multi-virtual tables should be created in order to obtain the final virtual table that can be used to create the required BI reports. Refer to appendix H for the related pseudo code.

5.3.6 The benefit

There are many benefits of applying the proposed DVDeM model in this organization. Mainly, the proposed model and its prototype can support decision making process in near real time data to data consumers, as well as the capability to add and delete data sources based on business needs without the needs to change the end user application. Furthermore, the model reduces the infrastructure and implementation cost.

Universiti Utara Malaysia

5.3.7 The Prototype Evaluation

As discussed in Chapter 3, the usability testing for the prototype was conducted. The measurements were made through an instrument named Q-U, which comprises of six main dimensions, visibility, flexibility, learnability, application behavior, error control and help, and near real-time decision making as a usability attributes. This instrument is composed of 36 items spread over the six attributes (refer to Chapter 3 section 3.4.3.1.1). During the usability test, 30 respondents with computer science, software engineering, and IT background participated. The overall Q-U instrument is available in Appendix C. The respondents were required to answer questions after trying the DVDeM prototype.

Descriptive statistics was used. in order to describe the basic features of the data in this study, to provide summaries about the sample and the measures, together with graphic analysis, and to form the basis of virtually every quantitative analysis of data. The justification for using descriptive statistics is because it is useful to summarize the group of data using a combination of tabulated description (i.e., tables), graphical description (i.e., graphs and charts) and statistical commentary (i.e., a discussion of the results). Refer to appendix F for all descriptive statistics. Moreover, the findings are adequately discussed in the next sections.

5.3.7.1 The Visibility of DVDeM Prototype

As clearly indicated by the Figure 5.33 and Table 5.14 that all respondents agreed that the DVDeM prototype is able to extract useful information. Particularly, the majority of respondents (16.66 % strongly agree and 83.34% agree) confirm that DVDeM prototype can display the information (virtual tables, virtual data mart, and BI reports) in an uncluttered and well-structured manner; but 3.33% of them took the side of neutrality. Besides, all the respondents (6.67% and 93.33%) either strongly agree and agree respectively that all DVDeM prototype instructions are visible and self- explanatory. Regarding the navigation options in to DVDeM prototype, the findings shows (13.34% strongly agree and 83.33% agree) of the respondents think that the navigation options such as (Links, shortcuts, home, back, forward, etc.) are displayed in a visible manner, but 3.33% of them took the side of neutrality. Likewise, all respondents (20% strongly agree and 80 % agree) found that DVDeM prototype has ability to communicate the status at all time (whether resting, processing etc.). Furthermore, the majority of respondents (10% strongly agree and 86.66% agree) affirmed that the data in DVDeM prototype is concisely presented, while 3.34% of them took the side of neutrality. Noteworthy, all respondents (26.66% strongly agree and 73.34% agree) state that the DVDeM prototype meets the user expectation in terms of functions and capabilities. Besides, (13.34 strongly agree and 86.66 % agree) that the DVDeM prototype has an accepted interface; while 16.66% of them took the side of neutrality. On the same aspect, (30% strongly agree and 63.33% agree) stated that the DVDeM prototype has a pleasant interface, while the remaining 6.67% took the side of neutrality. Eventually, all respondents (33.33% strongly agree and 66.67 % agree) found that the organization information in DVDeM prototype visualizes in a clear manner All the findings above are supported by the bar chart and statistics in Figure 5.33 and Table 5.14.

Table 5.14

The Visibility of DVDeM Prototype

Visibility Items	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Mean**	**Std. Deviation**	**Total**
Item 1	N= 5 16.66%	N=25 83.34%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	3.170	0.397	**30**
Item 2	N= 2 06.67%	N= 28 93.33%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	4.070	0.254	**30**
Item 3	N= 4 13.34%	N= 25 83.33%	N= 1 03.33%	N= 0 00.00%	N= 0 00.00%	4.130	0.346	**30**
Item 4	N= 6 20.00%	N= 24 80.00%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	4.200	0.407	**30**
Item 5	N= 3 10.00%	N= 26 86.66%	N= 1 03.34%	N= 0 00.00%	N= 0 00.00%	4.070	0.365	**30**
Item 6	N= 8 26.66%	N= 22 73.34%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	4.270	0.450	**30**
Item 7	N= 4 13.34%	N= 26 86.66%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	4.130	0.346	**30**
Item 8	N= 9 30.00%	N= 19 63.33%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.230	0.568	**30**
Item 9	N= 10 33.33%	N= 20 66.67%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	4.330	0.479	**30**

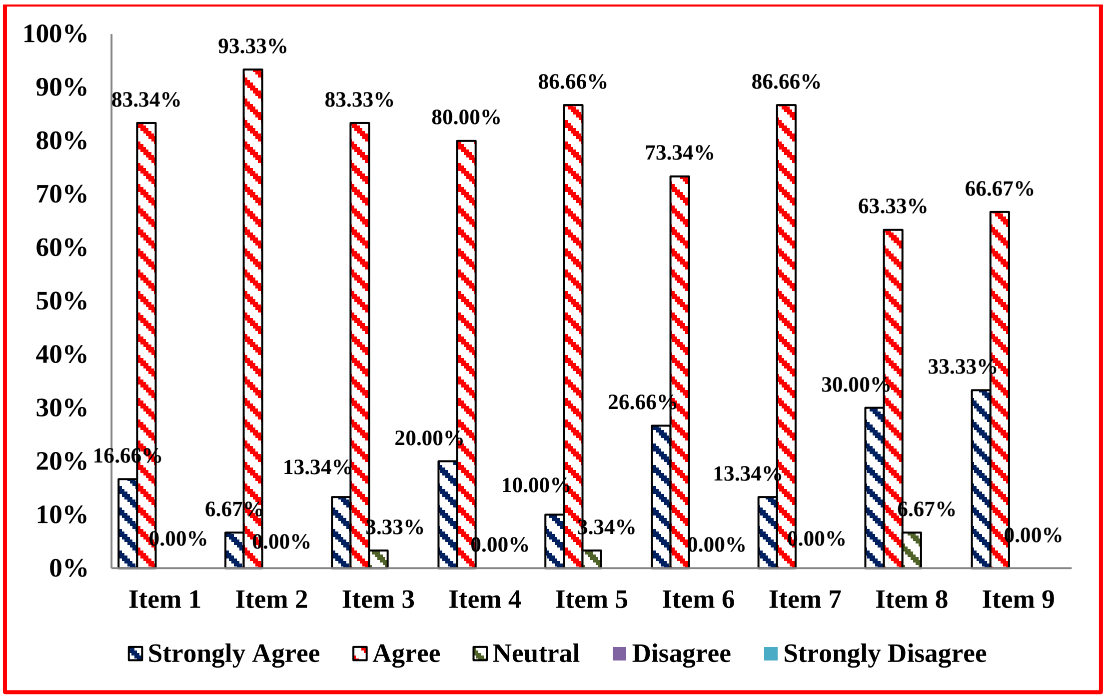
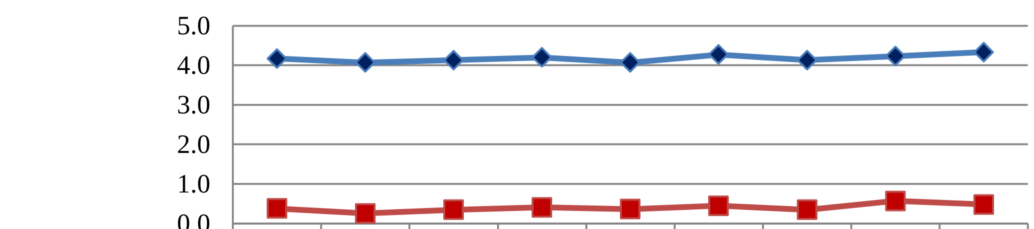


Figure 5.33. The Visibility of DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.34 shows the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have positive reaction about the DVDeM prototype usability in terms of visibility; and thereby strengthens the conclusion. The implication of this finding is that DVDeM and its prototype are visible. Besides, the standard deviation values for each measurement item are less than 0.7, indicating how close to the average the data is clustered.



	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9
Mean	4.170	4.070	4.130	4.200	4.070	4.270	4.130	4.230	4.330
Std. Deviation	0.379	0.254	0.346	0.407	0.365	0.450	0.346	0.568	0.479

Figure 5.34. The Descriptive Statistic of DVDeM Prototype Visibility

5.3.7.2 The Flexibility of DVDeM Prototype

The results reveal that the DVDEM prototype is flexible. Particularly, the majority of the respondents (16.66% strongly agree and 80% agree) with feeling in control during usage of DVDeM prototype; while 3.34% tend to neutrality. Besides, the majority of the respondents (20% and 73.33%) strongly agree and agree respectively that the DVDeM prototype has customizable feature; while 6.67% of them preferred to be in neutral. Regarding the data entry in DVDeM prototype, the finding shows (20% strongly agree and 76.66% agree) of the respondents think that the design data entry is flexible; but 3.34% of them believe in stay in the neutral side. Likewise, all of the respondents (23.33% strongly agree and 76.67 % agree) found that by using DVDeM prototype the data can be used, manipulated, and processed in an easy manner. Eventually, most of the respondents (6.66% strongly agree and 80% agree) found that the DVDeM prototype can handle user-specified windows; while the remaining 13.34% took the side of neutrality. All aforementioned findings are supported by the bar chart and statistics in Figure 5.35 and Table 5.15. Respondents have affirmed that the DVDeM prototype is flexible.

Table 5.15
The Flexibility of DVDeM Prototype

Flexibility Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Std. Deviation	Total
Item 1	N= 5 16.66%	N=24 80.00%	N= 1 03.34%	N= 0 00.00%	N= 0 00.00%	3.130	0.434	**30**
Item 2	N= 6 20.00%	N= 22 63.33%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.170	0.481	**30**
Item 3	N= 6 20.00%	N= 23 73.33%	N= 1 03.34%	N= 0 00.00%	N= 0 00.00%	4.170	0.461	**30**
Item 4	N= 7 23.33%	N= 23 76.67%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	4.170	0.430	**30**
Item 5	N= 2 06.66%	N= 24 80.00%	N= 4 13.34%	N= 0 00.00%	N= 0 00.00%	4.870	0.507	**30**

Universiti Utara Malaysia

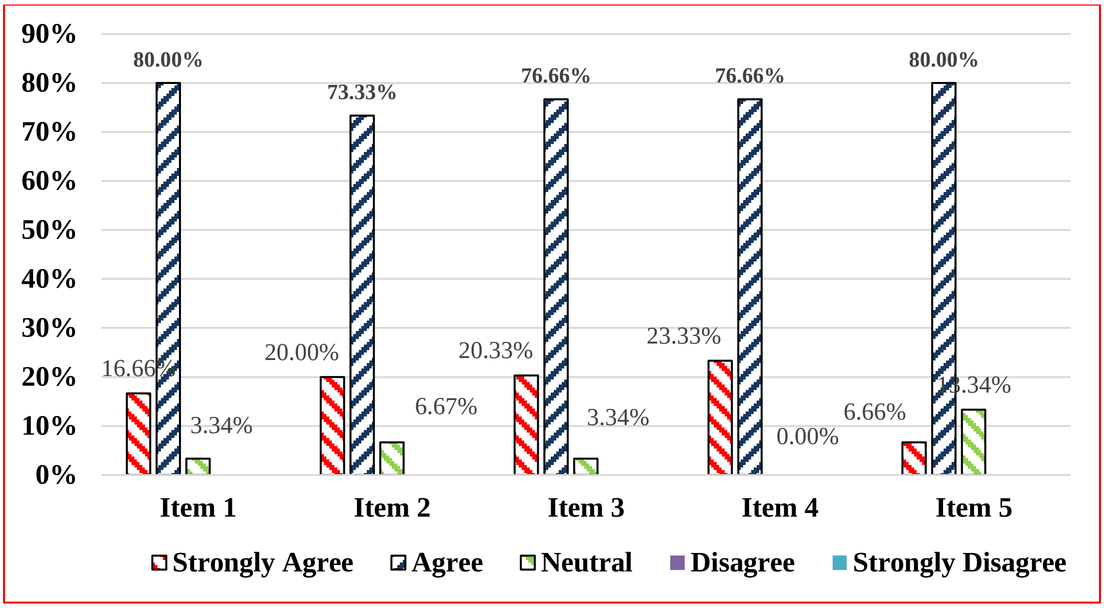
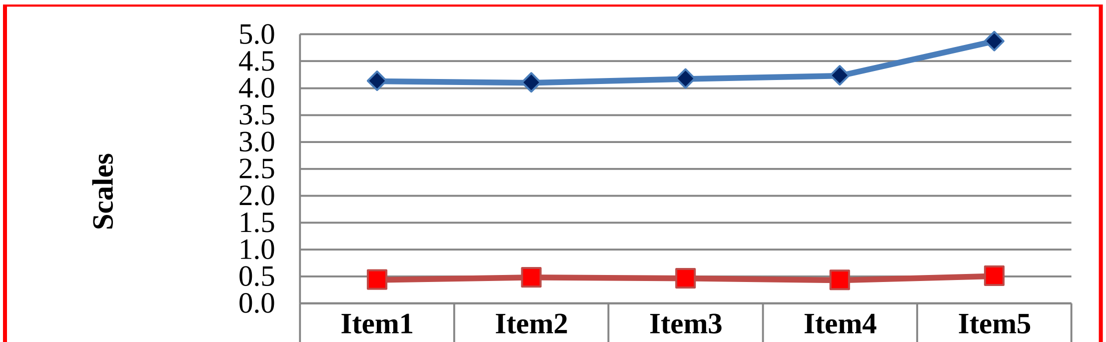


Figure 5.35. The Flexibility of DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.36 showed the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype usability in terms of flexibility; and thereby strengthens the conclusion. The implication of this finding is that DVDeM and its prototype are workable in practice in term of flexibility. Besides, the standard deviation values for each measurement item are less than 0.6, this indicates how close to the average the data is clustered.



	Item1	Item2	Item3	Item4	Item5
Mean	4.130	4.100	4.170	4.230	4.870
Std. Deviation	0.434	0.481	0.461	0.43	0.507

Figure 5.36. The Descriptive Statistic of DVDeM Prototype Flexibility

5.3.7.3 The Learnability of DVDeM Prototype

As clearly indicated by the Figure 5.37 and Table 5.16, the results reveal that the learnability is high. The majority of the respondents (6.66% strongly agree and 83.34% agree) found the data grouping of DVDeM prototype is reasonable for easy learning; but 10% tend to neutrality. Besides, 26.66% and 63.34% strongly agree and agree respectively, agree that the DVDeM prototype promotes learnability to make it accessible for infrequent usage; but 10% of them preferred to be in neutral. Regarding the grouping of menu options in DVDeM prototype, the finding shows all of the respondents (16.66% strongly agree and 83.34% agree) think that it is logical. Eventually, majority of respondents (23.33% strongly agree and 70 % agree) found that by using DVDeM prototype could become productive quickly; but 6.67% of them preferred to be in neutral. All aforementioned findings above are supported as shown in the bar chart and statistics in Figure 5.37 and Table 5.16.

Universiti Utara Malaysia

Table 5.16
The Learnability of DVDeM Prototype

Items	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Mean**	**Std. Deviation**	**Total**
Item 1	N= 2 06.66%	N=25 83.34%	N= 3 10.00%	N= 0 00.00%	N= 0 00.00%	3.970	0.414	**30**
Item 2	N= 8 26.66%	N= 19 63.34%	N= 3 10.00%	N= 0 00.00%	N= 0 00.00%	4.170	0.592	**30**
Item 3	N= 5 16.66%	N= 25 83.34%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	4.170	0.379	**30**
Item 4	N= 7 23.33%	N= 21 70.00%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.170	0.531	**30**

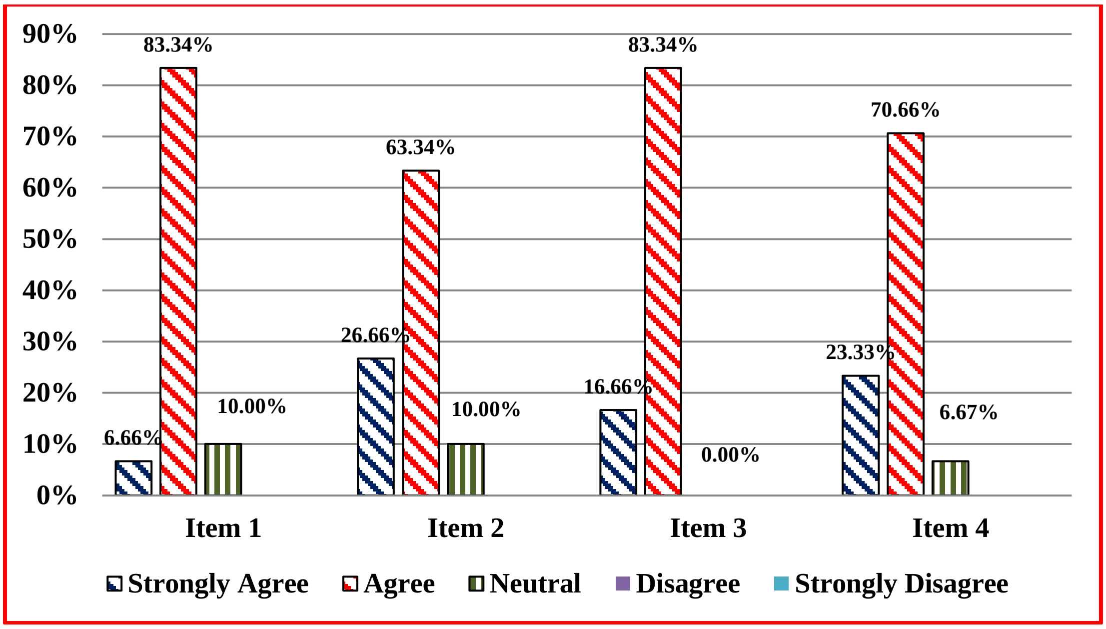
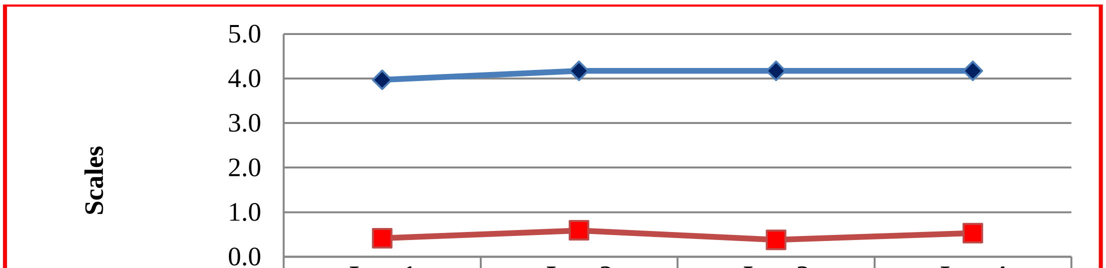


Figure 5.37. The Learnability of DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.38 showed the mean values for each measurement item are greater than 4 this indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype usability in terms of learnability; and thereby strengthens the conclusion. The implication of this finding is that DVDeM and its prototype are highly learnable. Besides, the standard deviation values for each measurement item are less than 0.6. This indicates how close to the average the data is clustered.

Universiti Utara Malaysia



	Item1	Item2	Item3	Item4
Mean	3.970	4.170	4.170	4.170
Std. Deviation	0.414	0.592	0.379	0.531

Figure 5.38. The Descriptive Statistic of DVDeM Prototype Learnability

5.3.7.4 The Application Behavior of DVDeM Prototype

The findings indicate that all of the respondents (13.33% strongly agree and 86.67% agree) that by using DVDeM prototype will enhance their efficiency through a consistently rapid response rate. Likewise, all respondents (30% and 70%) either strongly agree and agree respectively, believes that the DVDeM prototype is consistent. Besides, the majority of respondents (13.33% strongly agree and 80% agree) think that the information provides by DVDeM prototype was effective in helping them complete the tasks and scenarios; but 6.67% of them preferred to be in neutral. In regard to the availability of information, the majority of respondents (23.33% strongly agree and 70% agree) think that by using DVDeM prototype it was easy to find, however, 6.67% of them preferred to be in neutral. When any mistake occurred, the majority of the respondents (20% strongly agree and 73.33% agree) believe that it can recover easily and quickly; further, 6.67% of them prefer to be on the neutral side. Also, the majority of the respondents (30% strongly agree and 63.33% agree) believe that they are able to complete the tasks and scenarios quickly via using DVDeM prototype, further, 6.67% of them prefer to be on the neutral side. Refer to Figure 5.39 and Table 5.17.

Table 5.17
The Application Behaviour of DVDeM Prototype

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Std. Deviation	Total
Item 1	N=4 13.33%	N=26 86.67%	N= 0 00.00%	N=0 00.00%	N=0 00.00%	4.130	0.346	**30**
Item 2	N= 9 30.00%	N= 21 70.00%	N= 0 00.00%	N= 0 00.00%	N= 0 00.00%	4.300	0.466	**30**
Item 3	N= 4 13.33%	N= 24 80.00%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.070	0.450	**30**
Item 4	N= 7 23.33%	N= 21 70.00%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.170	0.531	**30**
Item 5	N= 6 20.00%	N= 22 73.33%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.130	0.507	**30**
Item 6	N= 9 30.00%	N= 19 63.33%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.230	0.568	**30**

Universiti Utara Malaysia

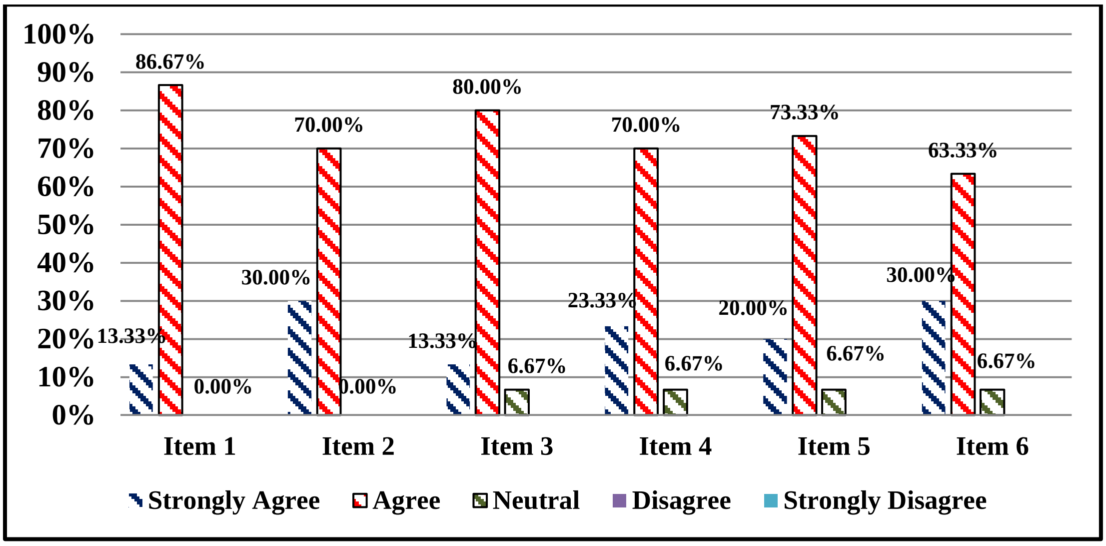
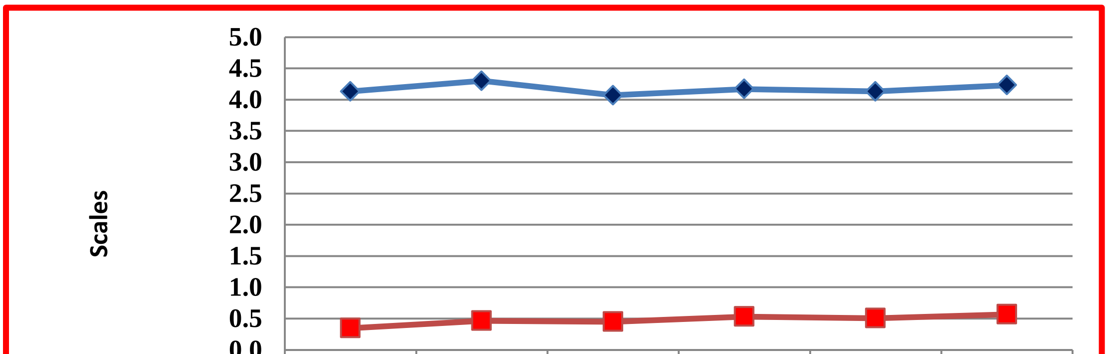


Figure 5.39. The Application Behaviour for DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.40 showed the mean values for each measurement item are greater than 4 this indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype usability in terms of application behavior; and thereby strengthens the conclusion. Besides, the standard deviation values for each measurement item are less than 0.6, which indicates how close to the average the data is clustered.



	Item1	Item2	Item3	Item4	Item5	Item6
Mean	4.130	4.300	4.070	4.170	4.130	4.230
Std. Deviation	0.346	0.466	0.450	0.531	0.507	0.568

Figure 5.40. The Descriptive Statistics Application Behaviour

5.3.7.5 The Error Control and Help of DVDeM Prototype

The majority of the respondents (20% strongly agree and 60% agree) think that DVDeM prototype has the ability for error prevention and error recovery; but 20% tend to neutrality. Besides, (30% and 56.66%) of them strongly agree and agree respectively think that DVDeM prototype provides online help and onscreen messages in a clear manner, but 13.34% tend to neutrality. Likewise, the majority of the respondents (46.66% strongly agree and 46.66% agree) state that DVDeM prototype can give error messages that clearly inform what is the problem and how to fix. In terms of on-demand help, but 6.67% of them tend to neutrality. The majority of the respondents (30.33% strongly agree and 63% agree), found that DVDeM prototype provides such function, but 6.67% tend to neutrality. Noteworthy, the majority of the respondents (40 % strongly agree and 53.33% agree) feel satisfied with DVDeM Prototype, but 6.67% of them tend to neutrality. The majority also (33.33% strongly agree and 60% agree), feel satisfied with how easy to use DVDeM prototype; but 6.67% tend to neutrality. Figure 5.41 and Table 5.18, has affirmed that DVDeM prototype provides error control and help.

Table 5.18
The Error Control and Help for DVDeM Prototype

Items	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	Mean	Std. Deviation	Total
Item 1	N=6 20.00%	N=18 60.00%	N=6 20.00%	N=0 00.00%	N=0 00.00%	4.270	0.583	**30**
Item 2	N= 9 30.00%	N= 17 56.66%	N= 4 13.34%	N= 0 00.00%	N= 0 00.00%	4.230	0.568	**30**
Item 3	N= 14 46.66%	N= 14 46.66%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.330	0.606	**30**
Item 4	N= 9 30.00%	N= 19 63.33%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.230	0.430	**30**
Item 5	N= 12 40.00%	N= 16 53.33%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.330	0.479	**30**
Item 6	N= 10 33.33%	N= 18 60.00%	N= 2 06.67%	N= 0 00.00%	N= 0 00.00%	4.100	0.607	**30**

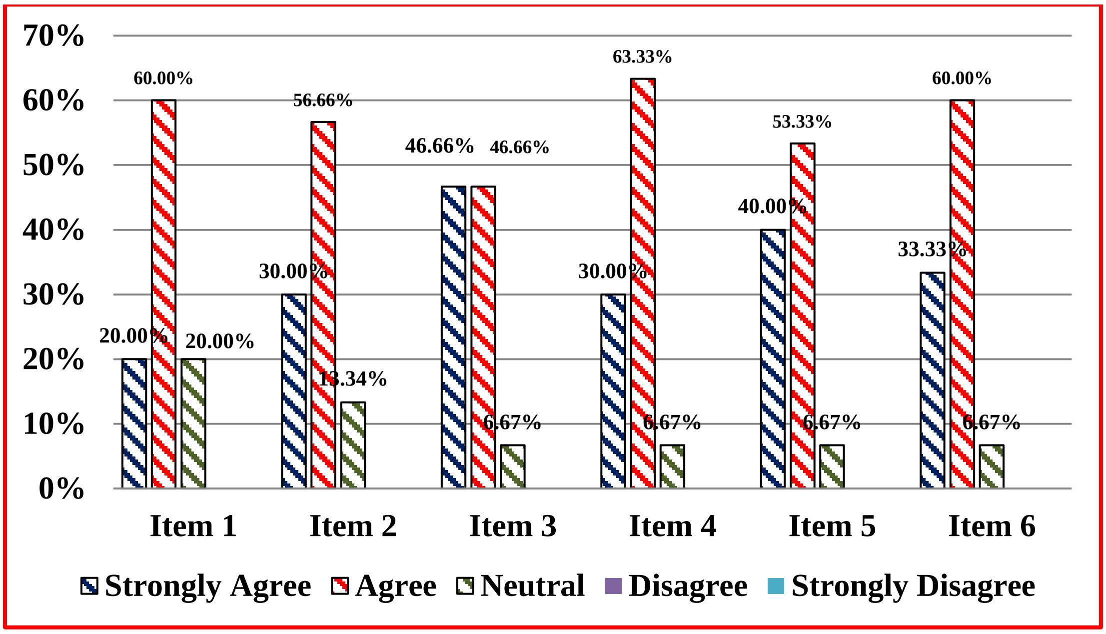
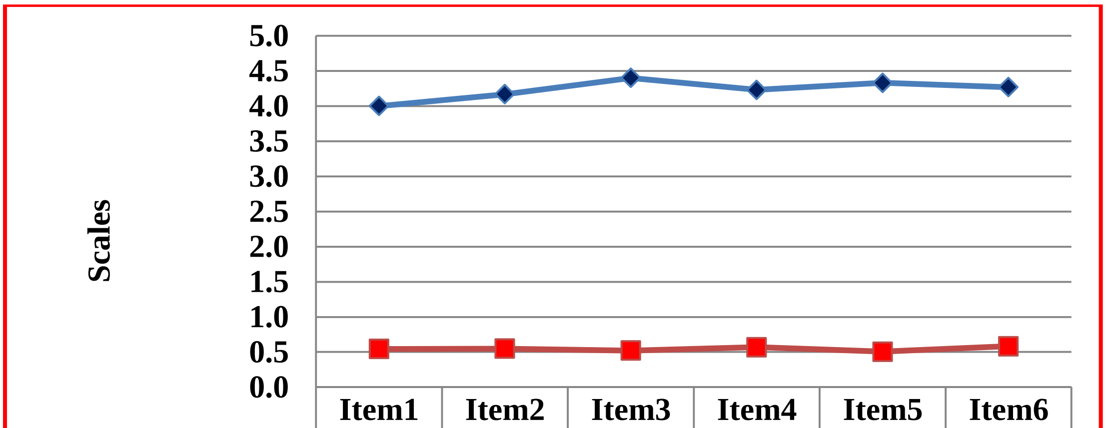


Figure 5.41. The Error Control and Help for DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.42 show the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype in terms of error control and help; and thereby strengthens the conclusion. Besides, the standard deviation values for each measurement item are less than 0.6, which indicates how close to the average the data is clustered.



	Item1	Item2	Item3	Item4	Item5	Item6
Mean	4.000	4.170	4.400	4.230	4.330	4.270
Std. Deviation	0.543	0.548	0.521	0.568	0.506	0.583

Figure 5.42. The Descriptive Statistic for Error Control and Help

5.3.7.6 Near Real-Time Decision Making of DVDeM Prototype

As clearly indicated by the results visualize in Table 5.19, that the DVDeM prototype is able to provide near real-time data for decision-making. Particularly, the majority of the respondents (36.66% strongly agree and 56.66% agree) agree that the knowledge sharing is allowed in DVDeM prototype; while 6.67% tend to neutrality. Likewise, most of the respondents (30% and 63.33%) strongly agree and agree respectively, that DVDeM prototype assists the decision-making process by providing the information visualization functionality (comparison charts, graphs in revealing trends etc.); but 6.67% tend to neutrality. Besides, the majority of the respondents (23.33% strongly agree and 70% agree) believe that DVDeM prototype provides sufficient coverage for all data resources; while 6.67% tend to neutrality. Furthermore, all the respondents (23.33% strongly agree and 76.67% agree) affirm that DVDeM prototype provides data on time to take suitable actions and decisions. likewise, all the respondents (33.33% strongly agree and 66.67% agree) affirm that the data provided by DVDeM prototype are always live data and up to date, which means DVDeM proposed model and its prototype provides fully supported for near real-time decision-making. The majority of them (23.33 % strongly agree and 63.33% agree) found that DVDeM prototype provides high-level data quality; however, 13.34% tend to neutrality. Refer to Figure 5.43 and Table 5.19, which affirmed that the DVDeM prototype is believe to provide near real-time data for decision making.

Universiti Utara Malaysia

Table 5.19

Near Real-Time Decision Making for DVDeM Prototype

Near Real-Time Decision Making Items	**Strongly Agree**	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**	**Mean**	**Std. Deviation**	**Total**
Item 1	N=11 36.66%	N=17 56.66%	N=2 06.67%	N=0 00.00%	N=0 00.00%	4.270	0.583	**30**
Item 2	N=9 30.00%	N=19 63.33%	N=2 6.67%	N=0 00.00%	N=0 00.00%	4.230	0.568	**30**
Item 3	N=7 23.33%	N=21 70.00%	N=2 6.67%	N=0 00.00%	N=0 00.00%	4.330	0.606	**30**
Item 4	N=7 23.33%	N=23 76.67%	N=0 0.00%	N=0 00.00%	N=0 00.00%	4.230	0.430	**30**
Item 5	N=10 33.33%	N=20 66.66%	N=0 0.00%	N=0 00.00%	N=0 00.00%	4.330	0.479	**30**
Item 6	N=7 23.33%	N=19 63.33%	N=4 13.34%	N=0 00.00%	N=0 00.00%	4.100	0.607	**30**

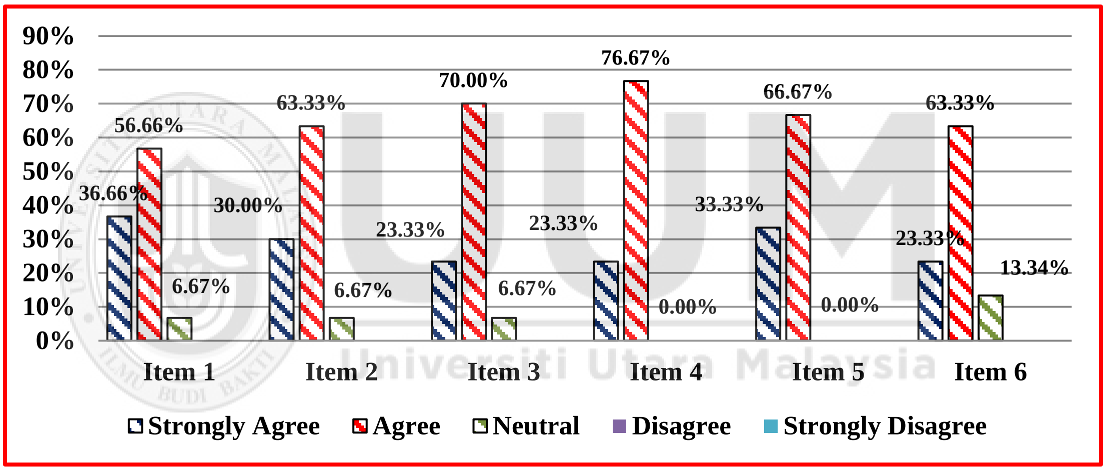


Figure 5.43. Near Real-Time Decision Making for DVDeM Prototype

On the same aspects, the descriptive analysis presented in Figure 5.44 showed the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item, hence, the respondents have a positive reaction about the DVDeM prototype usability in terms of near real-time decision making, and thereby strengthens the conclusion. Besides, the standard deviation values for each measurement item are less than 0.7, which indicates how close to the average the data is clustered.

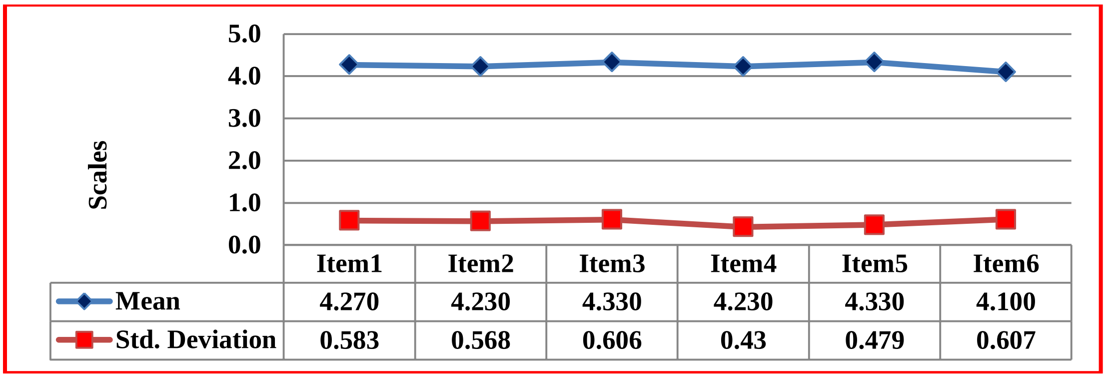


Figure 5.44. The Descriptive Statistic for Near Real-Time Decision Making of DVDeM Prototype

5.3.7.7 Overall Usability Finding

In order to obtain the overall finding, the average of findings for each usability attribute was calculated. The overall usability is made of six dimensions as depicted in Table 5.20 and Figure 5.45. Calculating the average agreement value for each dimension, the figure show that for each dimension, the majority of respondents agree that the DVDeM prototype provide data visibility, is flexible, easy to learn, behave as expected, provide necessary error control and help, and data to allow near real time decision making. Overall, on total average 94.62 %of the respondents agree to this.

Universiti Utara Malaysia

Table 5.20
The Overall Usability Finding

	Usability Dimension	**Strongly Agree + Agree**
1	Visibility	98.51%
2	Flexibility	94.66%
3	Learnability	93.49%
4	Application Behavior	95.55%
5	Error Control & Help	90.00%
6	Near Real-Time Decision Making	95.55%
	The Average	**94.62**

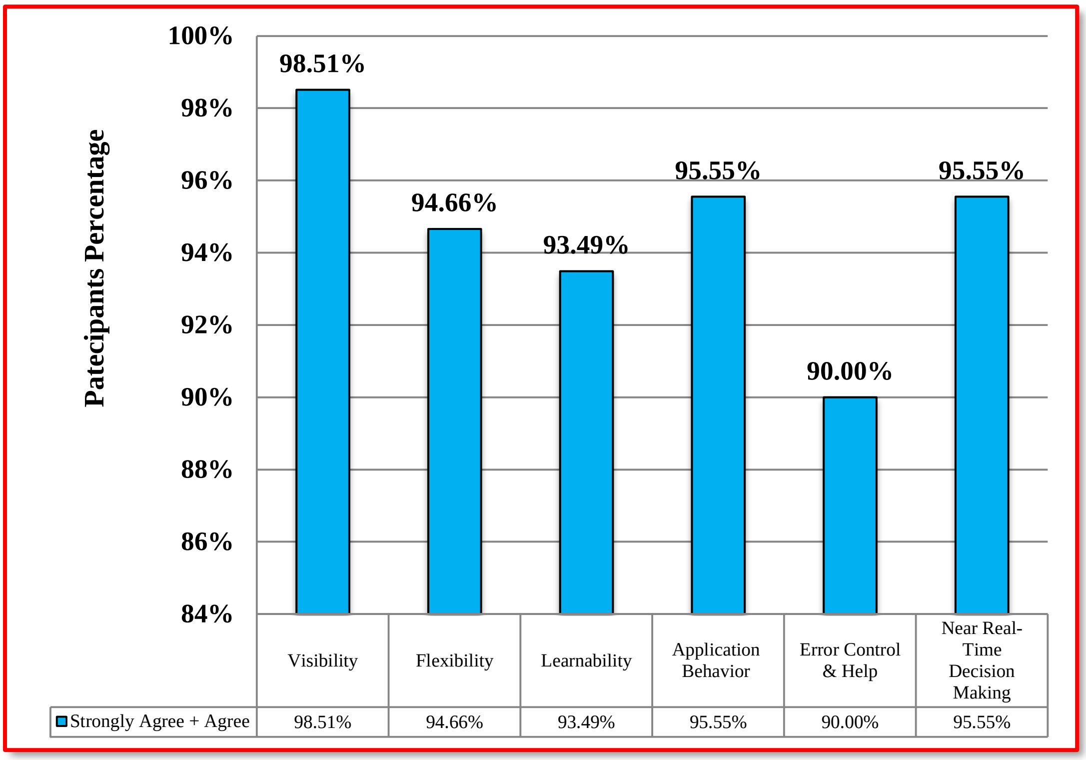


Figure 5.45. Overall Finding Education Sector

5.4 Summary

In a nutshell, this chapter discusses the implementation of the DVDeM prototype in two case studies, as well as measuring the usability of DVDeM prototype. the results of user experience of the DVDeM prototype spread over these case studies; (business sector and education sector). The purpose was to measure users' perception towards usability of the proposed DVDeM model.

The measurements were made through an instrument named Q-U, which comprises of six main attributes, visibility, flexibility, learnability, application behavior, error control and help, and finally, near real-time decision making.

Furthermore, 60 participants(divided into 30 participants for each case) participated in the usability test. In order to present the data in a very meaningful manner, the descriptive statistics was utilized, both the mean and standard deviation were

UUM Universiti Utara Malaysia

calculated, tabulated and presented. Besides, the calculation of the average of all Q-U usability attributes findings was calculated to obtain user satisfaction percentage.

From the studies, the findings also showed that the majority of the participants (more than 90%) in both of education and business sectors, have affirmed the workability of the DVDeM and the usability of the prototype in particular able to deliver near real-time decision-making data. Findings also indicate theoretical and practical contributions for developers to develop efficient BI applications using DV technique.

Also, the mean values for each measurement item are greater than 4 indicating that the respondents agreed with the statement for each measurement item. Meanwhile, it was found that the mean scores for overall usability attributes of DVDeM design model fall under “High” or "Fairly High”. Therefore, the results show sufficient indications that by adopting DVDeM model in developing a system, the usability of the produced system is perceived by the majority of respondents as high and is able to support near real time decision making data.

Universiti Utara Malaysia

CHAPTER SIX
CONCLUSION

6.1 Introduction

The details of research background, research objectives and research questions are discussed in Chapter 1. The research focuses on design and development of Data Virtualization Design Model for Near- Real Time Decision Making in Business Intelligence Environment (DVDeM). Hence, to accomplish the main aim, five specific research questions were carried out:

i. What are the main components of the DVDeM design model?

ii. How to design DVDeM for BI environment using data virtualization technique?

iii. How should such model be validated in BI environment?

iv. Based on DVDeM, how can a prototype be developed in BI environment?

v. How should such prototype be tested in BI environment?

Also, this study formed a main aim to meet its expectation, which is to propose a Data virtualization design model for real time decision making in business intelligence environment that is able to ensure the helpfulness of the organizational decision making. To accomplish this, five objectives were outlined:

i. To identify the main components of DVDeM model.

ii. To design DVDeM model using data virtualization technique for business intelligence environments.

iii. To validate DVDeM in BI environment using expert review and focus group discussion.

iv. To develop prototype based on DVDeM model.

v. To test DVDeM prototype in BI environment.

The research objectives in Chapter 1 and literatures reviewed in Chapter 2 lead to a research structural framework, as discussed in Chapter 3, which involves the three components of triangulation methods combined of five phases adapted from DSRM: theoretical (preliminary investigation and elicit work), development (proposed design model and determine usability strategies), and empirical (analysis, deduction and evaluation process to test users' experience). In Chapter 4, an analysis of components and element through two comparative approaches of existing studies as well as real world DV solutions, initiated the formation and definition of a figure of the proposed design model DVDeM. All these were translated into working prototype of DVDeM and were discussed in detail in Chapter 4 and 5. In addition, the prototype of the DVDeM was tested in two case studies were discussed in detail in Chapter 5.

Therefore, this chapter highlights the answers of research questions, and describes the solutions proposed for each research question. In addition, this chapter also provides the summaries, discussions and describes research contributions to the body of knowledge. This chapter then ends with a discussion of future recommendation of this research, and conclusions of the study.

6.2 Overall Discussion and Conclusion

This section answers the research objectives and the research questions. There are five research questions; five research objectives are formulated to answer the questions. Obviously, the findings show that all the research questions were successfully answered through the design model and the usability testing that have been evaluated in BI environment. Furthermore, the discussions in details for each research question will be further explained in the next sections.

6.2.1 Research Question 1:

What are the main components of such design model?

In answering this question, few activities were conducted and have been discussed in this thesis. The main purpose of these activities was to identify components, phases, activities, and flows involved in developing the proposed model. The activities include expert consultation, content analysis of literature, two comparative studies one of them focused on existing BI models, while the second one focused on DV solutions for real-world. The phases and components of the proposed model (DVDeM) stand beyond three phases as mentioned in chapter 4, these phases named: data virtualization requirement gathering, data virtualization development, and data virtualization presentation. There are three phases, nine main components, and twenty-six sub components or the formed in the proposed model. Details of the components are presented in Chapter 4. Eventually, all these model phases and components were then combined and made up the DVDeM design model. The proposed model with its phases and components were also reviewed and found well accepted by the experts. And, therefore, the objective 1 has been achieved. Refer to Table 6.1

Table 6.1
Main Phases and Main Components of the DVDeM

1	**Phase 1. Data Virtualizations Requirements Gathering**	
	DVDeM Model Components	**Sub- Components and Processes**
1.1	Organization and Business Requirements	✓ Enterprise Architecture which linking with ✓ Enterprise Business Architecture (EBA). Which include: {strategic plans, goals, objectives, Measures}. ✓ Enterprise Information Architecture (EIA). ✓ Enterprise Service Component Architecture (ESA). ✓ Enterprise Technical Architecture which include: {Software Requirement and Hardware Requirement}. ✓ Organization Sponsorship and Involvement, which include: {Management and Potential Users}. ✓ Business Requirements, which include :{ Strategic Plan and Performance Measures}.
1.2	Data Sources Requirements	✓ Data Classification • High effort data integration. • Medium effort data integration. • Low effort data Integration.
1.3	Document Requirements Specification	✓ Document all Requirements and use these documentation in: ✓ Selection data integration based on • Time. • Cost. • Data quality. • Data sources status. • Complexity of data model. • User needs.
2	**Phase 2. Data Virtualization Development**	
2.1	Data Preparation and Data Connection	✓ Data Understanding and Data Connection: • Collect initial data. • Explore data. • Describe data. • Verify data quality. ✓ Data Preparation: • Data selection: {decides which data will be use}. • Data profiling: {convert to flattening form}. • Data cleaning:{ solve data quality issues}. ✓ Imported Data Sources.

Universiti Utara Malaysia

Table6.1 Continued

2.2	Data Manipulation and Data Management	✓ Create Meta data for data sources. ✓ Create wrappers. ✓ Define Mapping. ✓ Create Virtual Tables. ✓ Create virtual data marts. ✓ Apply GODV Approach. ✓ Create virtual table based on GODV. ✓ Create virtual data mart based on GODV
3	Phase 3. Data Virtualization Presentation	
		✓ Graphical User Interface (GUI). • View virtual tables. • View reports based on the virtual table. • View virtual data marts. • View reports based on virtual data marts. • View virtual table based on GODV. • View reports based on GODV. • View virtual data marts based on GODV. • View reports based on GODV.

6.2.2 Research Question 2:

How to design a system for BI environments using data virtualization technique?

The proposed model DVDeM was obtained through literature review (Chapter 2), two comparative studies one of them focused on existing business intelligence models, while the second one focused on ten DV solutions for real-world case studies. Review of literature analyzed appropriate theories and content for design and implementing BI environment, data integration design models, gathering requirements models, as well as DV solutions in real-world cases studies. This resulted in extraction of the model phases, model components as well as all model functions. Eventually, the first version of the proposed model was produced. In terms of model validation, two validation methods were conducted (expert review and focus group discussion). Their valuable comments and corrections were added value to the proposed DVDeM model and has guided the modification of the first version. Finally, the final version was produced; therefore, the objective 2 has been achieved.

6.2.3 Research Question 3:

How should such model be validated in BI environment?

The proposed model was validated through expert review (Chapter 4 Section 4.5.1) and focus group discussion (Chapter 4 Section 4.5.2).

Firstly, the proposed model (first version) was validated by expert review. It was conducted by gathering (12) experts in the field of BI, IT and multimedia. This review obtained feedbacks to improve the proposed model. Then these feedbacks were used to refine the proposed model to obtain DVDeM (final version). Next, the conceptual model was validated in focus group discussion.

Secondly, the focus group discussion was conducted with other (6) experts to review the conceptual model. Then, this review obtained feedbacks to improve the conceptual model. Then, these feedbacks were used to refine the final version of DVDeM proposed model. Accordingly, the final version of DVDeM proposed model was sent to prototyping. The prototype was developed based on the DVDeM proposed model. This step was completed to validate the DVDeM proposed model.

6.2.4 Research Question 4:

Based on the model, how can a prototype be developed for BI environment?

To achieve the objective 3, a prototype based on DVDeM proposed model has been developed and implemented. As mentioned in Chapter 4, the intended prototype consists of three main phases named: DV requirement gathering, DV development, and DV presentation. Details explanation about prototype development is provided in the sections 4.1. Meanwhile, as mentioned in section Chapter 1 and Chapter 3, the prototype was implemented in two cases studies for both business sector and education sector. Besides, the usability tests were conducted, in order to find out how the users can reach the prototype objectives. Details explanation about prototype evaluation is provided in the Chapter 5.

6.2.5 Research Question 5:

How should such prototype be tested in BI environment?

In answering this question, A prototype based on DVDeM was implemented in two real-world case studies, meanwhile, (Q-U) evaluation instrument was constructed which is used to test the prototype usability. The Q-U consists of 6 dimensions to measure the usability of the DVDeM prototype, which are visibility, flexibility, learnability, application behavior, error control and help, and business intelligence. Furthermore, 36 items on these dimensions were adapted from relevant studies. The instrument was found highly reliable in the pilot study with Cronbach's Alpha for each dimension was greater than 0.7. In addition, when the DVDeM prototype was tested, two general findings were obtained; (i) (93.00 %) (both strongly agree and agree) of the participants have affirmed the workability of the DVDeM and its prototype in terms of usability in general and particularly, supporting decision making process in providing near real time data (ii) the descriptive statistics (mean and standard deviation) have confirmed that there are sufficient indications that by adopting DVDeM does work in practice besides providing theoretical and practical contributions for developers to develop efficient BI applications using DV technique.

The results were obtained, showing enough evidence to conclude that the proposed DVDeM design model and its prototype are workable in practice in BI environment as well as a guideline in the academic field.

In line with above situations, the main aim of this research has been accomplished after all specific objectives were achieved successfully.

6.3 Contribution

The impact of this study can be perceived in two perspectives; the tangible application, and to the body of knowledge and theory. Chapter 1, Section 1.12 discussed in detail the contributions of this study.

The study introduces the concept of DVDeM; this is a type of model that can use as guidance to develop BI applications. It includes comprehensive phases and components for developing BI applications in a virtual manner. Meanwhile, The DVDeM concept is complimented with a design model, to guide BI developers on how to develop BI application based on DVDeM. The proposed DVDeM model includes the requirements and steps to make DVDeM Prototype.

Universiti Utara Malaysia

Prototype of DVDeM was developed in this study and found to satisfy users in two organizations. It was implemented in two case studies both (education and business sectors). The DVDeM prototype was also tested in terms of usability using Q-U instrument which was adapted.

In addition, this study contributed to the theories and body of knowledge through the findings of user testing which can assist BI developers in developing BI application in BI environment. Though the existing BI models and approaches were found as adequately providing BI applications guideline, however, they are not considered providing near real-time for data integration to support decision-making process. Consequently, existing BI models, approaches, and methods are not designed to meet the requirement of organizations. That if designed carefully by considering all DVDeM phases and their components as part of the BI application development and in mind, and in mind the outcome of the system output to work and be workable.

6.4 Limitations of the Study and Recommendations for Future Works

There are few limitations encountered in this study. Even though they did not have a direct bearing on the findings of this study, future researchers are encouraged for further improvements.

The design model for DVDeM is proposed as part of BI environment using virtual data integration. Guideline for usable BI environment for developing BI application in a virtual manner has not been found in any literature. Three main phases which associated with their components and sub-components were considered in the model: (i) data virtualization requirement gathering; (ii) data virtualization development; (iii) data virtualization presentation. All aforementioned above were proposed as the outcome of considered and implemented many methods in this study, which are elicit work, content analysis, comparative analysis, expert review, and focus group discussion. Several different targets and BI prototypes were considered to form the features for DVDeM prototype. Although, the prototypes used were not meticulously designed for a universal purpose. Indeed, consideration of universal model regarding the different BI applications might come out with different design model.

Universiti Utara Malaysia

Besides, in the proposed model, the focusing on a technical perspective of BI only and focusing on the first stage of BI stages, while the comprehensive BI model should be handling all BI stages as well as the organizational perspective of BI.

A future research should study the impact all BI phases on decision making process, which start from data sourcing and ending with data visualization will be necessary. In this, apart from evaluating the overall functionality of the BI model which this study did, data integration execution time can be studied.

6.5 Significance of Study

The significant of this study is in two areas: (1) Business Intelligence developers and Organizations. Furthermore, finding obtained shown some indications that the prototype which is based on the proposed DVDeM design model has the advantages which could give benefits to the following stakeholders:

Universiti Utara Malaysia

6.5.1 Business intelligence developers

Business intelligence and data integration developers can benefit from the proposed model (those work in decision support systems, data integration, and business intelligence developers). They can take related components from the proposed model and apply them in their intended applications. There are many ways developers can use this study:

i. The developers, who do not know about data virtualization, can use the proposed model as an introduction to the topic of data virtualization. In terms of practical side, the developers would drill into the two case studies to obtain the ideas about how the organizations can apply the data virtualization. While in terms of the theory behind the practice, the developers would study the literature review and comparative analysis in this study.

ii. The developers who already understand about data virtualization in general, in terms of learning about adoption best practices the developers refer to the chapter five, to obtain the ideas about how the organizations can apply the data virtualization.

iii. The developers, who are implementing data virtualization now, would use the proposed model as a reference for those times when understanding the path taken by others could provide new insights to help them and their organizations to achieve their business agility objectives sooner.

6.5.2 Organizations

In terms of organizations, the proposed DVDeM model can deliver the following benefits.

6.5.2.1 Intend to Adopt Data Virtualization

If the organization is new to data virtualization, the benefits that can be obtained by using the proposed DVDeM model is to demonstrate with real-world examples how they can move beyond traditional data integration and use data virtualization to improve the organization's business agility. This requires an understanding of what data virtualization is and a roadmap for its effective implementation.

Universiti Utara Malaysia

6.5.2.2 Already Adopting Data Virtualization

If the organization is already adopting data virtualization, the benefits that can be obtained by using the proposed DVDeM model is to help them successfully accelerate and expand their adoption, compound their business agility gains and achieve other business benefits from data virtualization.

6.6 Conclusion

The research has systematically identified the main phases, components, sub-components and the relevant processes between them, and came up in proposing a DVDeM model which was validated through expert review and focus group discussion, also, was developed a prototype based on DVDeM which was implemented in two case studies (business and education sectors), and finally, tested the usability of the prototype. From the findings, there are indications that the proposed DVDeM model could be utilized to design and develop BI systems in supporting decision making process by providing near real-time data for decision makers in BI environment.

Universiti Utara Malaysia

REFERENCES

Abdullah, M. F., & Ahmad, K. (2015). *Business intelligence model for unstructured data management.* Paper presented at the The International Conference on Electrical Engineering and Informatics (ICEEI).

Aguilar, E. R., Ruiz, F., García, F., & Piattini, M. (2006). *Evaluation measures for business process models.* Paper presented at the Proceedings of the 2006 ACM symposium on Applied computing.

Ahuja, A., Kumar, A., & Singh, R. (2012). An Approach for Virtualization and Integration of Heterogeneous Cloud Databases. *International Journal of Engineering Research and Applications, 2*(5), 352-355.

Akanmu, S. A., & Jamaludin, Z. (2016). Students' data-driven decision making in HEI: The explicit knowledge involved. *International Journal of Information and Education Technology, 6*(1), 71.

AlSuwaidan, L., & Zemirli, N. (2015). *Toward a knowledge-based model for real-time business intelligence.* Paper presented at the Science and Information Conference (SAI), 2015.

Amor, H. (2014). Top 5 Criteria for Evaluating Business Intelligence Reporting and Analytics Software. Retrieved from http://www.arcplan.com/en/blog/2014/07/top-5-criteria-for-evaluating-business-intelligence-reporting-and-analytics-software-2/#comments

Anderson-Lehman, R., Watson, H. J., Wixom, B. H., & Hoffer, J. A. (2008). Flying high with real-time business intelligence *Handbook on Decision Support Systems 2* (pp. 443-462): Springer.

Andriessen, D. (2006). *Combining design-based research and action research to test management solutions.* Paper presented at the 7th World Congress Action Research.

Anton, A. I. (1996). *Goal-based requirements analysis.* Paper presented at the Requirements Engineering, 1996., Proceedings of the Second International Conference on.

Ariffin, A. M. (2009). *Conceptual design of reality learning media (RLM) model based on entertaining and fun constructs.* Universiti Utara Malaysia.

Azvine, B., Cui, Z., & Nauck, D. D. (2005). Towards real-time business intelligence. *BT Technology Journal, 23*(3), 214-225.

Universiti Utara Malaysia

Azvine, B., Cui, Z., Nauck, D. D., & Majeed, B. (2006). *Real time business intelligence for the adaptive enterprise.* Paper presented at the E-Commerce Technology, 2006. The 8th IEEE International Conference on and Enterprise Computing, E-Commerce, and E-Services, The 3rd IEEE International Conference on.

Baharuddin, R., Singh, D., & Razali, R. (2013). Usability dimensions for mobile applications—A review. *Res. J. Appl. Sci. Eng. Technol, 5*, 2225-2231.

Barnum, C. M., & Dragga, S. (2001). *Usability testing and research*: Allyn & Bacon, Inc.

Barone, D., Yu, E., Won, J., Jiang, L., & Mylopoulos, J. (2010). Enterprise modeling for business intelligence *The practice of enterprise modeling* (pp. 31-45): Springer.

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with computers, 23*(1), 4-17.

Benbunan-Fich, R. (2001). Using protocol analysis to evaluate the usability of a commercial web site. *Information & management, 39*(2), 151-163.

Bostrom, R. P., & Heinen, J. S. (1977). MIS problems and failures: a socio-technical perspective, part II: the application of socio-technical theory. *MIS quarterly*, 11-28.

Botta-Genoulaz, V., & Millet, P.-A. (2006). An investigation into the use of ERP systems in the service sector. *International journal of production economics, 99*(1), 202-221.

Bresciani, P., Perini, A., Giorgini, P., Giunchiglia, F., & Mylopoulos, J. (2004). Tropos: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems, 8*(3), 203-236.

Bruckner, R., List, B., & Schiefer, J. (2002). *Striving towards Near Real-Time Data Integration for Data Warehouses. Data Warehousing and Knowledge Discovery.* Paper presented at the 4th International Conference, DaWaK.

Bucher, T., Gericke, A., & Sigg, S. (2009). Process-centric business intelligence. *Business Process Management Journal, 15*(3), 408-429.

Burstein, F., & Holsapple, C. (2008). *Handbook on decision support systems 2: variations*: Springer Science & Business Media.

Castellanos, M., Simitsis, A., Wilkinson, K., & Dayal, U. (2009). *Automating the loading of business process data warehouses.* Paper presented at the

Universiti Utara Malaysia

Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology.

Cavana, R. Y., Delahaye, B. L., & Sekaran, U. (2001). *Applied business research: Qualitative and quantitative methods*: John Wiley & Sons Australia.

Chang, E., Hussain, F., & Dillon, T. (2006). *Trust and reputation for service-oriented environments: technologies for building business intelligence and consumer confidence*: John Wiley & Sons.

Chu, M. Y. (2004). *Blissful Data: Wisdom and Strategies for Providing Meaningful, Useful, and Accessible Data for All Employees*: AMACOM Div American Mgmt Assn.

Churchman, C. W. (1971). The Design of Inquiring Systems Basic Concepts of Systems and Organization.

Cicchetti, D. V., Shoinralter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of interrater reliability: A Monte Carlo investigation. *Applied Psychological Measurement, 9*(1), 31-36.

Clemmensen, T., Hertzum, M., Hornbæk, K., Shi, Q., & Yammiyavar, P. (2009). Cultural cognition in usability evaluation. *Interacting with computers, 21*(3), 212-220.

Codd, E. F., Codd, S. B., & Salley, C. T. (1993). Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate. *Codd and Date, 32*.

Cody, W. F., Kreulen, J. T., Krishna, V., & Spangler, W. S. (2002). The integration of business intelligence and knowledge management. *IBM systems journal, 41*(4), 697-713.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative science quarterly*, 1-25.

Conrad, C., Gasman, M., Lundberg, T., Nguyen, T.-H., Commodore, F., & Samayoa, A. C. (2013). Using educational data to increase learning, retention, and degree attainment at minority serving institutions (MSIs): A Research Report of Penn Graduate School of Education, GSE.

Coursaris, C. K., & Kim, D. J. (2011). A meta-analytical review of empirical mobile usability studies. *Journal of usability studies, 6*(3), 117-171.

Craik, A. D., & Leibovich, S. (1976). A rational model for Langmuir circulations. *Journal of Fluid Mechanics, 73*(03), 401-426.

Cummins, R. A., & Gullone, E. (2000). *Why we should not use 5-point Likert scales: The case for subjective quality of life measurement.* Paper presented at the Proceedings, second international conference on quality of life in cities.

D'Souza, E., & White, E. (2006). Demand Forecasting for the Net Age: From Thought to Fulfillment in One Click: Global Integrated Supply Chain Systems, Idea Group Inc., pp.

Davis, J. R., & Eve, R. (2011). *Data Virtualization: Going Beyond Traditional Data Integration to Achieve Business Agility*.

Dayal, U., Castellanos, M., Simitsis, A., & Wilkinson, K. (2009). *Data integration flows for business intelligence.* Paper presented at the Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology.

Dayal, U., Wilkinson, K., Simitsis, A., & Castellanos, M. (2009). Business Processes Meet Operational Business Intelligence. *IEEE Data Eng. Bull., 32*(3), 35-41.

Devlin, B. A., & Murphy, P. T. (1988). An architecture for a business and information system. *IBM Systems Journal, 27*(1), 60-80.

Dix, A. (2002). *Beyond intention-pushing boundaries with incidental interaction.* Paper presented at the Proceedings of Building Bridges: Interdisciplinary Context-Sensitive Computing, Glasgow University.

Dwivedi, Y. K., Papazafeiropoulo, A., & Metaxiotis, K. (2009). Exploring the rationales for ERP and knowledge management integration in SMEs. *Journal of Enterprise Information Management, 22*(1/2), 51-62.

Dwolatzky, B., Kennedy, I., & Owens, J. (2002). Modern software engineering methods for developing courseware.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of management review, 14*(4), 532-550.

Ellis, G., & Dix, A. (2006). *An explorative analysis of user evaluation studies in information visualisation.* Paper presented at the Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization.

Eriksson, H.-E., & Penker, M. (2000). *Business modeling with UML*: Wiley Chichester.

Eve, R., & Davis, J. R. (2011). *Data Virtualization: Going Beyond Traditional Data Integration to Achieve Business Agility*: Composite Software.

Ferguson, M. (2011). Succeeding with Data Virtualization High Value Use Cases for Analytical Data Services. *Business Intelligence Journal, 4*, 15.

Fiora, B. (1998). Ethical business intelligence is NOT Mission Impossible. *Strategy & Leadership, 26*(1), 40-41.

Folkes, C., & Quintas, P. (2004). Knowledge mapping: map types, contexts and uses. *KM-SUE Working Paper. The Open University, Milton Keynes*.

Franconi, E., & Sattler, U. (1999). A Data Warehouse Conceptual Data Model for Multidimensional Aggregation:a preliminary report.

Frandsen-Thorlacius, O., Hornbæk, K., Hertzum, M., & Clemmensen, T. (2009). *Non-universal usability?: a survey of how usability is understood by Chinese and Danish users.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Gacenga, F., Cater-Steel, A., Toleman, M., & Tan, W.-G. (2012). A Proposal and Evaluation of a Design Method in Design Science Research. *Electronic Journal of Business Research Methods, 10*(2).

Ghosh, R., Haider, S., & Sen, S. (2015). *An integrated approach to deploy data warehouse in business intelligence environment.* Paper presented at the Computer, Communication, Control and Information Technology (C3IT), 2015 Third International Conference on.

Gill, B., Borden, B. C., & Hallgren, K. (2014). A conceptual framework for data-driven decision making. *Final Report of Research conducted by Mathematica Policy Research, Princeton, submitted to Bill & Melinda Gates Foundation, Seattle, WA.*

Giorgini, P., Rizzi, S., & Garzetti, M. (2008). GRAnD: A goal-oriented approach to requirement analysis in data warehouses. *Decision Support Systems, 45*(1), 4-21.

GROUP, G. (1998). Introducing the Zero-Latency Enterprise: Research Note COM-04-3770.

Guo, S.-S., Yuan, Z.-M., Sun, A.-B., & Yue, Q. (2015). A New ETL Approach Based on Data Virtualization. *Journal of Computer Science and Technology, 30*(2), 311-323.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*: Pearson College Division.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis 6th Edition. *New Jersey: Pearson Education*.

Hair Jr, J. F. (2007). Knowledge creation in marketing: the role of predictive analytics. *European Business Review, 19*(4), 303-315.

Hall, D. J. (2008). Decision makers and their need for support *Handbook on Decision Support Systems 1* (pp. 83-102): Springer.

Herschel, R. T., & Jones, N. E. (2005). Knowledge management and business intelligence: the importance of integration. *Journal of Knowledge Management, 9*(4), 45-55.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly, 28*(1), 75-105.

Hill, J., & Scott, T. (2004). A consideration of the roles of business intelligence and e-business in management and marketing decision making in knowledge-based and high-tech start-ups. *Qualitative Market Research: An International Journal, 7*(1), 48-57.

Hopkins, B. (2011). Data virtualization reaches the critical mass. *Business Intelligence Journal, 4*(4), 12.

Hou, C.-K. (2012). Examining the effect of user satisfaction on system usage and individual performance with business intelligence systems: An empirical study of Taiwan's electronics industry. *International Journal of Information Management, 32*(6), 560-573.

Iacono, J. C., Brown, A., & Holtham, C. (2011). The use of the Case Study Method in Theory Testing: The Example of Steel eMarketplaces. *The Electronic Journal of Business Research Methods, 9*(1), 57-65.

IBM. (2016, 2016). Today's business settings. Retrieved January 03, 2016, 2016, from http://www.inside-erp.com/

Inmon, W. H. (1996). Building the data warehouse. *John Whiley & Sons, NY*.

Inmon, W. H. (2005). *Building the data warehouse*: John wiley & sons.

Inmon, W. H., & Hackathorn, R. D. (1994). *Using the data warehouse*: Wiley-QED Publishing.

Jones, D. G., & Malik, J. (1992). Computational framework for determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing, 10*(10), 699-708.

Jooste, C., van Biljon, J., & Mentz, J. (2013). *Usability evaluation guidelines for business intelligence applications.* Paper presented at the Proceedings of the

South African Institute for Computer Scientists and Information Technologists Conference.

Jossen, C., Blunschi, L., Mori, M., Kossmann, D., & Stockinger, K. (2012). *The Credit Suisse Meta-data Warehouse.* Paper presented at the Data Engineering (ICDE), 2012 IEEE 28th International Conference on.

Karim, A. J. (2011a). The value of competitive business intelligence system (CBIS) to stimulate competitiveness in global market. *International Journal of Business and Social Science, 2*(19).

Karim, A. J. (2011b). The value of competitive business intelligence system (CBIS) to stimulate competitiveness in global market. *International Journal of Business and Social Science, 2*(19), 196-203.

Katsis, Y., & Papakonstantinou, Y. (2009). View-based data integration *Encyclopedia of Database Systems* (pp. 3332-3339): Springer.

Keen, P. G., & Morton, M. S. S. (1978). *Decision support systems: an organizational perspective* (Vol. 35): Addison-Wesley Reading, MA.

Kemper, H.-G., & Baars, H. (2009). *From data warehouses to transformation hubs-A conceptual architecture.* Paper presented at the ECIS.

Khraibet, H. N., Mousa, A. H., Bakar, A., & Shahbani, M. (2013). Intelligent Iraqi Health System (IIHS) using Online Analytical Process (OLAP) model.

Kimball, R. (1998). *The data warehouse lifecycle toolkit: expert methods for designing, developing, and deploying data warehouses*: John Wiley & Sons.

Kimball, R., & Caserta, J. (2004). *The data warehouse ETL toolkit*: John Wiley & Sons.

Kimball, R., & Ross, M. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*: John Wiley & Sons.

Kimball, R., Ross, M., Thorthwaite, W., Becker, B., & Mundy, J. (2008). *The data warehouse lifecycle toolkit*: John Wiley & Sons.

Landers, T., & Rosenberg, R. L. (1986). *An overview of multibase.* Paper presented at the Distributed systems, Vol. II: distributed data base systems.

Lans, R. v. d. (2013). Data Virtualization: Where Do We Stand Today? , from http://www.b-eye-network.com/view/16996

Universiti Utara Malaysia

Lau, L. K. (2005). *Managing business with SAP: planning, implementation and evaluation*: IGI Global.

Lavery, D., Cockton, G., & Atkinson, M. (1996). Heuristic evaluation. Usability evaluation materials. *Glasgow, United Kingdom: Department of Computing Science, University of Glasgow*.

Li, S.-T., Shue, L.-Y., & Lee, S.-F. (2008). Business intelligence approach to supporting strategy-making of ISP service management. *Expert Systems with Applications, 35*(3), 739-754.

Lin, H. X., Choong, Y.-Y., & Salvendy, G. (1997). A proposed index of usability: a method for comparing the relative usability of different software systems. *Behaviour & information technology, 16*(4-5), 267-277.

List, B., & Korherr, B. (2006). *An evaluation of conceptual business process modelling languages.* Paper presented at the Proceedings of the 2006 ACM symposium on Applied computing.

Liutong Xu, J. L., Ruixue Zhao, Bin Wu. (2011). A PAAS BASED METADATA-DRIVEN ETL FRAMEWORK. *IEEE*, 5.

Liyang, T., Zhiwei, N., Zhangjun, W., & Li, W. (2011). *A conceptual framework for business intelligence as a service (saas bi).* Paper presented at the Intelligent Computation Technology and Automation (ICICTA), 2011 International Conference on.

Loebbert, A. P. J. (2011). *Multi agent enhanced business intelligence for localized automatic pricing in grocery chains.* School of Information Technology, Bond University.

Lönnqvist, A., & Pirttimäki, V. (2006). The measurement of business intelligence. *Information Systems Management, 23*(1), 32.

Mantel, M. (1994). A basic framework for cost-justifying usability engineering. *Cost-justifying usability,* 9.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems, 15*(4), 251-266.

Marjanovic, O. (2007). *The next stage of operational business intelligence: Creating new challenges for business process management.* Paper presented at the System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on.

Marren, P. (2004). The father of business intelligence. *Journal of Business Strategy, 25*(6).

Universiti Utara Malaysia

Marsden, J. R. (2008). The Internet and DSS: massive, real-time data availability is changing the DSS landscape. *Information Systems and E-Business Management, 6*(2), 193-203.

McGregor, C., & Kumaran, S. (2002). *Business Process Monitoring Using Web Services in B2B e-Commerce.* Paper presented at the Proceedings of the 16th International Parallel and Distributed Processing Symposium.

McGregor, C., & Scheifer, J. (2003). *A framework for analyzing and measuring business performance with web services.* Paper presented at the E-Commerce, 2003. CEC 2003. IEEE International Conference on.

Mendling, J., Neumann, G., & Nüttgens, M. (2005). A comparison of XML interchange formats for business process modelling. *Workflow handbook*, 185-198.

Meredith, R., O'Donnell, P., & Arnott, D. (2008). Databases and data warehouses for decision support *Handbook on Decision Support Systems 1* (pp. 207-230): Springer.

Moore, M. G. (1973). Toward a theory of independent learning and teaching. *The Journal of Higher Education*, 661-679.

Morgan, D. L. (1996). Focus groups. *Annual review of sociology*, 129-152.

Moss, L. T., & Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications*: Addison-Wesley Professional.

Mousa, A. H., & Shiratuddin, N. (2015). *Data Warehouse and Data Virtualization Comparative Study*. Paper presented at the Developments of E-Systems Engineering (DeSE), 2015 International Conference on.

Mousa, A. H., Shiratuddin, N., & Bakar, M. S. A. (2014a). Generic Framework for Better Choosing Between Data Integration Types (GFCBDIT) During Build Business Intelligence Applications. *International Journal of Digital Content Technology and its Applications, 8*(5), 27.

Mousa, A. H., Shiratuddin, N., & Bakar, M. S. A. (2014b). Generic Framework for Better Choosing Between Data Integration Types (GFCBDIT) During Build Business Intelligence Applications. *International Journal of Digital Content Technology & its Applications, 8*(5).

Mousa, A. H., Shiratuddin, N., & Bakar, M. S. A. (2015a). Process Oriented Data Virtualization Design Model for Business Processes Evaluation (PODVDM) Research in Progress. *Jurnal Teknologi, 72*(4).

Mousa, A. H., Shiratuddin, N., & Bakar, M. S. A. (2015b). *RGMDV: An approach to requirements gathering and the management of data virtualization projects.* Paper presented at the INNOVATION AND ANALYTICS CONFERENCE AND EXHIBITION (IACE 2015): Proceedings of the 2nd Innovation and Analytics Conference & Exhibition.

Nasir, J., & Shahzad, M. K. (2007). Architecture for virtualization in data warehouse *Innovations and advanced techniques in computer and information sciences and engineering* (pp. 243-248): Springer.

Negash, S. (2004). Business intelligence. *The communications of the Association for Information Systems, 13*(1), 54.

Negash, S., & Gray, P. (2008). *Business intelligence*: Springer.

Newell, Allen, Simon, & Alexander, H. (1972). *Human problem solving* (Vol. 104): Prentice-Hall Englewood Cliffs, NJ.

Nguyen, T. M., & Tjoa, A. M. (2006). *Zero-latency data warehousing (ZLDWH): the state-of-the-art and experimental implementation approaches.* Paper presented at the RIVF.

Nielsen, J. (1994). Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier. *Cost-justifying usability*, 245-272.

Nielsen, J. (2012). How many test users in a usability study. *Nielsen Norman Group, 4*(06).

Norshuhada, & Shahizan. (2010). Design Research in Software Development :Constracting Linking Research Questions,Objectives,Methods and Outcomes. *U.U.Malaysia Ed.Uneversiti Utara Malaysia*

O'Brien, V. F., & Fuld, L. M. (1991). Business intelligence and the new Europe. *Planning Review, 19*(4), 29-34.

O’Leary, D. E. (2008). Decision Support System Evolution: Predicting, Facilitating, and Managing Knowledge Evolution. *Handbook on Decision Support Systems 2*, 345-367.

Offermann, P., Levina, O., Schönherr, M., & Bub, U. (2009). *Outline of a design science research process.* Paper presented at the Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology.

Olszak, C. M., & Ziemba, E. (2007). Approach to building and implementing business intelligence systems. *Interdisciplinary Journal of Information, Knowledge, and Management, 2*(2007), 134-148.

Universiti Utara Malaysia

Olszak, C. M., & Ziemba, E. (2010). Business performance management for competitive advantage in the information economy. *Journal of Internet Banking and Commerce, 15*(3), 93-104.

OMG, B. (2009). BPMN 1.2: Final Specification: Technical report.

Paim, F. R. S., & De Castro, J. F. B. (2003). *DWARF: An approach for requirements definition and management of data warehouse systems.* Paper presented at the Requirements Engineering Conference, 2003. Proceedings. 11th IEEE International.

Pourshahid, A., Amyot, D., Peyton, L., Ghanavati, S., Chen, P., Weiss, M., & Forster, A. J. (2008). *Toward an Integrated User Requirements Notation Framework and Tool forBusiness Process Management.* Paper presented at the e-Technologies, 2008 International MCETECH Conference on.

Pourshahid, A., Amyot, D., Peyton, L., Ghanavati, S., Chen, P., Weiss, M., & Forster, A. J. (2009). Business process management with the user requirements notation. *Electronic Commerce Research, 9*(4), 269-316.

Pourshahid, A., Richards, G., & Amyot, D. (2011). Toward a goal-oriented, business intelligence decision-making framework *E-Technologies: Transformation in a Connected World* (pp. 100-115): Springer.

Preece, J., Rogers, Y., & Sharp, H. (2002). Interaction Design: Beyond Human-Computer Interaction.

Ramachandran, S., Rajeswari, S., Murty, S., Valsan, M., Dayal, R., Rao, R., & Raj, B. (2010). *Design of a dimensional database for materials data.* Paper presented at the Trendz in Information Sciences & Computing (TISC), 2010.

Ramanigopal, C., Palaniappan, G., & Hemalatha, N. (2012). Business intelligence for infrastructure and construction industry. *ZENITH International Journal of Business Economics & Management Research, 2*(6), 71-86.

Reinschmidt, J., & Francoise, A. (2000). Business intelligence certification guide. *IBM International Technical Support Organisation.*

Reynolds, K. M., Twery, M., Lexer, M. J., Vacik, H., Ray, D., Shao, G., & Borges, J. G. (2008). Decision support systems in forest management *Handbook on Decision Support Systems 2* (pp. 499-533): Springer.

Ricardo Jorge Santos, J. B., Marco Vieira. (2011). 24/7 Real-Time Data Warehousing: A Tool for Continuous Actionable Knowledge. *IEEE,* 10.

Universiti Utara Malaysia

Richter, J., McFarland, L., & Bredfeldt, C. (2012). CB4-03: An Eye on the Future: A Review of Data Virtualization Techniques to Improve Research Analytics. *Clinical medicine & research, 10*(3), 166-166.

Roscoe, J. T. (1975). *Fundamental research statistics for the behavioral sciences [by] John T. Roscoe*.

Rouibah, K., & Ould-Ali, S. (2002). PUZZLE: a concept and prototype for linking business intelligence to business strategy. *The Journal of Strategic Information Systems, 11*(2), 133-152.

Rumbaugh, J., Jacobson, I., & Booch, G. (2004). *Unified Modeling Language Reference Manual, The*: Pearson Higher Education.

Samoff, J. (1999). Education sector analysis in Africa: limited national control and even less national ownership. *International Journal of Educational Development, 19*(4), 249-272.

Siemens, G., Dawson, S., & Lynch, G. (2013). Improving the quality and productivity of the higher education sector. *Policy and Strategy for Systems-Level Deployment of Learning Analytics. Canberra, Australia: Society for Learning Analytics Research for the Australian Office for Learning and Teaching*.

Sargut, G., & McGrath, R. G. (2011). Learning to live with complexity. *Harvard Business Review, 89*(9), 68-76.

Sauter, V. L. (2014). *Decision support systems for business intelligence*: John Wiley & Sons.

Schläfke, M., Silvi, R., & Möller, K. (2012). A framework for business analytics in performance management. *International Journal of Productivity and Performance Management, 62*(1), 110-122.

Sekaran, U. (1992). *Research methods for business: A skill building approach*: John Wiley & Sons.

Sekaran, U., & Bougie, R. (2011). Research method for business: A skill building approach: Taylor & Francis.

Shahbani, M., & Shiratuddin, N. (2011). Conceptual Design Model Using Operational Data Store (CoDMODS) for Developing Business Intelligence Applications. *IJCSNS, 11*(3), 161.

Shahzad. (2010). *A Data Warehouse Model for Integrating Fuzzy Concepts in Meta Table Structures.* Paper presented at the 17th-International Conference on Engineering of Computer-Based Systems.

Universiti Utara Malaysia

Shahzad, & Giannoulis, C. (2011). *Towards a Goal-Driven Approach for Business Process Improvement Using Process-Oriented Data Warehouse.* Paper presented at the Business Information Systems.

Shiratuddin, N., & Hassan, S. (2010). *Design Research in Software Development*. Kedah, Malaysia: Universiti Utara Malaysia Press.

Simon, & Herbert. (1960). The new science of management decision.

Siti Mahfuzah, S. (2011). *Conceptual Design Model of Computerized Personal-Decision AID (ComPDA).* Universiti Utara Malaysia.

Sommerville, I., & Sawyer, P. (1997). *Requirements engineering: a good practice guide*: John Wiley & Sons, Inc.

Sureephong, P., Chakpitak, N., Ouzrout, Y., & Bouras, A. (2008). An ontology-based knowledge management system for industry clusters *Global Design to Gain a Competitive Edge* (pp. 333-342): Springer.

Syamsul Bahrin, Z. (2011). *Mobile game-based learning (mGBL) engineering model.* Universiti Utara Malaysia.

Ta'a, A., Bakar, M. S. A., & Saleh, A. R. (2006). *Academic business intelligence system development using SAS® tools.* Paper presented at the Workshop on Data Collection System for PHLI-MOHE.

Thomas Jr, J. H. (2001). Business intelligence–why. *eAI Journal*, 47-49.

Tiwana, A. (2000). *The knowledge management toolkit: practical techniques for building a knowledge management system*: Prentice Hall PTR.

Trivedi, M. (2011). Regional and categorical patterns in consumer behavior: revealing trends. *Journal of Retailing, 87*(1), 18-30.

Turban, D. B., Forret, M. L., & Hendrickson, C. L. (1998). Applicant attraction to firms: Influences of organization reputation, job and organizational attributes, and recruiter behaviors. *Journal of Vocational Behavior, 52*(1), 24-44.

Uygun, Ö., Öztemel, E., & Kubat, C. (2009). Scenario based distributed manufacturing simulation using HLA technologies. *Information Sciences, 179*(10), 1533-1541.

Vaishnavi, V. K., & Kuechler Jr, W. (2007). *Design science research methods and patterns: innovating information and communication technology*: CRC Press.

Universiti Utara Malaysia

Van der Lans, R. (2012). *Data Virtualization for Business Intelligence Systems: Revolutionizing Data Integration for Data Warehouses*: Morgan Kaufmann.

Viaene, S., & Van den Bunder, A. (2011). The secrets to managing business analytics projects. *MIT Sloan Management Review, 53*(1), 65.

Voss, C., Tsikriktsis, N., & Frohlich, M. (2002). Case research in operations management. *International journal of operations & production management, 22*(2), 195-219.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information systems research, 3*(1), 36-59.

Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *Computer, 40*(9), 96-99.

Watson, H. J., Wixom, B. H., Hoffer, J. A., Anderson-Lehman, R., & Reynolds, A. M. (2006). Real-time business intelligence: Best practices at Continental Airlines. *Information Systems Management, 23*(1), 7.

Webster, J., & Watson, R. T. (2002). ANALYZING THE PAST TO PREPARE FOR THE FUTURE: WRITING A. *MIS quarterly, 26*(2).

Weng, L., Agrawal, G., Catalyurek, U., Kur, T., Narayanan, S., & Saltz, J. (2004). *An approach for automatic data virtualization.* Paper presented at the High performance Distributed Computing, 2004. Proceedings. 13th IEEE International Symposium on.

Weske, M. (2012). *Business process management: concepts, languages, architectures*: Springer.

Wiegers, K. E. (2002). Seven truths about peer reviews. *Cutter IT Journal, 15*(7), 31-37.

Wirth, R., & Hipp, J. (2000). *CRISP-DM: Towards a standard process model for data mining.* Paper presented at the Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining.

Wixom, B., Ariyachandra, T., Goul, M., Gray, P., Kulkarni, U., & Phillips-Wren, G. (2011). The current state of business intelligence in academia. *Communications of the Association for Information System, 29*(16), 299-312.

Wu, L., Barash, G., & Bartolini, C. (2007). *A service-oriented architecture for business intelligence.* Paper presented at the Service-Oriented Computing and Applications, 2007. SOCA'07. IEEE International Conference on.

Universiti Utara Malaysia

Yen, P.-Y., & Bakken, S. (2012). Review of health information technology usability study methodologies. *Journal of the American Medical Informatics Association*, *19*(3), 413-422.

You, H. (2010). *A knowledge management approach for real-time business intelligence.* Paper presented at the Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on.

Yu, C., & Popa, L. (2004). *Constraint-based XML query rewriting for data integration.* Paper presented at the Proceedings of the 2004 ACM SIGMOD international conference on Management of data.

Zellner, G. (2011). A structure evaluation of business process improvement approaches. *Business Process Management Journal, 17(2)*, 203-237.

Zhang, D.-P. (2009). *A Data Warehouse Based on University Human Resource Management of Performance Evaluation.* Paper presented at the Information Technology and Applications, 2009. IFITA'09. International Forum on.

Zikopoulos, P., deRoos, D., Bienko, C., Buglio, R., & Andrews, M. (2015). *What is big data?* (IBM Ed.). IBM: IBM.

Zur Muehlen, M., & Rosemann, M. (2004). *Multi-Paradigm Process Management.* Paper presented at the CAiSE Workshops (2).

Universiti Utara Malaysia

Appendix A

Expert Review Form

INSTRUMENT FOR EXPERT REVIEW:

DATA VIRTUALIZATION DESIGN MODEL FOR NEAR- REAL TIME DECISION MAKING IN BUSINESS INTELLIGENCE ENVIRONMENT

Dear Prof. / Dr. / Sir / Ma,

I am Ayad Hameed Mousa Albadri who is currently pursuing his PhD study in Multimedia at University Utara Malaysia. I am delighted to inform you that you have been selected to participate in this research on the reason as follows:

- Your qualifications either in higher education or database data integration or human computer interaction or instructional design or computer science areas, and/or technology or data virtualization or business intelligence.
- You have been studying/researching/teaching in data virtualization or data integration or data mining or business intelligence or computer science areas for many years.

My PhD research proposes the **Data Virtualization Design Model for Near- Real Time Decision Making in Business Intelligence Environment.** It aims is to propose an appropriate structure, layout and navigation as part of how to integrate and manage data in virtualize manner and deliver them in data consumers(end users). These proposed model will provide on-demand data and data on fly, live data (real data in real time to real users) in order to support decision makers to make better decisions.

One part of this research is to evaluate the proposed appropriate design strategies in a few dimensions as listed in the review form.

You will see the review questions give you ample opportunity to use your expertise, experiences, interest, and creativity. It would be greatly appreciated if you could complete this evaluation form.

The information supplied will be treated as confidential and will be used for the research purposes, which will be reported anonymously in academic publications.

Please feel free to contact me by e-mail: maryemayad@yahoo.com in regard to any queries.

Instructions:

The model development consist of three main phases the first one is Data virtualization requirement gathering, while the second one represent Data virtualization development, finally, the third one represent Data virtualization presentation. The first figure shows the proposed model in general while the others show each phase in details. Please read and go through them carefully. Once this is done, with the expertise you possess, please provide feedback for the entire question in the provided spaces. Finally, regarding other components in this model (Data virtualization validation and control and Data consumer feedback). Regarding Data virtualization validation and control: - In Data virtualization environments, traceability and change management must be carried out in both requirements and architectural spheres. The former is concerned with managing changes to agreed requirements and its impact to other requirements inside the same or in external documents. The Data consumer Feedback is important to overcome the shortcomings that may occur at all levels of this model.

EXPERT REVIEW DETAILS

Name* :______________________________

Age : ______________________________

Gender : Male ☐ Female ☐

Highest education level* :______________________________

Years of Working Experience* :______________________________

Signature & Stamp* :______________________________

ITEMS TO REVIEW

Based on the attached figures of **proposed appropriate structure, phases and navigation** (as depicted in the attached), please tick (√) your choice.

	Items	All are relevant	Some are definitely not relevant	Some may not relevant	Total
Q1	**Relevancy of proposed model phases**				
1.1	Model main phases				

Universiti Utara Malaysia

1.2	Linking between Phases				
1.3	DV Requirements Gathering				
1.4	DV Model Development				
1.5	DV Model Presentation				
1.6	Content Structure				
1.7	theory				
Q2	**Relevancy of proposed model phases and their components**				
DV Requirements Gathering					
2.1	Organization and Business Requirements				
2.2	Data Sources requirements				
2.3	Infrastructure requirements				
2.4	Linking between components				
DV Development					
2.5	Data Preparation and connection				
2.6	Data manipulation and management				
2.7	GODV approach				
2.8	Linking between components				
DV Presentation					
2.9	View data sources				
2.10	View virtual table				
2.11	View reports				
2.12	View virtual data mart				
2.13	View report based on virtual data marts				
2.14	Publishing virtual table				
2.15	Linking between components				
Q3	**Items**	**Need very details explanation**	**Needs some explanation**	**Is easy to understand**	**Total**
1	**Clarity of terminology**				
3.1	Model main phases				
3.2	Linking between Phases				
3.3	DV Requirements Gathering				
3.4	DV Model Development				
3.5	DV Model Presentation				

Universiti Utara Malaysia

3.6	Content Structure				
3.7	Theory				
	Data Virtualization Requirements Gathering				
3.8	Organization and Business Requirements				
3,9	Data Sources Requirements				
3,10	Infrastructure Requirements				
3.11	Requirements Specifications				
3.12	Linking between components				
	Data Virtualization Development				
3.13	Data Preparation & Connection				
3.14	Data Manipulation & Management				
3.15	GODV approach				
3.16	Linking between components				
	Data Virtualization Presentation				
3.17	View data sources				
3.18	View Virtual Tables				
3.19	View Virtual data mart				
3.20	View report based on virtual table				
3.21	View report based on virtual data mart				
3.22	Publishing virtual table				
3.23	View virtual table based on GODV approach				
3.24	View virtual data mart based on GODV				
3.25	Create and view reports based on GODV approach (virtual table)				
3.26	Create and view reports based on GODV approach (virtual data marts)				
3.27	Linking between components				

Interpretations of the scales: 4= Strongly Agree, 3= Agree, 2= Disagree, 1= Strongly Disagree

	Question	4	3	2	1
4.	I found that terminologies used are easy to understand				
5.	I found the connections and flows of all components are logically appropriate				
6.	I found that proposed phases and components will guide designers to develop of data virtualization system in Business Intelligence				

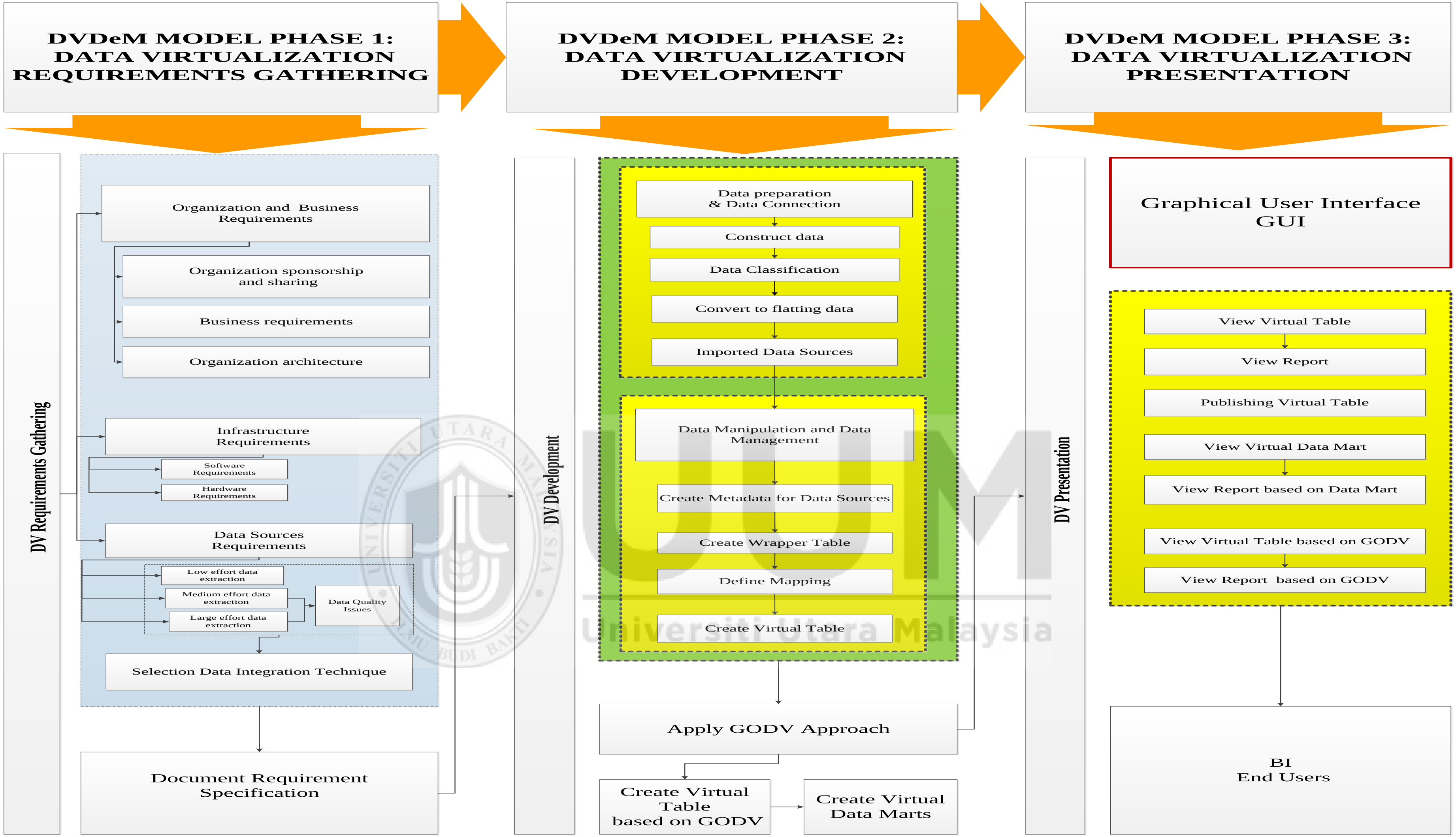
Universiti Utara Malaysia

7.	Generally, I found that the proposed appropriate design data virtualization model are readable and understanding				

Please write your further comments below:

Thanks for your usual cooperation
Ayad Hameed Mousa Albadri

Universiti Utara Malaysia



The Proposed Model

Universiti Utara Malaysia

Appendix B

Focus Group Discussion Form

INSTRUMENT FOR FOCUS GROUP DISCUSSION:

DATA VIRTUALIZATION DESIGN MODEL FOR NEAR- REAL TIME DECISION MAKING IN BUSINESS INTELLIGENCE ENVIRONMENT

Dear Prof. / Dr. / Sir / Ma,

I am Ayad Hameed Mousa Albadri who is currently pursuing his PhD study in Multimedia at University Utara Malaysia. Firstly, I would like to thank you for your attending and listening to my presentation.

Universiti Utara Malaysia

My PhD research proposes the **Data Virtualization Design Model for Near- Real Time Decision Making in Business Intelligence Environment.** It aims is to propose an appropriate structure, layout and navigation as part of how to integrate and manage data in virtualize manner and deliver them in data consumers(end users). These proposed model will provide on-demand data and data on fly, live data (real data in real time to real users) in order to support decision makers to make better decisions.

One part of this research is to evaluate the proposed appropriate design strategies in a few dimensions as listed in the review form.

You will see the review questions give you ample opportunity to use your expertise, experiences, interest, and creativity. It would be greatly appreciated if you could complete this evaluation form.

The information supplied will be treated as confidential and will be used for the research purposes, which will be reported anonymously in academic publications.

Instructions:

The model development consist of three main phases the first one is Data virtualization requirement gathering, while the second one represent Data virtualization development, finally, the third one represent Data virtualization presentation. The first figure shows the proposed model in general while the others show each phase in details. Please read and go through them carefully. Once this is done, with the expertise you possess, please provide feedback for the entire question in the provided spaces. Finally, regarding other components in this model (Data virtualization validation and control and Data consumer feedback). Regarding Data virtualization validation and control: - In Data virtualization environments, traceability and change management must be carried out in both requirements and architectural spheres. The former is concerned with managing changes to agreed requirements and its impact to other requirements inside the same or in external documents. The Data consumer Feedback is important to overcome the shortcomings that may occur at all levels of this model.

Universiti Utara Malaysia

PARTICIPANTS DETAILS

Name* :___________________________________

Age : ___________________________________

Gender : Male [ ] Female [ ]

Highest education level* :___________________________________

Years of Working Experience* :___________________________________

Signature & Stamp* :___________________________________

ITEMS TO REVIEW

Based on the attached figures of **proposed, phases and linking between phases and their components** (as depicted in the attached), please tick (√) your choice.

	Items	**Need very details explanation**	**Needs some explanation**	**Is easy to understand**	**Total**
Q1	How clear are the terminologies in terms of :				
1.1	Model main phases				
2.1	Linking between Phases				
Q2	To what extent the understanding the DV requirements gathering in terms of :				
2.1	Organization and Business Requirements				
2.2	Data Sources Requirements				
2.3	Infrastructure Requirements				
2.4	Requirements Specifications				
2.5	Linking between them				
Q3	**To what extent the understanding the DV Development gathering in terms of :**				
3.1	Data Preparation & Connection.				
3.2	Data Manipulation & Management				
3.3	Linking between them				
Q4	**To what extent the understanding DV Presentation in terms of:**				
4.1	View data sources				
4.2	View virtual table				
4.3	View virtual data marts				
4.4	Publishing virtual table				
4.5	View virtual table based on GODV				

4.6	View virtual data marts based on GODV.				
4.7	Create and view report based on GODV (virtual table).				
4.8	Create and view report based on GODV(virtual data mart).				

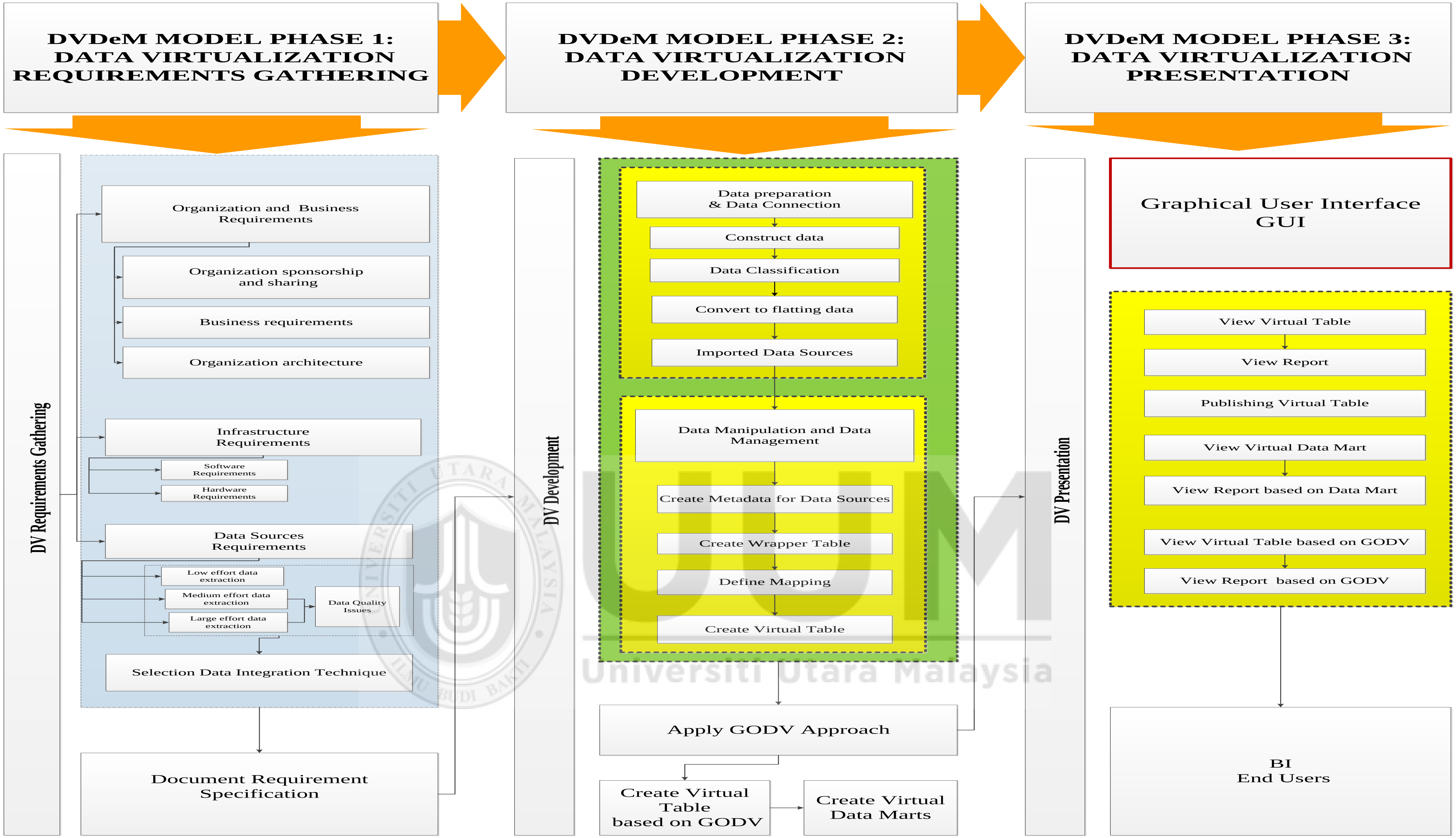
	Questions	**Strongly Agree**	**Agree**	**Dis-agree**	**Strongly Disagree**
5	I found the connections and flows of all elements are logically appropriate				
6	Generally, I found that the proposed Model are readable and understanding				
7	I found that proposed model can be used as guideline by BI developers.				

Please write your further comments below:

Universiti Utara Malaysia

Thanks for your usual cooperation

Ayad Hameed Mousa Albadri



The Proposed Model

Appendix C

Q-U Instrument

Q-U Evaluation Instrument for Measuring Usability of the DVDeM Prototype

PARTICIPANTS DETAILS

Age: ______________________________

Gender : Male [ ] Female [ ]

Highest education level* :______________________________

Years of Working Experience* :______________________________

Interpretations of the scales: Strongly Agree (5); Agree (4); neutral (3); Disagree (2); and Strongly Disagree (1).

	Q-U INSTRUMENT ITEMS		
1	**Visibility Items**		
	1	*This system can display the information (Virtual Tables, Virtual Data Mart, and BI Reports) in an uncluttered and well-structured manner.*	5 4 3 2 1
	2	*All instructions are visible and self-explanatory.*	5 4 3 2 1
	3	*Navigation options in this application such as (Links, shortcuts, home, back, forward, etc.) are displayed in visible manner.*	5 4 3 2 1

	4	*The system is able to communicate the status at all times (whether resting, processing etc.).*	5 4 3 2 1
	5	*In this system the data is concisely presented.*	5 4 3 2 1
	6	*This system has all the functions and capabilities I expect it to have.*	5 4 3 2 1
	7	*I like using the interface of this system.*	5 4 3 2 1
	8	*The interface for this system is pleasant.*	5 4 3 2 1
	9	*The organization of information in the system screen was clear.*	5 4 3 2 1
2	**Flexibility Items**		
	1	*I felt fully in control when using this system.*	5 4 3 2 1
	2	*This system has customizable feature.*	5 4 3 2 1
	3	*In this system the design for data entry is flexible.*	5 4 3 2 1
	4	*The data can be used, manipulated, and/or processed in easy manner.*	5 4 3 2 1
	5	*This system handles user-specified windows.*	5 4 3 2 1
3	**Learnability Items**		
	1	*In this system the data grouping reasonable for easy learning.*	5 4 3 2 1

Universiti Utara Malaysia

	2	*In this system promotes learnability to make it accessible for infrequent usage.*	5 4 3 2 1
	3	*In this system the grouping of menu options is logical.*	5 4 3 2 1
	4	*I believe I could become productive quickly using this system.*	5 4 3 2 1
4	**The Application Behavior Items**		
	1	*This system enhances user efficiency through a consistently rapid response rate.*	5 4 3 2 1
	2	*The system behavior is consistent.*	5 4 3 2 1
	3	*The information was effective in helping me complete the tasks and scenarios.*	5 4 3 2 1
	4	*It was easy to find the information I needed.*	5 4 3 2 1
	5	*Whenever I made a mistake using the system, I could recover easily and quickly.*	5 4 3 2 1
	6	*I was able to complete the tasks and scenarios quickly using this application.*	5 4 3 2 1
5	**Error Control & Help**		
	1	*This system has ability for error prevention and error recovery.*	5 4 3 2 1
	2	*The information (such as online help, onscreen messages and other documentation) provided with this system was clear.*	5 4 3 2 1
	3	*The system gave error messages that clearly told me how to fix problems.*	5 4 3 2 1
	4	*The system provides a help on demand.*	5 4 3 2 1

Universiti Utara Malaysia

	5 *Overall, I am satisfied with this system.*	5 4 3 2 1
	6 *Overall, I am satisfied with how easy it is to use this system.*	5 4 3 2 1
6	**Near Real Time Decision Making**	
	1 *In this system, the knowledge sharing is allowed.*	5 4 3 2 1
	2 *This System provides the information visualization functionality (comparison charts, graphs to reveal trends etc.) to assist in decision making.*	5 4 3 2 1
	3 *In this system, the breadth and depth of the data provide sufficient coverage for all data resources.*	5 4 3 2 1
	4 *In this system, data is received on time to take suitable actions and decisions.*	5 4 3 2 1
	5 *In this system, the data is always live data and up to date.*	5 4 3 2 1
	6 *In this system, the data was cleaning, cleansing, and profiling, so there are almost no errors in the data.*	5 4 3 2 1

Universiti Utara Malaysia

1. I agree that the system based on DVDeM can help users to create the intended reports in right time. Yes ☐ No ☐
2. I agree that the system based on DVDeM can be used in business intelligence environment. Yes ☐ No ☐

Appendix D
Detail Results Of The Pilot Study

1. Reliability test of Measurement for Visibility Items

Reliability Statistics	
Cronbach's Alpha	N of Items
0.709	9

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
46.42	3.468	1.862	9

Item Statistics			
Items	Mean	Std. Deviation	N
V_Q1	5.12	.454	67
V_Q2	5.07	.252	67
V_Q3	5.18	.390	67
V_Q4	5.20	.403	67
V_Q5	5.10	.354	67
V_Q6	5.27	.446	67
V_Q7	4.98	.537	67
V_Q8	5.17	.740	67
V_Q9	5.33	.475	67

Universiti Utara Malaysia

Item-Total Statistics				
Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
V_Q1	41.30	2.417	0.597	0.761
V_Q2	41.35	3.181	0.249	0.755
V_Q3	41.23	3.029	0.210	0.766
V_Q4	41.22	3.529	-0.148	0.758
V_Q5	41.32	2.864	0.399	0.750
V_Q6	41.15	3.248	0.013	0.737
V_Q7	41.43	2.589	0.342	0.759
V_Q8	41.25	2.597	0.135	0.760
V_Q9	41.08	2.790	0.285	0.760

2. Reliability test of Measurement for Flexibility Items

Reliability Statistics	
Cronbach's Alpha	N of Items
.719	5

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
25.55	1.736	1.318	5

Item Statistics			
Items	Mean	Std. Deviation	N
FL10	5.12	.409	67
FL11	5.10	.431	67
FL12	5.18	.458	67
FL13	5.25	.438	67
FL14	4.90	.526	67

Universiti Utara Malaysia

Item-Total Statistics				
Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
FL10	20.43	1.310	.276	.740
FL11	20.45	1.221	.346	.748
FL12	20.37	1.328	.187	.733
FL13	20.30	1.303	.240	.746
FL14	20.66	1.077	.350	.737

3. Reliability test of Measurement for Learnability Items

Reliability Statistics	
Cronbach's Alpha	N of Items
0.725	4

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
20.57	1.673	1.294	4

Item Statistics			
Item	Mean	Std. Deviation	N
LR15	5.00	.426	67
LR16	5.19	.584	67
LR17	5.18	.386	67
LR18	5.19	.529	67

Item-Total Statistics				
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
LR15	15.57	1.007	.567	.750
LR16	15.37	.722	.616	.739
LR17	15.39	1.726	.199	.748

Universiti Utara Malaysia

LR18	15.37	.844	.566	.734

1. Reliability test of Measurement for Application Behavior Items

Reliability Statistics	
Cronbach's Alpha	N of Items
.771	6

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
31.03	2.908	1.705	6

Item Statistics			
Item	Mean	Std. Deviation	N
AB19	5.15	.359	67
AB20	5.31	.467	67
AB21	4.97	.651	67
AB22	5.21	.538	67
AB23	5.15	.500	67
AB24	5.24	.553	67

Universiti Utara Malaysia

Item-Total Statistics				
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
AB19	25.88	2.258	.483	.727
AB20	25.72	1.812	.697	.730
AB21	26.06	2.299	.094	.729
AB22	25.82	2.149	.298	.737
AB23	25.88	2.379	.181	.745
AB24	25.79	2.350	.149	.729

2. Reliability test of Measurement for Error Control & Help Items

Reliability Statistics	
Cronbach's Alpha	N of Items
.723	6

Item Statistics			
Item	Mean	Std. Deviation	N
EC25	4.94	.795	67
EC26	5.18	.650	67
EC27	5.45	.610	67
EC28	5.25	.560	67
EC29	5.36	.620	67
EC30	5.27	.592	67

Universiti Utara Malaysia

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
31.45	6.190	2.488	6

Item-Total Statistics				
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
EC25	26.51	4.526	.305	.745
EC26	26.27	4.260	.563	.717
EC27	26.00	5.394	.150	.740

EC28	26.19	4.492	.584	.721
EC29	26.09	4.477	.506	.743
EC30	26.18	4.089	.731	.732

3. Reliability test of Measurement for Near Real Time Decision Making Items

Reliability Statistics	
Cronbach's Alpha	N of Items
.745	6

Scale Statistics			
Mean	Variance	Std. Deviation	N of Items
31.52	4.950	2.225	6

Item Statistics			
Item	Mean	Std. Deviation	N
BI31	5.31	.583	67
BI32	5.25	.560	67
BI33	5.37	.599	67
BI34	5.13	.600	67
BI35	5.34	.478	67
BI36	5.10	.581	67

Item-Total Statistics					
Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
BI31	26.21	3.319	.608	.970	.734
BI32	26.27	3.715	.427	.754	.732
BI33	26.15	3.402	.539	.970	.738
BI34	26.39	3.393	.541	.906	.720

BI35	**26.18**	**4.270**	**.228**	**.136**	**.742**
BI36	**26.42**	**3.641**	**.438**	**.761**	**.720**

UUM
Universiti Utara Malaysia

Appendix E

The Terminologies Definitions

	Terminologies	Definitions
1	**Extract, Transform, Load (ETL).**	*Three database functions that are combined into one tool to pull data out of one database and place it into another database.*
2	**Extract.**	*The process of reading data from a database.*
3	**Transform.**	*The process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. Transformation occurs by using rules or lookup tables or by combining the data with other data.*
4	**Load.**	*The process of writing the data into the target database.*
5	**Line-of-business (LOB).**	*General term that describes the products or services offered by a business or manufacturer. A company that manufactures solid state disk drives, for example, might claim their LOB is data storage.*
6	**Operational Data Store (ODS).**	*An operational data store (or "ODS") is a database designed to integrate data from multiple sources for additional operations on the data. Unlike a master data store, the data is not passed back to operational systems. It may be passed for further operations and to the data warehouse for reporting.*
7	**Key Performance Indicator (KPI).**	*A business metric used to evaluate factors that are crucial to the success of an organization. KPIs differ per organization; business KPIs may be net revenue or a customer loyalty metric, while government might consider unemployment rates.*
8	**Comparative Analysis.**	*Comparative analysis refers to several existing models from previous studies being analyzed and*

UUM Universiti Utara Malaysia

		compared.
9	**Graphical User Interface (GUI).**	*A visual way of interacting with a computer using items such as windows, icons, and menus, used by most modern operating systems.*
10	**Evaluation Instrument (Q-U).**	*Evaluation Instrument used to measuring prototype usability in BI environment compressed of six attributes named: visibility, flexibility, learnability, application behavior, error control and help, and near real time decision making used for test usability of business intelligence prototype in business intelligence environment.*
11	**Information system (IS).**	*Any organized system for the collection, organization, storage and communication of information. More specifically, it is the study of complementary networks that people and organizations use to collect, filter, process, create and distribute data.*
12	**Artificial Intelligence (AI).**	*The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.*
13	**Online Transaction Processing (OLTP).**	*A class of software programs capable of supporting transaction-oriented applications on the Internet. Typically, OLTP systems are used for order entry, financial transactions, customer relationship management (CRM) and retail sales.*
14	**Zero-latency enterprise (ZLE).**	*Any strategy that exploits the immediate exchange of information across technical and organizational boundaries to achieve business benefit*
15	**Conceptual Graph (CG).**	*A conceptual graph (CG) is a graph representation for logic based on the semantic networks of artificial intelligence and the*

Universiti Utara Malaysia

		existential graphs of Charles Sanders Peirce. The research CGs have explored novel techniques for reasoning, knowledge representation, and natural language semantics.

UNIVERSITI UTARA MALAYSIA
ILMU BUDI BAKTI
UUM
Universiti Utara Malaysia

Appendix F
The Descriptive Statistics for Education Sector

Q-U Instrument Attributes	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Visibility Items									
Item 1	30	1	4	5	125	4.170	0.069	0.379	0.144
Item 2	30	1	4	5	122	4.070	0.046	0.254	0.064
Item 3	30	1	4	5	124	4.130	0.063	0.346	0.120
Item 4	30	1	4	5	126	4.200	0.074	0.407	0.166
Item 5	30	2	3	5	122	4.070	0.067	0.365	0.133
Item 6	30	1	4	5	128	4.270	0.082	0.450	0.202
Item 7	30	1	4	5	124	4.130	0.063	0.346	0.120
Item 8	30	2	3	5	127	4.230	0.104	0.568	0.323
Item 9	30	1	4	5	130	4.330	0.088	0.479	0.230
Flexibility Items									
Item 1	30	2	3	5	124	4.130	0.079	0.434	0.189
Item 2	30	2	3	5	123	4.100	0.088	0.481	0.231
Item 3	30	2	3	5	125	4.170	0.084	0.461	0.213
Item 4	30	1	4	5	127	4.230	0.079	0.430	0.185
Item 5	30	2	3	5	118	3.930	0.082	0.450	0.202
Learnability Items									
Item 1	30	2	3	5	119	3.970	0.076	0.414	0.171
Item 2	30	2	3	5	125	4.170	0.108	0.592	0.351

Universiti Utara Malaysia

Item 3	30	1	4	5	125	4.170	0.069	0.379	0.144
Item 4	30	2	3	5	125	4.170	0.097	0.531	0.282
Application Behavior Items									
Item 1	30	1	4	5	124	4.130	0.063	0.346	0.120
Item 2	30	1	4	5	129	4.300	0.085	0.466	0.217
Item 3	30	2	3	5	122	4.070	0.082	0.450	0.202
Item 4	30	2	3	5	125	4.170	0.097	0.531	0.282
Item 5	30	2	3	5	124	4.130	0.093	0.507	0.257
Item 6	30	2	3	5	127	4.230	0.104	0.568	0.323
Error Control & Help Items									
Item 1	30	2	3	5	120	4.000	0.117	0.643	0.414
Item 2	30	2	3	5	125	4.170	0.118	0.648	0.420
Item 3	30	2	3	5	132	4.400	0.113	0.621	0.386
Item 4	30	2	3	5	127	4.230	0.104	0.568	0.323
Item 5	30	2	3	5	130	4.330	0.111	0.606	0.368
Item 6	30	2	3	5	128	4.270	0.106	0.583	0.340
Near Real Time Decision Making Items									
Item 1	30	2	3	5	128	4.270	0.106	0.583	0.340
Item 2	30	2	3	5	127	4.230	0.104	0.568	0.323
Item 3	30	2	3	5	130	4.330	0.111	0.606	0.368
Item 4	30	1	4	5	127	4.230	0.079	0.430	0.185
Item 5	30	1	4	5	130	4.330	0.088	0.479	0.230
Item 6	30	2	3	5	123	4.100	0.111	0.607	0.369

Universiti Utara Malaysia

Appendix G
The Descriptive Statistics for Business Sector

Q-U Instrument Attributes	N	Range	Min	Max	Sum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
Visibility Items									
Item1	30	2	3	5	123	4.100	0.088	0.481	0.231
Item2	30	1	4	5	122	4.070	0.046	0.254	0.064
Item3	30	2	3	5	122	4.070	0.082	0.450	0.202
Item4	30	1	4	5	126	4.200	0.074	0.407	0.166
Item5	30	2	3	5	122	4.070	0.082	0.450	0.202
Item6	30	1	4	5	127	4.230	0.079	0.430	0.185
Item7	30	2	3	5	119	3.970	0.102	0.556	0.309
Item8	30	2	3	5	124	4.130	0.115	0.629	0.395
Item9	30	2	3	5	128	4.270	0.117	0.640	0.409
Flexibility Items									
Item1	30	2	3	5	127	4.230	0.133	0.728	0.530
Item2	30	2	3	5	131	4.370	0.131	0.718	0.516
Item3	30	2	3	5	122	4.070	0.117	0.640	0.409
Item4	30	2	3	5	129	4.300	0.128	0.702	0.493
Item5	30	2	3	5	126	4.200	0.121	0.664	0.441
Learnability Items									
Item1	30	2	3	5	121	4.030	0.102	0.556	0.309
Item2	30	2	3	5	125	4.170	0.097	0.531	0.282

Universiti Utara Malaysia

Item3	30	2	3	5	124	4.130	0.093	0.507	0.257
Item4	30	2	3	5	124	4.130	0.093	0.507	0.257
Application Behavior Items									
Item1	30	1	4	5	125	4.170	0.069	0.379	0.144
Item2	30	2	3	5	127	4.230	0.092	0.504	0.254
Item3	30	2	3	5	124	4.130	0.079	0.434	0.189
Item4	30	2	3	5	126	4.200	0.101	0.551	0.303
Item5	30	2	3	5	125	4.170	0.097	0.531	0.282
Item6	30	2	3	5	125	4.170	0.097	0.531	0.282
Error Control & Help Items									
Item1	30	2	3	5	126	4.200	0.111	0.610	0.372
Item2	30	2	3	5	129	4.300	0.119	0.651	0.424
Item3	30	2	3	5	130	4.330	0.111	0.606	0.368
Item4	30	2	3	5	124	4.13	0.115	0.629	0.395
Item5	30	2	3	5	129	4.30	0.128	0.702	0.493
Item6	30	2	3	5	127	4.23	0.124	0.679	0.461
Near Real Time Decision Making Items									
Item1	30	2	3	5	130	4.33	0.111	0.606	0.368
Item2	30	2	3	5	126	4.20	0.101	0.551	0.303
Item3	30	2	3	5	131	4.37	0.112	0.615	0.378
Item4	30	1	4	5	126	4.20	0.074	0.407	0.166
Item5	30	1	4	5	131	4.37	0.089	0.490	0.240
Item6	30	2	3	5	125	4.17	0.108	0.592	0.351

Universiti Utara Malaysia

Appendix H
The DVDeM Prototype Pseudocode

The main procedure to achieve the organization requirements:

1- Connect To the server
2- Connect to sources databases
3- Understand and analysis the all tables and attributes in order to select which attribute are should import.
4- Verify data quality issues.
5- Create wrapper table by joined three tables (business, loan , and loan applicant), however, the wrapper table named as (MST_ relation) and it's consist of ten attributes named: [NoIC],[Nama],[AccNo],[SIS_Id],[SIS_Desc],[Status],[Status_Id],[Sp],[Upd_Date],and[Tkh_Ekstrak].
6- Create virtual Tables.
7- Apply GODV approach.
8- Create virtual tables and virtual data marts based on GODV.

Pseudocode

```
Start
1. Remove incorrect data like (Null data, out of rang data) if
   found.
   For all table row;
   If the row has null value then delete row
  End if
2. Select the NoIC, Nama, AccNo, SIS_Desc , SIS_Id , Status,
Status_Id, Sp , Upd_Date , and Tkh_Ekstrak by Joining business,
loan, and loan applicant. Group them if necessary Order them by
if necessary'.
3. Create the wrapper table.
End
```

Universiti Utara Malaysia

```
CREATE VIEW [dbo].[v_PerformanceCategoriesbyBusinessSector]
AS
SELECT              COUNT(*) AS qty, LEFT(BUSNSS, 1) AS bussid, CASE WHEN
LEFT(BUSNSS, 1) = '1' THEN 'Argriculture'
                                                              WHEN
LEFT(BUSNSS, 1) = '2' THEN 'Manufacturing'

                              WHEN LEFT(BUSNSS, 1) = '3' THEN 'Wholesale'

                              WHEN LEFT(BUSNSS, 1) = '4' THEN 'Commerce'

                              WHEN LEFT(BUSNSS, 1) = '5' THEN 'Construction'

                              WHEN LEFT(BUSNSS, 1) = '6' THEN 'Services'

                              WHEN LEFT(BUSNSS, 1) = '7' THEN 'Others'

                              WHEN LEFT(BUSNSS, 1) = '8' THEN 'Transportation'

                              WHEN LEFT(BUSNSS, 1) = '9' THEN 'Others
Services'                                     END AS CATEGORIES
FROM            dbo.BUSNINFO_D
WHERE        (LEFT(BUSNSS, 1) NOT IN ('A', ' '))
```

```
GROUP BY LEFT(BUSNSS, 1)
GO
```

Therefore, the Pseudocode to create virtual tables for this organization is:

Pseudocode Virtual Table1 V1:

```
Craete virtual table V1 as SELECT COUNT(*) AS count, SIS_Desc
FROM  dbo.tbl_MST_Relation
WHERE (SIS_Desc NOT IN ('NULL', 'Blacklist', 'Komputer', 'Pendahuluan', 'Pelajaran'))
GROUP BY SIS_Desc,
/////////////////////////////////////////////////////////////////////////////
```

Pseudocode Virtual Table1 V2:

```
Create virtual table V2 as SELECT COUNT(*) AS bil, LEFT(BUSNSS, 1) AS bussid,
CASE WHEN LEFT(BUSNSS, 1) = '1' THEN 'Agriculture'
     WHEN LEFT(BUSNSS, 1) = '2' THEN 'Premises'
     WHEN LEFT(BUSNSS, 1) = '3' THEN 'Manufacturing'
     WHEN LEFT(BUSNSS, 1) = '4' THEN 'Commerce'
     WHEN LEFT(BUSNSS, 1) = '5' THEN 'Contractor'
     WHEN LEFT(BUSNSS, 1) = '6' THEN 'Services'
     WHEN LEFT(BUSNSS, 1) = '7' THEN 'Trading'
     WHEN LEFT(BUSNSS, 1) = '8' THEN 'Transportation'
     WHEN LEFT(BUSNSS, 1) = '9' THEN 'Other Services' END AS bussdesc
  FROM  dbo.BUSNINFO_D
  WHERE (LEFT(BUSNSS, 1) NOT IN ('A', ' '))  GROUP BY LEFT(BUSNSS, 1)
/////////////////////////////////////////////////////////////////////////////
```

Pseudocode Virtual Table1 V3:

```
Create virtual table V3 as SELECT TOP (100) PERCENT dbo.LEDGER_D.TRAN_AMT * -0.01 AS AMT_withPoint, LEFT(CAST(dbo.LEDGER_D.PROC_DTE AS CHAR(8)), 4) AS Tahun, SUBSTRING(CAST(dbo.LEDGER_D.PROC_DTE AS CHAR(8)), 5, 2) AS Bulan, dbo.LEDGER_D.TRN_CENT_CODE, dbo.LOAN_M.DAERAH AS pusat, dbo.LEDGER_D.[TRAN]
FROM  dbo.LEDGER_D LEFT OUTER JOIN
dbo.LOAN_M ON dbo.LEDGER_D.LOANEE# = dbo.LOAN_M.LOANEE#
WHERE (dbo.LEDGER_D.[TRAN] = N'KA') OR
         (dbo.LEDGER_D.[TRAN] = N'DA') OR
         (dbo.LEDGER_D.[TRAN] = N'KB') OR
         (dbo.LEDGER_D.[TRAN] = N'DB') OR
         (dbo.LEDGER_D.[TRAN] = N'KR') OR
         (dbo.LEDGER_D.[TRAN] = N'DR')
ORDER BY bulan, tahun
```

(Education Sector)

The main procedure in Order to achieve this requirement:

1- Connect To the server
2- Connect to sources databases
3- Understand and analysis the all tables and attributes in order to select which attribute are should import.
4- Verify data quality issues.
5- Create wrapper table by joined three tables ['RAW DATA$'] INNER JOIN lec_Information INNER JOIN lec_Assessment ON lec_Information.Lecturer = lec_Assessment.Lecturer INNER JOIN lec_Resources ON lec_Assessment.Lecturer = lec_Resources.Lecturer INNER JOIN lec_Activities ON lec_Assessment.Lecturer = lec_Activities.Lecturer ON ['RAW DATA$'].Lecturer = lec_Assessment.Lecturer.

```
CREATE VIEW [dbo].[WRAPPER_ ACCOUNTING] AS SELECT
lec_Information.Lecturer,lec_Information.Lecturer_School,lec_Inform
ation.Information, lec_Resources.Resources,
lec_Activities.Activities, lec_Assessment.Assessment
FROM lec_Assessment INNER JOIN
                    lec_Activities INNER JOIN
                    lec_Resources INNER JOIN
                    lec_Information ON lec_Resources.Lecturer
= lec_Information.Lecturer ON lec_Activities.Lecturer =
lec_Information.Lecturer ON lec_Assessment.Lecturer =
lec_Information.Lecturer
GO
```

6- Create virtual Tables.
7- Apply GODV approach.
8- Create virtual tables and virtual data marts based on GODV.

In order to know the lecturer blended or not blended, we should calculate the assignment value, information, activities, and resources. It’s difficult to calculate all in one SQL-statement. Therefore, as mentioned in Chapter 5, there are multi-level databases tables should be joined.

Universiti Utara Malaysia

Pseudocode Virtual Table1 V1:

```
CREATE VIEW [dbo].[FINAL] AS SELECT
DISTINCT(TEST_ACCOUNTING1.CourseShort), TEST6.Activities,
TEST6.Assessment, TEST6.Resources, TEST6.Information,
TEST_ACCOUNTING1.Lecturer
FROM            TEST6 INNER JOIN
                         TEST_ACCOUNTING1 ON TEST6.Lecturer =
TEST_ACCOUNTING1.Lecturer
GO
```

Pseudocode Virtual Table1 V2:

```
CREATE VIEW [dbo].[FINAL1] AS SELECT
VIRTUAL_TABLE_LAST_RESULT1.IS_BLENDED, FINAL.CourseShort
FROM            FINAL INNER JOIN
                         VIRTUAL_TABLE_LAST_RESULT1 ON
FINAL.Lecturer = VIRTUAL_TABLE_LAST_RESULT1.Lecturer
GO
```

Pseudocode Virtual Table1 V3:

```
CREATE VIEW [dbo].[virtual table-_ACCOUNTING1] AS SELECT
DISTINCT (TEST_ACCOUNTING.CourseFull),TEST_ACCOUNTING.Lecturer,
TEST_ACCOUNTING.Lecturer_School, TEST6.Information,
TEST6.Resources, TEST6.Assessment, TEST6.Activities
FROM            TEST_ACCOUNTING INNER JOIN
                         TEST6 ON TEST_ACCOUNTING.Lecturer =
TEST6.Lecturer
GO
```

Universiti Utara Malaysia