A NEW AUDITING MECHANISM FOR OPEN SOURCE NOSQL DATABASE – A CASE STUDY ON OPEN SOURCE MONGODB DATABASE

HANY HEIDAR HUSSEIN MOHAMED

MASTER OF SCIENCE (INFORMATION TECHNOLOGY)  
SCHOOL OF COMPUTING  
COLLEGE OF ARTS AND SCIENCES  
UNIVERSITI UTARA MALAYSIA  
2015
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


Kata Kunci: data big, NoSQL, MongoDB, MongoDB pengauditan
Abstract

MongoDB as a NoSQL database management system is relatively new on the database market and it is used in many important projects and products. Security analysis for MongoDB revealed that it doesn’t provide any facilities for auditing actions performed in the database. Recently, MongoDB company tried to rectify the auditing gap by providing MongoDB new enterprise version 2.6 (8th of April 2014). The auditing system logs operations information including; schema data definition language operations and operations related to replica set in addition to operations of authentication and authorization, and eventually general operations. But unfortunately still cannot record Data Manipulation Language (DML). Thus, this study aims to improve the auditing functionality in MongoDB by presenting a new mechanism for auditing NoSQL MongoDB database to include Data Manipulation Language (DML)/ CRUD (Create, Read, Update and delete) operations.

Keywords: Big data, NoSQL, MongoDB, MongoDB auditing
Acknowledgement

All praises and thanks due to Almighty Allah, the most gracious and the most merciful for lightening my way throughout the completion of this valuable thesis. I adore His benevolence and mercy, without his kindness, I will not be able to complete this study especially as I was thousand miles away from my beloved country (Egypt). Also I would like to thanks to my Supervisor Dr. Massudi Mahmuddin. Without his patient support, enlightened guidance, it is impossible for me to complete and enhance the quality of my work.

My deepest and heartfelt gratitude, loves, thanks and appreciation for my dearest parents and my beloved siblings who are a part of my happiness, success, and the inspiration that led me for the quest for knowledge and self-empowerment through night and day. I hope I can put a smile on their faces for giving back their remendous support and encouragement, patience, unconditional love, and prayers for me. Thank you for giving me the strength to chase and reach my dreams.

Thank You All.

“This Thesis is only the beginning of my journey.”

HANY HEIDAR HUSSEIN MOHAMED
UUM University, Kedah, Malaysia
Monday, January 12, 2015
# Table of Contents

Permission to Use ..................................................................................................................i
Abstrak ..................................................................................................................................ii
Abstract .................................................................................................................................iii
Acknowledgement ..................................................................................................................iv
Table of Contents ....................................................................................................................v
List of Tables ............................................................................................................................vii
List of Figures ..........................................................................................................................viii
List of Abbreviations ...............................................................................................................ix

## CHAPTER ONE : INTRODUCTION .................................................................................1

1.1 Background ......................................................................................................................1
1.2 Problem Statement .........................................................................................................3
1.3 Research Questions .........................................................................................................5
1.4 Research Objectives .......................................................................................................5
1.5 Research Scope ..............................................................................................................6
1.6 Contributions ..................................................................................................................6
1.7 Report Organization .......................................................................................................6

## CHAPTER TWO: LITERATURE REVIEW .......................................................................7

2.1 NoSQL Database ...........................................................................................................7
  2.1.1 Overview ..................................................................................................................7
  2.1.2 NoSQL Data Model .................................................................................................7
    2.1.2.1 Key-value Data Model ......................................................................................11
    2.1.2.2 Column Data Model .......................................................................................11
    2.1.2.3 Document Data Model ...................................................................................12
    2.1.2.4 Graph Data Model .........................................................................................13
  2.1.3 Importance of NoSQL in Big Data Applications .......................................................13
  2.3 NoSQL Database Security Issues ................................................................................14
    2.3.1 Threats Posed By Distributed Environments ......................................................15
    2.3.2 Safeguarding Integrity .........................................................................................15
    2.3.3 Communication between Nodes .........................................................................15
    2.3.4 Sharded Data/Fragmented Data .........................................................................16
    2.3.5 Compromised Clients .........................................................................................16
    2.3.6 Protection of Data at Rest ....................................................................................17
    2.3.7 Challenges in Enforcing Access Control ............................................................17
    2.3.8 Administrative Data Access ...............................................................................17
    2.3.9 Configuration and Patch Management ..................................................................18
CHAPTER THREE: RESEARCH METHODOLOGY ........................................... 37
3.1 Introduction .......................................................................................... 37
3.2 Research Methodology ........................................................................ 37
   3.2.1 Stage 1: Identifying MongoDB auditing features ........................................ 38
      3.2.1.1 Analysis of Logging Techniques in MongoDB ...................................... 39
      3.2.1.2 MongoDB Auditing Features .............................................................. 41
      3.2.1.3 Auditing Events and Filters ............................................................... 41
   3.2.2 Stage 2: Develop Auditing Mechanism ............................................... 44
   3.2.3 Stage 3: Evaluation ........................................................................... 51
3.3 Conclusion ............................................................................................ 53

CHAPTER FOUR: RESULTS AND EVALUATION ....................................... 54
4.1 Introduction .......................................................................................... 54
4.2 Auditing Mechanism Prototype ............................................................ 54
4.3 The Experiment Results ....................................................................... 56
   4.3.1 Results of Auditing the CRUD/DML Operations ..................................... 61
4.4 Performance Evaluation ......................................................................... 62
4.5 Summary ............................................................................................... 75

CHAPTER FIVE: CONCLUSION AND FUTURE WORK ........................... 77
5.1 Introduction .......................................................................................... 77
5.2 Conclusion ........................................................................................... 77
5.3 Limitations ........................................................................................... 78
4.4 Future Work ......................................................................................... 79

REFERENCES ............................................................................................. 80

APPENDIX A ............................................................................................. 87
List of Tables

Table 2.1: List of companies using NoSQL database with its categories ........................................8
Table 2.2: NoSQL data stores ...........................................................................................................9
Table 2.3: MongoDB vs SQL terms ..................................................................................................23
Table 2.4: Auditing types and descriptions .......................................................................................26
Table 2.5: Auditing in NoSQL databases ..........................................................................................28
Table 2.6: Examples of DML/CRUD operations .................................................................................30
Table 2.7: Sample of MongoDB related work ......................................................................................33
Table 3.1: Description of event message fields .................................................................................40
Table 3.2: MongoDB auditing system records the following actions .................................................41
Table 3.3: DML/CRUD operations: compare MongoDB vs MySQL ..................................................51
Table 3.4: Description of data sets used in the evaluation stage .......................................................52
Table 4.1: Time of the select operations in both data sets (ms) ........................................................63
Table 4.2: Time of the insert operations in both data sets (ms) ..........................................................66
Table 4.3: Time of the remove/delete operations in both data sets (ms) ..........................................69
Table 4.4: Time of the update operations in both data sets (ms) .......................................................64
List of Figures

Figure 2.1: MongoDB architecture ................................................................. 24
Figure 3.1: Stages of research methodology .................................................. 35
Figure 3.2: MongoDB security architecture .................................................... 38
Figure 3.3: Sample of MongoDB auditing system messages ............................... 39
Figure 3.4: The architecture of auditing mechanism ........................................ 42
Figure 3.5: MongoDB Database ...................................................................... 43
Figure 3.6: Sample of audit trail records ......................................................... 44
Figure 3.7: Sample of data recorded in MongoDB database. ............................ 47
Figure 3.8: MongoDB auditing algorithm flowchart ......................................... 49
Figure 3.9: Libraries used in the C# code ....................................................... 50
Figure 4.1: Auditing mechanism prototype .................................................... 55
Figure 4.2: Fields of prototype records .......................................................... 55
Figure 4.3: Details field of the prototype output ............................................. 55
Figure 4.4: Data Set 1 “AuditData” ................................................................. 56
Figure 4.5: Sample of data in “Department” collection in data set one “AuditData” .57
Figure 4.6: Sample of data in “Employee” collection in data set one “AuditData” ... 57
Figure 4.7: Sample of data in “Department” collection in data set one “AuditData” .57
Figure 4.8: Dataset 2 “AuditData2” description ............................................... 57
Figure 4.9: Sample of data in “Course” collection in data set two “AuditData2” ..... 58
Figure 4.10: Sample of data in “Lecturer” collection in data set two “AuditData2” .... 58
Figure 4.11: Sample of data in “Student” collection in data set two “AuditData2” .... 58
Figure 4.12: Auditing of the query (select) operation for data set 1 ................... 59
Figure 4.13: Auditing of the update operation for data set 1 ............................ 59
Figure 4.14: Auditing of the insert operation for data set 1 .............................. 60
Figure 4.15: Auditing of the remove (delete) operation for data set 1 ............... 62
Figure 4.16: Auditing of the query (select) operation for data set 2. .................................61
Figure 4.17: Auditing of the update operation for data set 2. ........................................61
Figure 4.18: Auditing of the insert operation for data set 2. ...........................................61
Figure 4.19: Auditing of the remove operation for data set 2. .........................................62
Figure 4.20: Time of the select operation for Data Set 1 before and after applying the
proposed auditing mechanism.......................................................................................64
Figure 4.21: Time of the select operation for Data Set 2 before and after applying the
proposed auditing mechanism.......................................................................................65
Figure 4.22: Time of the insert operation for Data Set 1 before and after applying the
proposed auditing mechanism.......................................................................................67
Figure 4.23: Time of the insert operation for Data Set 2 before and after applying the
proposed auditing mechanism.......................................................................................68
Figure 4.24: Time of the remove/delete operation for Data Set 1 before and after applying
the proposed auditing mechanism...............................................................................70
Figure 4.25: Time of the remove/delete operation for Data Set 2 before and after applying
the proposed auditing mechanism...............................................................................71
Figure 4.26: Time of the update operation for Data Set 1 before and after applying the
proposed auditing mechanism.....................................................................................73
Figure 4.27: Time of the update operation for Data Set 2 before and after applying the
proposed auditing mechanism.....................................................................................74
List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACID</td>
<td>Atomicity, Consistency, Isolation, Durability</td>
</tr>
<tr>
<td>BSON</td>
<td>Binary JavaScript Object Notation</td>
</tr>
<tr>
<td>CRUD</td>
<td>Create Read Update Delete</td>
</tr>
<tr>
<td>DDL</td>
<td>Data Definition Language</td>
</tr>
<tr>
<td>DML</td>
<td>Data Manipulation Language</td>
</tr>
<tr>
<td>DBA</td>
<td>Database Administrators</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>NoSQL</td>
<td>Not Only SQL</td>
</tr>
<tr>
<td>RBAC</td>
<td>Role Based Access Control</td>
</tr>
<tr>
<td>RDMS</td>
<td>Relational Database Management System</td>
</tr>
<tr>
<td>RFID</td>
<td>Radio Frequency Identification</td>
</tr>
<tr>
<td>RPC</td>
<td>Remote Procedure Call</td>
</tr>
<tr>
<td>OS</td>
<td>Operating System</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>TCP/IP</td>
<td>Transmission Control Protocol /Internet Protocol</td>
</tr>
</tbody>
</table>
CHAPTER ONE
INTRODUCTION

1.1 Background

The term NoSQL is used first time by Mr. Carlo Strozzi (1998) to name his lightweight open source relational database. The system did not expose the standard SQL (Structure Query Language) interface. There is a series of database following NoSQL (Not Only SQL) standards. The term “Not Only SQL” is also used for these databases that provide storage and retrieval mechanism with less constrained consistency models than traditional relational databases (Mohamed, Altrafi, & Ismail, 2014).

The last three decades were ruled by the traditional relational database management systems such as DB2, MS SQL Server and Oracle (Bonnet, Laurent, Sala, Laurent, & Sicard, 2011). They have the standard SQL. Due to the growing web scale applications such as Facebook, mobile applications and RFID (Radio Frequency Identification) the Internet has become an essential part of the world today.

Everyday zettabytes of data are being generated due to these applications. Due to changing need of applications and databases, the traditional relational databases are proved to be weak in distributed environment. This made NoSQL databases to get importance and preference. Being schema free, elastic and scalable, NoSQL databases appeared to be effective (Kanade, Gopal, & Kanade, 2013).
The contents of the thesis is for internal user only
REFERENCES


