INDEXING STRATEGY FOR BIG DATA PROCESSING: A CASE STUDY OF PINGER

FATIMA BINTA ADAMU

MASTER OF SCIENCE INFORMATION TECHNOLOGY
UNIVERSITI UTARA MALAYSIA
2015

Perakuan Kerja Tesis/Disertasi

(To be substituted with signed document for this page)

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

Abstrak

Dengan adanya jumlah data yang besar yang dikumpul secara berterusan dan dikongsikan oleh organisasi, maka telah menjadi keperluan untuk memenuhi pemprosesan yang baru muncul dan mendapatkan semula keperluan yang berkaitan dengan jumlah data yang besar ini. Hal ini boleh dicapai melalui mengindeks set data dan mengurangkan pengiraan berat biasa kepada strategi pengindeksan yang terkini semasa pemprosesan jumlah set data yang sangat besar. Kajian ini mencadangkan satu strategi Pengindeksan yang dinamakan sebagai Big Data INDexing Stratergy (BIND), yang menggunakan konsep pengkomputeran selari berprestasi tinggi. BIND menyokong pengedaran data selari dan menjalankan pemprosesan dalam bentuk MapReduce. Bagi membina strategi BIND, konsep penjadualan tugas lan foster untuk pemprosesan selari digunakan. Cadangan strategi pengindeksan yang pertama kali diuji ke atas persekitaran kelompok 2-nod mempunyai pelbagai saiz set data digunakan untuk mengambil perhatian samada prestasinya bertambah baik atau merosot mengikut pertambahan saiznya. Selepas itu, ia telah diuji pada kelompok 3-nod untuk diambil perhatian prestasinya apabila bilangan sumber pengiraan ditambah. Keputusan menunjukkan BIND mengurangkan pemprosesan dan query time berbanding dengan strategi semasa. Hasil kajian mempunyai implikasi yang signifikan dalam keberkesanannya menguruskan Big Data dan memudahkan penyimpanan data serta dapatan semula maklumat untuk pengguna dan organisasi yang menguruskan Big Data ini.

Kata kunci: Big Data; Pemprosesan; Pengindeksan; MapReduce; Dapatan Semula Maklumat

Abstract

With the huge amount of data continuously accumulated and shared by individuals and organizations, it has become necessary to meet the emerging processing and retrieval requirements associated with these large volumes of complex data. This could be achieved by indexing the data sets and reducing heavy computational overhead accustomed to most current indexing strategies during processing of very large amount of data sets. This study proposed a novel Indexing strategy called Big Data INDexing Strategy (BIND), using a concept of high performance parallel computing. BIND supports parallel distribution of data and performs processing in a MapReduce fashion. To develop BIND strategy, Ian foster's task-scheduling concept for parallel processing is applied. The proposed indexing strategy was first tested on a 2-node cluster environment where varying sizes of datasets were used to note if the performance improves or declines as the size of the data increases. Subsequently, it was tested on a 3-node cluster to note the performance when the number of computation resources are increased. The results demonstrate that BIND minimizes the processing and query time as compared to the current strategy. The findings have significant implication in efficiently managing Big Data and facilitating data storage and information retrieval for users and organizations that manage Big Data.

Keywords: Big Data; Processing; Indexing; MapReduce; Information Retrieval

Acknowledgements

In the name of ALLAH, Most Gracious, Most Merciful:

"*Work; so Allah will see your work and (so will) His Messenger and the believers*;"

(The Holy Quran - AtTawbah 9:105)

All thanks and praises to Allah (SWT) for the blessings of life and for guiding me through studies and life.

My sincere appreciation goes to my able mentor and supervisor, Dr. Adib Habbal, for stirring my interest to Big Data and for diligently going through this work to make it a success.

To my dearest Parents, you are my everyday teachers. Thank you is simply not enough for your prayers, your support, your guidance, your patience, and everything you stand for me. May Allah continue to bestow wisdom upon you and shower you with immense blessings.

To all my instructors who guided me through studies and life, may God continue to guide you and reward you with more beneficial knowledge. My profound appreciation goes to the team of InterNetWorks Research Laboratory, UUM; the SLAC Laboratory, Stanford University, US; Professor Bebo White, Dr. Les Cottrell, Dr. Anjum Naveed, and the entire PingER team. Most of all, I say a big thank you to the School of Computing, Universiti Utara Malaysia.

To my family, friends, and well wishers, there is no greater support than you. Thank you for all the love and encouragement you shower on me.

My unrelented prayers and appreciation goes to my adviser, my guardian, my guru,

my counselor, my consultant, my confidant, and my number one friend - my husband, Ibrahim Abdullahi. May Allah continue to guide and reward you.

And to my dearest little one, my darling Hanan, who went through a masters degree before the age of two for mama, I LOVE YOU. May Allah grant you a blessed life and aljannat firdaus.

Dedication

To my family.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BIND	-	Big Data INDexing
GPS	-	Global Positioning System
RFID	-	Radio Frequency Identification System
IT	-	Information Technology
MB	-	Megabyte
ZB	-	Zettabyte
TB	-	Terabyte
EX	-	Exabyte
SQL	-	Structured Query Language
RDBMS	-	Relational Database Management System
IDC	-	International Data Corporation
PingER	-	Ping End-to-end Reporting
ABDC	-	Analytics and Big Data Committee
SDAV	-	Scalable Data Management, Analysis and Visualization Institute
MA	-	Measurement Agents
PC	-	Personal Computer
IEPM	-	Internet End-to-end Performance Measurement
SLAC	-	Standard Linear Accelerator Center
HDFS	-	Hadoop Distributed File System
CSV	-	Comma Separated Values
JT	-	JobTracker
RM	-	ResourceManager
TT	-	TaskTracker
NM	-	NodeManager
DN	-	DataNode
NN	-	NameNode
SNN	-	SecondaryNameNode
RAM	-	Random Access Memory
CPU	-	Central Processing Unit

NNS	-	Nearest Neighbor Search
AI	-	Artificial Intelligence
NAI	-	Non-Artificial Intelligence
LSI	-	Latent Semantic Indexing
HMM		Hidden Markov Model
SVD		Singular Value Decomposition
RDF		Resource Description Framework
GiST		Generalized Search Tree
GIN		Generalized Inverted Index
PT		Posting Tree
ET		Entries Tree
PL		Posting List
RTT		Round Trip Time
ICMP		Internet Control Message Protocol
QoS		Quality of Service
AIS		Asynchronous Index Strategy
ETL		Extract, Transform, Load
LFS		Local File System
K-V		Key-Value
K-L(V)		Key-List (Value)
UI		User Interface
SSH		Secure Shell
JAR		Java Archive
IDE		Integrated Development Environment
HQL		Hive Query Language
IS		Indexing Strategy
TTL		Time-To-Leave
IR		Information Retrieval

CHAPTER ONE
INTRODUCTION

1.1 OVERVIEW

Big Data is a term used to describe very large data sets that are of different forms or structure (complex), generated at a very high speed, and cannot be managed by traditional database management systems [6]. This definition explains the three (3) main characteristics associated with Big Data: volume, variety and velocity (3Vs), and the value that can be extracted from it is seen as a fourth characteristic (4V's) [7]. Big Data is sourced from so many end devices such as Personal Computers (PC), smart phones, Global Positioning System (GPS) devices [8], sensors, Radio Frequency Identification (RFID) devices, etc. Also, online applications such as social networks and applications that involve video streaming are great sources that generate Big Data.

According to Zhou et al. in [9], the total size of data generated will surpass 7.9 Zettabytes (ZB) by the end of 2015, and predicted to reach 35ZB in 2020. Significant interest have been taken in this area lately, because of the ubiquitous and pervasive nature of data, and the emergence of mobile and cloud computing which results in the generation of huge amounts of data. The application of Big Data analytic has been of great value in terms of decision making in areas concerned with development [10, 11], security [12], and health care [13, 14, 15]. Big Data has also been used in the prediction of natural disaster, energy consumption, economic productivity, social network (for customer or consumer interest), Information Technology (IT), business[16], and other domains. Cisco related that organizations such as Facebook, Yahoo, Google, Twitter, etc., deal with Big Data on a daily basis, and have realized the need to analyze historical data for decision making [17]. The use of so many wired and wireless devices has generated voluminous data that are to be collected, transferred, processed, The contents of the thesis is for internal user only

REFERENCES

[1] D. Bouquin. (2015, May) R and data mining. [Online]. Available: http://med.cornell.libguides.com/HINF5008

[2] A. Marco. (2012,, October.) Driving big data. driving big data.

[3] Vishnur. (2014,, March) Hadoop. [Online]. Available: https://hadoop.apache.org/

[4] a. M. R. Sivaraman, E., "High performance and fault tolerant distributed file system for big data storage and processing using hadoop." in *Intelligent Computing Applications (ICICA), 2014 International Conference on, (pp. 32-36).*, 2014, March.

[5] R. Zhang, D. Hildebrand, and R. Tewari, "In unity there is strength: Showcasing a unified big data platform with mapreduce over both object and file storage," in *Big Data (Big Data), 2014 IEEE International Conference on*, Oct 2014, pp. 960–966.

[6] J. Li, Z. Xu, Y. Jiang, and R. Zhang, "(2014, aug). the overview of big data storage and management. cognitive informatics cognitive computing (icci*cc),," in *IEEE 13th International Conference on, (pp. 510-513*, 2014.

[7] C. Liu, J. Y. Chen, L. Z. Ranjan, X. R., Zhang, C. Yang, D. Georgakopoulos, and J. Chen, "Public auditing for big data storage in cloud computing -," in *A Survey. Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on, (pp. 1128-1135).*, 2013, Dec.

[8] H. Tan, W. Luo, and L. M. Ni, "Clost: A hadoop-based storage system for big spatio-temporal data analytics." in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (pp. 2139-2143). New York, NY, USA: ACM.*, 2012.

[9] W. Zhou, C. Yuan, R. Gu, and Y. Huang, "Large scale nearest neighbors search based on neighborhood graph," in *Advanced Cloud and Big Data (CBD), 2013 International Conference on*, Dec 2013, pp. 181–186.

[10] M. Cheminod, L. Durante, L. Seno, and A. Valenzano, "On the description of access control policies in networked industrial systems," in *Factory Communication Systems (WFCS), 2014 10th IEEE Workshop on*, May 2014, pp. 1–10.

[11] A. Desai, P. Garg, and P. Madhusudan, "Natural proofs for asynchronous programs using almost-synchronous reductions," in *Proceedings of the 2014 ACM International Conference on Object Oriented Programming Systems Languages & Applications*, ser. OOPSLA '14. New York, NY, USA: ACM, 2014, pp. 709–725. [Online]. Available: http://doi.acm.org/10.1145/2660193.2660211

[12] E. Bertino and B. Samanthula, "Security with privacy - a research agenda," in *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2014 International Conference on*, Oct 2014, pp. 144–153.

[13] M. Herland, T. Khoshgoftaar, and R. Wald, "Survey of clinical data mining applications on big data in health informatics," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 2, Dec 2013, pp. 465–472.

[14] J. Patel and P. Sharma, "Big data for better health planning." in *Advances in Engineering and Technology Research (ICAETR), 2014 International Conference on, (pp. 1-5).*, 2014, Aug.

[15] M. Ali-ud-din Khan, M. Uddin, and N. Gupta, "Seven v's of big data understanding big data to extract value," in *American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the*, April 2014, pp. 1–5.

[16] W. Li, W. Changyao, H. Pengyu, and a. A. A. Kaifen, S., "Cotton area estimation using muti-sensor rs data and big plot survey in xinjiang. (pp. 1-5). aug," in *Agro-geoinformatics (Agro-geoinformatics 2014), Third International Conference on,*, 2014,.

[17] H. Nakada, H. Ogawa, and T. Kudoh, "Stream processing with bigdata: Sss-mapreduce," in *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on*, Dec 2012, pp. 618–621.

[18] A. Datta and F. Oggier, "Storage codes: Managing big data with small overheads," in *Network Coding (NetCod), 2013 International Symposium on*, June 2013, pp. 1–6.

[19] A. Desai, P. Garg, and P. Madhusudan, "Natural proofs for asynchronous programs using almost-synchronous reductions," in *Proceedings of the 2014 ACM International Conference on Object Oriented Programming Systems Languages & Applications*, ser. OOPSLA '14. New York, NY, USA: ACM, 2014, pp. 709–725. [Online]. Available: http://doi.acm.org/10.1145/2660193.2660211

[20] R. Zhang and R. Hildebrand, D.and Tewari, "In unity there is strength: Showcasing a unified big data platform with mapreduce over both object and file storage." in *Big Data (Big Data), 2014 IEEE International Conference on, (pp. 960-966).*, (2014, Oct).

[21] N. Hu, B.and Carvalho, L. Laera, and T. Matsutsuka, "Towards big linked data: A large-scale, distributed semantic data storage." in *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*, (2012).

[22] T. Chardonnens, "Big data analytics on high velocity streams," Master's thesis, University of Fribourg (Switzerland), June 2013. [Online]. Available: http://exascale.info/students_projects/Chardonnens.pdf

[23] C. Liu, J. Chen, L. Yang, X. Zhang, C. Yang, and R. Ranjan, "Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates." *Parallel and Distributed Systems, IEEE Transactions on, 25(9), 2234-2244.*, 2014, Sept.

[24] A. Alnafoosi and T. Steinbach, "An integrated framework for evaluating big-data storage solutions - ida case study," in *Science and Information Conference (SAI), 2013*, Oct 2013, pp. 947–956.

[25] R. Grunzke, R. Muller-Pfefferkorn, R. Jakel, J. Hesser, N. Kepper, M. Hausmann, J. Starek, S. Gesing, M. Hardt, V. Hartmann, J. Potthoff, and S. Kindermann, "Device-driven metadata management solutions for scientific big data use cases," in *Parallel, Distributed and Network-Based Processing (PDP), 2014 22nd Euromicro International Conference on*, Feb 2014, pp. 317–321.

[26] C. Wei-Chun, C. Yu-Jung, C. Chien-Chih, L. Der-Tsai, and H. Jan-Ming, "Optimizing a mapreduce module of preprocessing high-throughput dna sequencing data," in *Big Data, 2013 IEEE International Conference on*, Oct 2013, pp. 1–6.

[27] K. Fasolin, R. Fileto, M. Krugery, D. Kaster, M. Ferreira, R. Cordeiro, A. Traina, and C. Traina, "Efficient execution of conjunctive complex queries on big multimedia databases," in *Multimedia (ISM), 2013 IEEE International Symposium on*, Dec 2013, pp. 536–543.

[28] R. Irudeen and S. Samaraweera, "Big data solution for sri lankan development: A case study from travel and tourism," in *Advances in ICT for Emerging Regions (ICTer), 2013 International Conference on*, Dec 2013, pp. 207–216.

[29] I. Giangreco, I. Al Kabary, and H. Schuldt, "Adam - a database and information retrieval system for big multimedia collections," in *Big Data (BigData Congress), 2014 IEEE International Congress on*, June 2014, pp. 406–413.

[30] T. Gollub, M. Volske, M. Hagen, and B. Stein, "Dynamic taxonomy composition via keyqueries," in *Digital Libraries (JCDL), 2014 IEEE/ACM Joint Conference on*, Sept 2014, pp. 39–48.

[31] Y. Hu, H.and Wen, T. Chua, and X. S. Li, "Toward scalable systems for big data analytics: A technology," *Tutorial. Access, IEEE, 2, 652-687.*, (2014).

[32] G. Press. (2013,, September) A very short history of big data. a very short history of big data.

[33] D. Garlasu, V. Sandulescu, I. Halcu, G. Neculoiu, O. Grigoriu, M. Marinescu, and V. Marinescu, "A big data implementation based on grid computing," in *Roedunet International Conference (RoEduNet), 2013 11th*, Jan 2013, pp. 1–4.

[34] J. Yu, F. Jiang, and T. Zhu, "Rtic-c: A big data system for massive traffic information mining." in *Cloud Computing and Big Data (CloudCom-Asia), 2013 International Conference on, (pp. 395-402).*, (2013, Dec).

[35] S. Bansal, "Towards a semantic extract-transform-load (etl) framework for big data integration," in *Big Data (BigData Congress), 2014 IEEE International Congress on*, June 2014, pp. 522–529.

[36] ——, "Towards a semantic extract-transform-load (etl) framework for big data integration," in *Big Data (BigData Congress), 2014 IEEE International Congress on*, June 2014, pp. 522–529.

[37] X. Mo and H. Wang, “Asynchronous index strategy for high performance real-time big data stream storage.” in *Network Infrastructure and Digital Content (IC-NIDC), 2012 3rd IEEE International Conference on, (pp. 232-236).*, 2012, Sept.

[38] S. Mariyah, “Identification of big data opportunities and challenges in statistics indonesia.” in *ICT For Smart Society (ICISS), 2014 International Conference on, (pp. 32-36).*, 2014, Sept.

[39] X. Tao, F. Ge, T. Huaiyuan, Z. Hong, and L. Xinran, “Thump storage: A management and analysis system for structured big data. mechatronic sciences,,” in *Electric Engineering and Computer (MEC), Proceedings 2013 International Conference on, (pp. 2424-2427).*, (2013, Dec).

[40] K. Grolinger, M. Hayes, W. Higashino, A. L’Heureux, D. Allison, and M. Capretz, “Challenges for mapreduce in big data,” in *Services (SERVICES), 2014 IEEE World Congress on*, June 2014, pp. 182–189.

[41] A. Patel, M. Birla, and U. Nair, “Addressing big data problem using hadoop and map reduce,” in *Engineering (NUiCONE), 2012 Nirma University International Conference on*, Dec 2012, pp. 1–5.

[42] R. Mao, H. Xu, W. Wu, J. Li, Y. Li, and M. Lu, “Overcoming the challenge of variety: big data abstraction, the next evolution of data management for aal communication systems.” *Communications Magazine, IEEE, 53(1), 42-47.*, 2015, January.

[43] A. Patel and . N. U. Birla, M., “Addressing big data problem using hadoop and map reduce.” in *Engineering (NUiCONE), 2012 Nirma University International Conference on, (pp. 1-5).*, 2012, Dec.

[44] I. Foster. (1995) Designing parallel algorithm. [Online]. Available: http://www.mcs.anl.gov/~itf/dbpp/text/node19.html#SECTION02350000000000000000

[45] J. Yu, F. Jiang, and T. Zhu, “Rtic-c: A big data system for massive traffic information mining,” in *Cloud Computing and Big Data (CloudCom-Asia), 2013 International Conference on*, Dec 2013, pp. 395–402.

[46] M. K. William. (2015, August) Research methods knowledge base: Types of data. [Online]. Available: http://www.socialresearchmethods.net/kb/datatype.php

[47] R. Hu, X. Zhang, and P. Wang, “Classification and evaluation of online indexing strategies,” in *Technologies and Applications of Artificial Intelligence (TAAI), 2011 International Conference on*, Nov 2011, pp. 233–238.

[48] A. Chacon, S. Marco-Sola, A. Espinosa, P. Ribeca, and J. Moure, “Boosting the fm-index on the gpu: effective techniques to mitigate random memory access,” *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[49] A. Gani, A. Siddiqa, S. Shamshirband, and F. Hanum, "A survey on indexing techniques for big data: taxonomy and performance evaluation," *Knowledge and Information Systems*, pp. 1–44, 2015.

[50] Y. Tang and L. Liu, "Multi-keyword privacy-preserving search in personal server networks," *Knowledge and Data Engineering, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[51] F. Amato, A. De Santo, F. Gargiulo, V. Moscato, F. Persia, A. Picariello, and S. Poccia, "Semtree: An index for supporting semantic retrieval of documents," in *Data Engineering Workshops (ICDEW), 2015 31st IEEE International Conference on*, April 2015, pp. 62–67.

[52] A. Matsui, S. Nishimura, and S. Katsura, "A classification method of motion database using hidden markov model," in *Industrial Electronics (ISIE), 2014 IEEE 23rd International Symposium on*, June 2014, pp. 2232–2237.

[53] Widodo and W. Wibowo, "Improving classification performance by extending documents terms," in *Data and Software Engineering (ICODSE), 2014 International Conference on*, Nov 2014, pp. 1–5.

[54] V. Alvarez, S. Richter, X. Chen, and J. Dittrich, "A comparison of adaptive radix trees and hash tables," in *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*, April 2015, pp. 1227–1238.

[55] I. Jaluta, "Transaction management in b-tree-indexed database systems," in *Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on*, vol. 3, April 2014, pp. 1968–1975.

[56] Y. Yu, Y. Zhu, W. Ng, and J. Samsudin, "An efficient multidimension metadata index and search system for cloud data," in *Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on*, Dec 2014, pp. 499–504.

[57] S. Puri and S. K. Prasad, "A parallel algorithm for clipping polygons with improved bounds and a distributed overlay processing system using mpi," in *Cluster, Cloud and Grid Computing (CCGrid), 2015 15th IEEE/ACM International Symposium on*, May 2015, pp. 576–585.

[58] A. Eldawy and M. Mokbel, "Spatialhadoop: A mapreduce framework for spatial data," in *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*, April 2015, pp. 1352–1363.

[59] H. Xu, N. Yao, W. Hu, H. Pan, and X. Gao, "The design and implementation of image information retrieval," in *Computer Science Service System (CSSS), 2012 International Conference on*, Aug 2012, pp. 1547–1550.

[60] GiST. (2015, March) Introduction to gist. [Online]. Available: http://www.sai.msu.su/~megera/postgres/gist/doc/intro.shtml

[61] PostgreSQL. (2015, March) Gin indexes. [Online]. Available: http://www.postgresql.org/docs/9.4/static/gin-intro.html

[62] M. Rouse. (March, 2015) B-tree definition. [Online]. Available: http://searchsqlserver.techtarget.com/definition/B-tree

[63] N. Du, J. Zhan, M. Zhao, D. Xiao, and Y. Xie, "Spatio-temporal data index model of moving objects on fixed networks using hbase," in *Computational Intelligence Communication Technology (CICT), 2015 IEEE International Conference on*, Feb 2015, pp. 247–251.

[64] SQlite. (2015, March) The sqlite r*tree module. [Online]. Available: https://www.sqlite.org/rtree.html

[65] G. Bui Cong and A. Duong-Tuan, "Improving sort-tile-recusive algorithm for r-tree packing in indexing time series," in *Computing Communication Technologies - Research, Innovation, and Vision for the Future (RIVF), 2015 IEEE RIVF International Conference on*, Jan 2015, pp. 117–122.

[66] Teradata. (2013, March) Hash indexing. [Online]. Available: http://www.info.teradata.com/HTMLPubs/DB_TTU_14_00/index.html#page/Database_Management/B035_1093_111A/ch02.021.26.html

[67] Elastic. (2015, March) Inverted index. [Online]. Available: http://www.elastic.co/guide/en/elasticsearch/guide/master/inverted-index.html

[68] C. U. Press. (2015, March) Indexes. [Online]. Available: http://nlp.stanford.edu/IR-book/html/htmledition/an-example-information-retrieval-problem-1.html

[69] A. Kucharik. (2015,, March) What is ping? what is ping?

[70] I. Areni, S. Tsuzuki, and Y. Yamada, "Packet size optimization of pps based radiation detection for aee-plc," in *Power Line Communications and Its Applications (ISPLC), 2012 16th IEEE International Symposium on*, March 2012, pp. 47–51.

[71] D. N. Quang, O. H. See, D. V. Nga, L. L. Chee, C. Y. Xuen, and e. a. Karuppiah, S., "Customized ping tool for smart grid communication network testing." in *Advanced Computer Science Applications and Technologies (ACSAT), 2012 International Conference on, (pp. 223-227).*, 2012, Nov.

[72] W.-C. Yang, J.-D. Jhan, D.-Y. Chen, K.-H. Lai, and R.-R. Lee, "Quality of service test mechanism and management of broadband access network." in *Network Operations and Management Symposium (APNOMS), 2014 16th Asia-Pacific, (pp. 1-4).*, 2014, Sept.

[73] P. Team. (2015,, March) Pinger ping end-to-end reporting. pinger ping end-to-end reporting.

[74] S. Team. (2015, March) Slac history. slac history.

[75] W. Jamal, A.and Pradani, N. Hasanati, A. Supriyanto, and R. Pujianto, "Scalability of dna sequence database on low-end cluster using hadoop. (pp. 50-55)." in *Information Technology Systems and Innovation (ICITSI), 2014 International Conference on,*, (2014, Nov).

[76] A. Babenko and V. Lempitsky, "The inverted multi-index," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 6, pp. 1247–1260, June 2015.