Big Data Management and Analytics Framework for IoT-Enabled Smart Buildings

by

Muhammad Rizwan Bashir

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Under the supervision of

Professor Ghassan Beydoun

Associate Professor Asif Gill

University of Technology Sydney

Faculty of Engineering and Information Technology

School of Computer Science

October 2022

Certificate of Original Authorship

I, Muhammad Rizwan Bashir, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature:

Date: 19 October 2022

Acknowledgements

I have been privileged to have had Prof. Ghassan Beydoun and A/Prof. Asif Qumer Gill as my supervisors. I would like to express my sincere gratitude to my supervisors for giving me an opportunity to work with them on this research project. Their highly valuable support, coaching, encouragement, quick feedback and guidance helped me to finish this research project.

This study, of course, would not have been possible without the support of my family. I am very grateful to my parents Prof. Bashir Latif and Tasneem Bashir for their precious love, prayers, blessings, support, encouragement and care since childhood. I am thankful to my wife Tajwar and her family, my dearest sister Iram and dearest brother Usman. Most importantly, this research is dedicated to my daughter Ayra Rizwan and my son Moosa Rizwan for being a major source of motivation for finalizing this research.

I would like to express my sincere thanks to my all colleagues and friends for their valuable feedback and support.

I also wish to thank the Surround Pty Ltd for the financial support. I am thankful to the people from both the research community and the software industry experts who helped me with their valuable feedback and experience throughout this research project. I am thankful to all the reviewers for the valuable feedback and comments.

Thank you all.

Preface

This thesis research was carried out in the School of Information, Systems and Modelling and the School of Computer Science at University of Technology Sydney. The main contributions of the thesis are discussed in Chapters 4 and 5. These are primarily based on the following publications:

- Bashir, M.R. & Gill, A.Q. 2016, 'Towards an IoT Big Data Analytics Framework: Smart Buildings Systems', *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, pp. 1325-32.
- Bashir, M.R. & Gill, A.Q. 2017, 'IoT-enabled smart buildings: A systematic review', *2017 Intelligent Systems Conference (IntelliSys)*, IEEE, pp. 151-9.
- Bashir, M.R., Gill, A.Q., Beydoun, G. & Mccusker, B. 2020, 'Big Data Management and Analytics Metamodel for IoT-Enabled Smart Buildings', *IEEE Access*, vol. 8, pp. 169740-58. The second halves of Chapters 4 and 5 form part of this publication.
- Bashir, M.R., Gill, A.Q. & Beydoun, G. 2021, 'A reference architecture for IoT-enabled Smart Buildings', *Complex and Intelligent Systems*, (submitted March 2021). The first halves of Chapters 4 and 5 form part of this publication.

TABLE OF CONTENTS

Chapter 1 : Introduction 1

1.1. Background and Research Problem 2

1.1.1. Data 2

1.1.2. Internet of Things (IoT) 3

1.1.3. Smart Buildings 5

1.1.4. Big Data Management and Analytics 6

1.2. Research Problem 11

1.3. Aims, Objectives, Significance and Scope 13

1.3.1. Aims 14

1.3.2. Objectives 14

1.3.3. Significance/Contribution to knowledge 14

1.3.4. Scope 15

1.4. Research Strategy 16

1.5. The Framework 17

1.6. Applications and Users 18

1.7. Thesis Organisation 19

1.8. Summary 20

Chapter 2 : Literature Review 21

2.1. Introduction 21

2.2. SLR Scope 22

2.3. Research Method for the Literature Review 23

2.3.1. Inclusion and exclusion criteria for selection 24

2.3.2. Data source selection and search strategies 25

2.3.3. Citation and inclusion decision management 28

2.3.4. Final selection and quality assessment 28

2.3.5. Data extraction and synthesis 28

2.4. Literature Review 29

2.4.1. Challenges 32

2.4.2. Solutions 37
2.5. Research Gap 41
2.6. Summary 43
Chapter 3 : Research Method 44
3.1. Review of Research Methods 44
3.1.1. Case study 44
3.1.2. Action design research 45
3.1.3. Design science research 45
3.1.4. Rationale for choosing DSR 46
3.2. DSR Method 49
3.2.1. Problem identification 51
3.2.2. Analysis 52
3.2.3. Design 52
3.2.4. Development 53
3.2.5. Evaluation 54
3.2.6. Output 57
3.3. Research Instruments 58
3.3.1. Resources 58
3.3.2. Research Ethics 59
3.4. Summary 59
Chapter 4 : IBDMA Framework 61
4.1. IBDMA Framework 61
4.1.1. People 63
4.1.2. Process 64
4.1.3. Technology 65
4.1.4. Information 67
4.1.5. Facility 68
4.2. IBDMA Reference Architecture 68

4.2.1. Reference Architecture Implementation 70
4.2.1.1. Monitoring 71
4.2.1.2. Data Sourcing 71
4.2.1.3. Data Ingestion 74
4.2.1.4. Data Storage (Big Data Management) 75
4.2.1.5. Data Analysis (Big Data Analytics) 76
4.2.1.6. Data Visualization 76
4.2.1.7. Decision Making 80
4.2.1.8. Action: Smart Building Control 80
4.3. IBDMA Metamodel 83
4.3.1. Defining metamodel concepts and relationships from the IBDMA framework reference architecture and the use case 84
4.3.2. Step 2: Mapping similar concepts and relationships onto relevant domain metamodels and architectures 87
4.3.3. Step 3: Reconciliation of Definitions 87
4.3.4. Step 4: Designation of Concepts 88
4.3.5. Step 5: Identification of Relationships and resultant IBDMA Metamodel 89
4.3.6. Step 6: Evaluation of Metamodel 90
4.4. Summary 93
Chapter 5 : Evaluation 94
5.1. IBDMA Framework Evaluation Overview 94
5.2. IBDMA Instantiation 96
5.3. IBDMA Evaluation 101
5.3.1. Evaluation of reference architecture using practical use cases 102
5.3.2. Evaluation of metamodel using practical use cases 110
5.3.3. Evaluation based on industry expert feedback 121
5.4. Summary 133
Chapter 6 : Discussion and Conclusion 135

6.1. Research Journey and Output 135
6.1.1. Research Journey 135
6.1.2. Research Output 137
6.1.2.1. Framework elements 137
6.1.2.2. Framework design 138
6.1.2.3. Framework Evaluation: 140
6.2. Key Contributions and Publications 140
6.3. Research Limitations 141
6.4. Summary 142
6.5. Conclusion 143

LIST OF FIGURES

FIGURE 1.1 TYPES OF DATA ANALYTICS (WALKER 2014) 10
FIGURE 1.2 THE ANALYTICS PROCESS OVERVIEW (CHEN ET AL. 2015) 11
FIGURE 1.3 RESEARCH GAP AND SCOPE 12
FIGURE 1.4 RESEARCH CONTEXT AND SCOPE 15
FIGURE 1.5 RESEARCH STRATEGY 17
FIGURE 1.6 IBDMA FRAMEWORK HIGH-LEVEL OVERVIEW 18
FIGURE 1.7 THESIS ORGANISATION 20
FIGURE 2.1 DESIGN SCIENCE RESEARCH (DYBÅ & DINGSØYR 2008) 24
FIGURE 2.2 STUDY SELECTION PROCESS (ALZOUBI, GILL & AL-ANI 2016) 27
FIGURE 3.1 DESIGN RESEARCH APPROACH (HEVNER & CHATTERJEE 2010) 47
FIGURE 3.2 DSR PROCESS 51
FIGURE 3.3 PROBLEM IDENTIFICATION 52
FIGURE 3.4 ANALYSIS 52
FIGURE 3.5 DESIGN 53
FIGURE 3.6 DEVELOPMENT 53
FIGURE 3.7 EVALUATION 55
FIGURE 3.8 EVALUATION CRITERIA FOR IBDMA EVALUATION 56
FIGURE 4.1 IBDMA FRAMEWORK 62
FIGURE 4.2 IBDMA FRAMEWORK – CONCEPTUAL ELEMENTS 62
FIGURE 4.3 THE FIRST ELEMENT OF IBDMA – PEOPLE 63
FIGURE 4.4 THE SECOND ELEMENT OF IBDMA – PROCESS 65
FIGURE 4.5 THE THIRD ELEMENT OF IBDMA – TECHNOLOGY 66
FIGURE 4.6 THE FOURTH ELEMENT OF IBDMA – INFORMATION 68
FIGURE 4.7 THE FIFTH ELEMENT OF IBDMA – FACILITY 68
FIGURE 4.8 ITERATION 5 – UPDATED AND IMPROVED NEAR REAL-TIME DATA ANALYSIS AND ACTUATION ARCHITECTURE 70
FIGURE 4.9 SENSOR DATA GENERATION CODE 73
FIGURE 4.10 FLUME CONFIGURATION FILE 74

FIGURE 4.11 DATA VISUALIZATION IN CLOUDERA USING HIVE TABLES 77
FIGURE 4.12 DATA VISUALIZATION IN TABLEAU (MIN AND MAX VALUES DISPLAYED TOP AND BOTTOM) 78
FIGURE 4.13 TEMPERATURE SENSOR DATA VISUALIZATION IN MS POWER BI 78
FIGURE 4.14 FORECASTING SENSOR DATA IN POWER BI USING R 79
FIGURE 4.15 FORECASTING SENSOR VALUES USING ARIMA MODEL 79
FIGURE 4.16 NEAR REAL-TIME SENSOR DATA VISUALIZATION IN KIBANA 80
FIGURE 4.17 SMART BUILDING CONTROL MESSAGES 82
FIGURE 4.18 FLOW CHART OF ACTUATION PROCESS 83
FIGURE 4.19 DESIGNATION OF IBDMA CONCEPTS INTO METAMODEL LAYERS 92
FIGURE 4.20 IBDMA METAMODEL 93
FIGURE 5.1 IBDMA EMPIRICAL EVALUATION OVERVIEW 95
FIGURE 5.2 OUTPUT OF EVALUATION SCENARIO 1 112
FIGURE 5.3 RESULTANT IBDMA METAMODEL FOR SCENARIO 1 113
FIGURE 5.4 RESULTANT IBDMA METAMODEL FOR SCENARIO 2 114
FIGURE 5.5 IBDMA METAMODEL CONCEPTS IN TOPBRAID 115
FIGURE 5.6 IBDMA METAMODEL RELATIONSHIPS 116
FIGURE 5.7 INSTANCES OF CONCEPTS IN TOPBRAID 116
FIGURE 5.8 IBDMA METAMODEL IMPORTED IN TOPBRAID 117
FIGURE 5.9 METAMODEL EVALUATION IN TOPBRAID – EVALUATION AND VALIDATION SCENARIO 3 118
FIGURE 5.10 METAMODEL EVALUATION SCENARIO 3 VERSION 119
FIGURE 6.1 RESEARCH JOURNEY 136
FIGURE 6.2 IBDMA FRAMEWORK ELEMENTS 137

LIST OF TABLES

TABLE 1.1 RESEARCH QUESTIONS, RESEARCH AIMS AND RESEARCH OBJECTIVES 14
TABLE 1.2 APPLICATIONS AND USERS OF IBDMA FRAMEWORK 19
TABLE 2.1 FOUNDATIONAL CONCEPTS 25
TABLE 2.2 SEARCH ITEMS 27
TABLE 2.3 ASSESSMENT METHOD (ALZOUBI, GILL & AL-ANI 2016) 27
TABLE 2.4 QUALITY CRITERIA (DYBÅ & DINGSØYR 2008) 29
TABLE 2.5 SUMMARY OF SELECTION PROCESS 29
TABLE 2.6 STUDY ASSESSMENT CRITERIA 30
TABLE 2.7 BIG DATA MANAGEMENT AND CHALLENGES 37
TABLE 2.8 BIG DATA MANAGEMENT AND ANALYTICS SOLUTIONS 41
TABLE 2.9 RESEARCH CHALLENGES, GAPS AND QUESTIONS 42
TABLE 3.1 DSR STEPS 47
TABLE 3.2 RESOURCES FOR IBDMA 58
TABLE 4.1: ELEMENTS IN THE IBDMA ARCHITECTURE AND THEIR PURPOSE. 69
TABLE 4.2IBDMA CONCEPTS 84
TABLE 4.3 IBDMA METAMODEL RELATIONSHIPS 86
TABLE 4.4 IBDMA METAMODEL CONCEPTS MAPPED WITH MODEL RELATIONSHIPS FOUND IN THE LITERATURE 88
TABLE 4.5 IBDMA METAMODEL RELATIONSHIPS MAPPED WITH THE METAMODEL RELATIONSHIPS FOUND IN LITERATURE 90
TABLE 4.6 IBDMA METAMODEL RELATIONSHIP BETWEEN CONCEPTS 90
TABLE 5.1 IBDMA EVALUATION METHOD AND EVALUATION CRITERIA 95
TABLE 5.2 IBDMA REFERENCE ARCHITECTURE TOOLS DESCRIPTION 96
TABLE 5.3 INSTRUCTIONS TO CREATE THE REFERENCE ARCHITECTURE INSTANCES 99
TABLE 5.4 IBDMA METAMODEL TOOLS DESCRIPTION 100
TABLE 5.5 INSTRUCTIONS TO CREATE THE METAMODEL INSTANCES 101

TABLE 5.6 TEST CASE 01 - DETECTION OF OXYGEN LEVELS 103
TABLE 5.7 TEST CASE 02 - DETECTION OF FIRE 104
TABLE 5.8 TEST CASE 03 - DETECTION OF LUMINOSITY 105
TABLE 5.9 TEST CASE 04 - DETECTION OF PARKING SPACE USAGE 106
TABLE 5.10 TEST CASE 05 - DETECTION OF GARBAGE 106
TABLE 5.11 REFERENCE ARCHITECTURE EVALUATION RESULTS FOR EVALUATING EACH TEST CASE AGAINST THE THIRTEEN ECS 107
TABLE 5.12: METAMODEL EVALUATION RESULTS FOR EVALUATING EACH TEST CASE AGAINST THE THIRTEEN ECS 119
TABLE 5.13 PARTICIPANTS DETAILS 121
TABLE 5.14 EVALUATION CRITERIA SATISFIED BASED ON PARTICIPANTS' FEEDBACK 128
TABLE 5.15 INDUSTRY EXPERTS' EVALUATION REPORT 132
TABLE 6.1 DESCRIPTION OF FRAMEWORK ELEMENTS 138
TABLE 6.2 REFERENCE ARCHITECTURE DESIGN ITERATIONS 138
TABLE 6.3 METAMODEL LAYERS DESCRIPTION 139
TABLE 6.4 RESEARCH CONTRIBUTION AND PUBLICATIONS 140

Abbreviations, Acronyms

5G	Fifth Generation
ARIMA	Autoregressive Integrated Moving Average
AWS	Amazon Web Services
BI	Business Intelligence
COVID	Coronavirus Disease of 2019
DSM	Demand Side Management
DSR	Design Science Research
EC	Evaluation Criteria
GCP	Google Cloud Platform
HDFS	Hadoop Distributed File System
HTTP	Hypertext Transfer Protocol
HVAC	Heating Ventilation and Air Conditioning
IBDMA	Integrated Big Data Management and Analytics
IDE	Integrated Development Environment
IoT	Internet of Things
IP	Internet Protocol
JSON	JavaScript Object Notation
MQTT	MQ Telemetry Transport
RAM	Random Access Memory
RDD	Resilient Distributed Dataset
RQ	Research Question
SLR	Systematic Literature Review
TCP	Transmission Control Protocol
TTN	The Things Network
VM	Virtual Machine

Abstract

The Internet of Things (IoT) is transforming how people communicate with each other and their environment. There is an increasing interest in IoT and its applications in various domains. One of the most significant applications of IoT is in the development and implementation of smart cities. Within a smart city, buildings consume a significant portion of energy and serve as a major building block for the successful development and implementation of a smart city. In today's world, not only is the demand to build new buildings as smart buildings, but the focus is also to transform existing buildings into smart buildings. Smart buildings can have hundreds or thousands of IoT sensors which generate huge amounts of data (also known as big data). The effective management and analysis of this big data is a huge challenge. The focus of this thesis is to address this challenge of efficiently and effectively managing and analysing big data generated by IoT sensors in smart buildings.

This research proposes the IBDMA (Integrated Big Data Management and Analytics) framework to address the challenge of the effective and efficient management of big data generated by IoT sensors deployed in smart buildings. The IBDMA framework is developed using the design science research (DSR) method. The IBDMA framework consists of a reference architecture and a metamodel. The framework has five conceptual level elements namely i) people, ii) process, iii) technology, iv) information, and v) facility. The reference architecture provides an architecture for ingesting, storing, and analysing the IoT data as well as controlling various facilities of the smart building in an automated way. The metamodel provides details of all the elements within the smart building that enable the management and analysis of big data and to identify the relationship between these elements.

The proposed IBDMA framework is evaluated by industry experts using an empirical evaluation comprising practical use cases. The results of the evaluation indicate that the proposed IBDMA framework can be considered reasonable for the efficient and effective management of data generated by IoT sensors in the context of smart buildings. The evaluation results indicate that the IBDMA framework is generic and can be scaled to different organisational contexts to enable the management and analysis of big data. The IBDMA framework is intended for use in IoT, 5G, the cloud and big data professionals as well as academics as a coherent framework with a reference architecture and a metamodel for the management and analysis of data generated by IoT sensors in smart buildings.