

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**EVENT BASED TRANSIENT NOTIFICATION
ARCHITECTURE AND NoSQL SOLUTION FOR
ASTRONOMICAL DATA MANAGEMENT**

A thesis presented in partial
fulfilment of the requirements
for the degree of

**DOCTOR OF PHILOSOPHY
IN COMPUTER SCIENCE**

at Massey University,
Albany (Auckland), New Zealand.



MASSEY UNIVERSITY
NEW ZEALAND

Yu Zhao

2015

Abstract

This thesis describes the development of a messaging system for astrophysical transient event notifications, and explores the use of NoSQL databases for astronomical data management. The Microlensing Observations in Astrophysics (MOA) Project is used as a case study for which respective implementations are provided.

The Java Message Service (JMS) was chosen as the most suitable message implementation. Full implementation details of this new generation transient event notification architecture are described. It was found that the JMS-based architecture can provide a high performance on data delivery and could be up-scaled to meet high event rate situations.

For NoSQL databasing, it was found that using a document-oriented database for astronomical data management is much easier than traditional relational databases, and significantly improves overall data access performance. In addition, a NoSQL database can be very easy to scale up in order to meet demands of large workload, and the scaling approach is very cost effective. Full design details of using a document-oriented NoSQL database for astronomical data management are presented in this thesis.

The approaches developed in this research will provide scalable and efficient solutions for resolving similar computational challenges in astronomy. There are great opportunities to further extend this research and to make additional contributions to the field.

Acknowledgements

I would like to thank my supervisor Ian Bond and my co-supervisor Winston Sweatman for their guidance over these four years. This thesis could not have been completed without their support, encouragement and help. I enjoyed every meeting and discussions of my research with them; they give me lots of fun during my PhD life.

I would also like to thank everyone at the Institute of Natural and Mathematical Sciences for their inspirational seminars, discussions, suggestions and helps.

My family is very important to me and I could not complete this thesis without their continued support. Special thanks to my mother for her sacrifices over the years in New Zealand. I would also like to especially thank my father for his support.

Table of Contents

Abstract	1
Acknowledgements	2
List of Figures	v
List of Tables	vii
Chapter 1	1
Introduction	1
1.1 Time critical astrophysics.....	1
1.2 Gravitational Microlensing	2
1.3 The MOA project and computational challenges	7
1.4 Gaps in computational knowledge in the astronomy domain	10
1.5 Objectives of the research and research questions.....	12
1.6 Research Scope	13
1.7 Research Contributions.....	14
Chapter 2	16
Microlensing and Other Time Critical Astrophysics Projects	16
2.1 MACHO collaboration and EROS collaboration	16
2.2 OGLE collaboration	18
2.3 Follow up observation groups.....	23
2.4 Next generation microlensing survey projects	24
2.4.1 KMTNet (Korea Microlensing Telescope Network)	24
2.4.2 LSST (Large Synoptic Survey Telescope)	25
2.5 Other Time Critical Astrophysics Projects.....	26
2.5.1 CSS (Catalina Sky Survey)	27
2.5.2 CRTS (Catalina Real Time Survey)	28
2.5.3 WFIRST (Wide Field Infrared Survey Telescope).....	29
2.6 Conclusions	30

Chapter 3.....	31
Computational Oriented Time Critical Astrophysics Projects and Relative Techniques for Implementing Real Time Transient Notification Architecture	31
3.1 IVOA Standards	31
3.1.1 IVOA (International Virtual Observatory Alliance).....	32
3.1.2 VO (Virtual Observatory).....	33
3.1.3 VOTable.....	36
3.1.4 VOEvent	39
3.1.5 Real Time Transient Notification Projects.....	42
3.2 Polling Mechanism.....	49
3.2.1 Short Polling Approach	50
3.2.2 Long Polling Approach	51
3.3 Push Mechanism.....	53
3.4 Publisher and Subscriber Mechanism	55
3.4.1 XMPP (Extensible Messaging and Persistence Protocol).....	59
3.4.2 JMS (Java Messaging Service)	62
3.4.3 Relative works of XMPP and JMS.....	65
3.5 Message Oriented Middleware	67
3.5.1 Openfire	68
3.5.2 ActiveMQ	69
3.6 Conclusions.....	71
Chapter 4.....	72
JMS Based Real Time Transient Notification Architecture Design and Implementation	72
4.1 Related works	73
4.2 Case Study and Scenarios.....	76
4.2.1 MOA Data Flow System	78
4.2.2 MOA Scenarios.....	82
4.3 Full Design of Transient Notification Architecture.....	85
4.3.1 Overview of Transient Notification Architecture.....	85
4.3.2 JMS Type Message oriented middleware Selection	89
4.3.3 ActiveMQ Deployment.....	90

4.3.4 Topic Design and Protection	96
4.4 New Transient Information Publishing Pipeline	112
4.4.1 The Concept of the New Transient Publisher Application	112
4.4.2 Email Publishing Task	117
4.4.3 Twitter Publishing Task	119
4.4.4 ActiveMQ Publishing Task	123
4.5 Subscriber Application for Receiving New Transients' Information	127
4.6 Using the ActiveMQ Server outside the Java World	131
4.7 Performance Benchmark of Transient Notification Architecture	134
4.8 Conclusions	139
Chapter 5	142
Receiving Transient Event Notifications on Mobile Platforms	142
5.1 Choice of Mobile OS (Operating System) Platform	142
5.2 Default Notification System of Android Application and Limitations	143
5.3 Connecting the ActiveMQ in the Android Application	146
5.4 Design of the Android Application for Receiving Alerts of Transient Events	149
5.4.1 Features of the User Interface	150
5.4.2 Handling Notifications	153
5.4.3 Network Issues	156
5.5 Performance Benchmark and Scalabilities Discussion	158
5.6 Conclusions	162
Chapter 6	163
NoSQL Database for Astronomical Data Management	163
6.1 Related Work	164
6.2 Data Management Approaches: File System vs. Relational Database System vs. NOSQL Database System	168
6.2.1 File System	168
6.2.2 Relational Database System (RDBMS)	171
6.2.3 NoSQL Database System	176
6.3 Suitable NoSQL Category for Transient Data Storage and Management	181
6.3.1 Discussion of Different NoSQL Categories for Storing Transient Data	181

6.3.2 Choice of Document Stores Implementation	183
6.4 Design & Implementation of NoSQL Based Astronomical Database	188
6.4.1 Database Structure Design	189
6.4.2 BSON Document Design	191
6.5 Database Querying Scenarios	204
6.5.1 Data Exploring Scenarios.....	204
6.5.2: Use Case of Alerting New Transient Event Notifications	217
6.6 Performance Measurements and Scalability Discussion	221
6.6.1 Performance Measurements of MongoDB.....	221
6.6.2 Additional Discussion of MongoDB Scalability.....	231
6.7 Conclusions	235
Chapter 7	236
Conclusions and Future Developments	236
7.1 Future development for real time notification architecture	236
7.2 Future work for the mobile application.....	237
7.3 Future development for the astronomical database.....	239
7.4 Concept of using the JSON format for transient notification	240
7.5 Conclusions	243
References.....	248

List of Figures

Figure 1: Theoretical geometry of gravitational microlensing	3
Figure 2: The image of a typical microlensing light curve	5
Figure 3: The 1.8-m telescope of MOA.....	7
Figure 4: The light curve of the first microlensing event.....	19
Figure 5: Data processes flow of CRTS	29
Figure 6: Processing flow of Virtual Observatory (VO) framework.	36
Figure 7: Example of using TABLEDATA in VOTable.....	37
Figure 8: Example of using FITS in VOTable.....	37
Figure 9: Example of using BINARY in VOTable.....	38
Figure 10: Components of a VOEvent document.....	40
Figure 11: The overview and processing flows of the GCN architecture.....	45
Figure 12: Flow chart of the short polling approach.....	51
Figure 13: Flow chart of the long polling approach.....	53
Figure 14: Flow chart of the push approach	55
Figure 15: Work flow of the Pub/Sub mode.	57
Figure 16: XMPP message format	60
Figure 17: Messaging object structure of the JMS	63
Figure 18: Processes of capturing observation images in the observatory	80
Figure 19: Flow chart of the MOA data flow system	82
Figure 20: A typical high magnification microlensing event - MOA 2013-BLG-220	83
Figure 21: The VOEvent fragment of an unclear characteristic transient event.....	84
Figure 22: Workflow of transient notification architecture	88
Figure 23: Adding new supported protocols in ActiveMQ configuration file.....	92
Figure 24: JAAS authentication mechanism illustration	102
Figure 25: Example data format of JAAS security configuration files.....	102
Figure 26: The code fragments of the JAAS authentication plugin.....	103
Figure 27: User registration table schema.....	106
Figure 28: The complete workflow of the servlet component	111
Figure 29: Class diagram of new transient publisher application	113
Figure 30: Example of notification email contents.....	118
Figure 31: Flow chart diagram of the VOEventSubscriber application.....	130
Figure 32: Performance comparison between JMS and STOMP approaches	137
Figure 33: Main UI and information UI of MOA's Android application	152
Figure 34: UI of removes received event files from the Android device storage	153
Figure 35: Screenshots of display new event notifications on the Android.....	155
Figure 36: Internal work flow of the Android application for receiving transient events	156
Figure 37: Measurement results of the Android application.....	159
Figure 38: Diagram of lookup particular data from a file system.....	169
Figure 39: The database structure of MongoDB.....	191
Figure 40: New MongoDB database structure.....	199

Figure 41: Flow chart of creating an event portfolio document for an interesting object.....	214
Figure 42: Diagram for generating VOEvent packets using the event portfolio document.....	220
Figure 43: Data insertion performance of MongoDB	223
Figure 44: Internal data interaction for accessing data files in MongoDB	225
Figure 45: Data querying performance of MongoDB.....	227
Figure 46: Data deletion performance of MongoDB.	230
Figure 47: MongoDB sharding cluster architecture.....	233
Figure 48: Holding transient notification data through using the JSON format.....	240

List of Tables

Table 1: New microlensing event candidates discovered by the new EWS system	21
Table 2: Spacecrafts and onboard instrumentation of the GCN system	44
Table 3: available event streams on SkyAlert.....	48
Table 4: Full protocols of the ActiveMQ.....	70
Table 5: List of sub directories of ActiveMQ jar file that needing to be deleted.....	148
Table 6: Major database implementations of four different NOSQL categories.	180
Table 7: Existing datasets that need to be stored in MongoDB.	189
Table 8: Time consumed for inserting data with different types of data storage media.	224