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

LIST OF FIGURES

LIST OF TABLES

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.

Chapter 1 : Introduction

An increasing number of IoT initiatives have been proposed in recent times to improve the quality of human life. However, these initiatives pose real-time big data management and analytics (BDMA) challenges which is why these challenges have been the focus of many researchers and practitioners in recent times (Al-Sai, Abdullah & Husin 2019; Cisco 2011; Desai 2018; Jamil et al. 2018; Varma 2018). Indeed, IoT and big data sources can be found in a number of applications such as smart homes (Caragliu, Del Bo & Nijkamp 2011; Vasicek et al. 2018), smart buildings (Arnone et al. 2016; El-Shafie & Fakeih 2018), smart grids (Chen et al. 2009), transportation (Adeli & Jiang 2009), healthcare (Demirkan 2013), disaster management (Saha et al. 2017), the financial sector (Dineshreddy & Gangadharan 2016), retail management (Dlamini & Johnston 2016) and smart cities (Chourabi et al. 2012; Zhang et al. 2017). The focus of this research is to deal with the data management and analytics challenges in smart buildings. IoT sensors can be deployed in a smart building environment to continuously monitor various environmental parameters, including smoke, parking lot usage, user comfort, energy consumption, waste management and many others.

Within a smart building, the number of sensors could range from a few hundred to thousands. Big data analytics and machine learning techniques can only be effective if the data from sensors is effectively managed and is made available and ready for real-time analytics. This real-time big data needs to be extracted and ingested into a centralized location from where it can be extracted, cleaned, transformed, analysed and visualized on-demand or in real-time (Nath, Narayanaswami & Mohan 2016) to obtain useful insights, to make effective decisions, and eventually trigger alerts and actuate various controls in a smart building.

To deal with the challenges of real-time IoT BDMA, the research proposes and evaluates the IBDMA framework. The framework will enable the real-time management and analytics of IoT data coming from various IoT sensors and then provide autonomous near real-time control of smart buildings. The framework is first presented at a conceptual level with five major integrated elements: (1) people, (2) process, (3) information, (4) technology and (5) facility. Then the framework is expanded into a BDMA metamodel for IoT-enabled smart buildings. The framework is presented at the logical and physical levels by developing a BDMA architecture using a smart building use case scenario. The applicability of the IBDMA framework is demonstrated by using it for smart building experimental scenarios.

This chapter provides an introduction and background to the research. Firstly, it introduces the research topic. In the second section, it provides the background and research problem. The third section discusses the aims, objectives, significance, and scope of the research. The fourth section presents the research strategy. The fifth section provides an overview of the IBDMA framework. The sixth section highlights the applications and the users of the framework. Finally, the seventh section provides details on how this dissertation is structured.

1.1. Background and Research Problem

The research presented in this thesis has been conducted in the integrated areas of data engineering, metamodeling, and data analytics in the context of IoT-enabled smart buildings. Data engineering involves designing and building pipelines that transform and transport data into a highly usable state for end users. These pipelines must collect data from a variety of sources and store it in a single warehouse that represents it uniformly as a single source of truth. Metamodeling is the design and development of a collection of concepts within a certain domain and the relationship between these concepts. The systematic computational analysis of data or statistics is known as data analytics. It is a technique to find, interpret, and communicate important patterns in data. It also implies the use of data trends to make more informed decisions. This research focuses on the design and development of the IBDMA framework by employing the above three domains. This section discusses the research background for the IBDMA and is sub-divided into four sections. The first four sections introduce basic concepts and terminologies important to understand the rest of the thesis. This includes concepts such as data, IoT, smart buildings, and BDMA.

1.1.1. Data

A collection of information, such as numbers, facts, words, observations, measurements, or descriptions of things, is referred to as data. In scientific research, data contains a set of values of qualitative or quantitative variables about one or more persons or objects. Over the years, data in various forms and formats has been used in a variety of domains including finance, business management, scientific research, governance and in almost all forms of human organizational activity (census data, rebates data, vaccination data etc). Data is measured, ingested, stored, reported, and analysed, and is used to create data visualizations such as images, tables or graphs. Data, as a generic concept, refers to the availability of some existing information in a form that is suited for better use or processing. Due to its importance and the business value it brings, data has been described as the new oil of the digital economy.

1.1.2. Internet of Things (IoT)

IoT is a network of interconnected devices where each device is connected to a sensor that generates data about a specific entity. In the last few years, IoT has been the focus of attention both in industry and academia for several reasons. One of the primary reasons for this is that the number of IoT devices has been growing exponentially each day. IoT devices are incorporated in wearable devices, in smart homes, smart buildings and smart cities where they are generating data every second. Reports by Ericsson (Ericsson 2011) suggest that the number of IoT devices across the globe would go above 50 billion by 2020. With the advent of 5G, smart electronic appliances and the advancements in technology of interconnected devices, the number of IoT devices will reach trillions (Uusitalo 2006). The quick adoption of IoT in various sectors is linked to its ease of use, improved consumer comfort and advancements in data processing, data management and data analytics technologies.

In industry, IoT has gained immense popularity in recent times in a wide variety of sectors and has resulted in the improved comfort and safety of human beings. Some of the sectors which have benefited most and will continue to benefit because of the IoT revolution include agriculture, energy, finance, healthcare, manufacturing, transportation and logistics, hospitality and retail. IoT is facilitating human lives in a significant way in various disciplines, including transportation, automation, disaster management and healthcare. This will in turn provide opportunities for research, innovation, investment and challenges (Zeng, Guo & Cheng 2011).

The scope of IoT is not only confined to a few sectors in certain countries. In fact, the adoption of IoT is global and a number of countries have developed policies and strategies for the effective and productive implementation of IoT for the comfort, safety and health of the general public. For instance, in Japan, broadband access is helping to establish communication between people, people and things, and things and people (Srivastava 2004). New IoT research centres have been setup in different parts of the globe to address critical issues around the IoT, such as privacy, ethics, trust, reliability and security (Mendoza 2016). The arrival of the COVID19 pandemic has also increased the adoption of IoT, particularly in healthcare. Due to the COVID19 pandemic, an increasing number of people are staying at home and working from home. As a result, they are buying smart devices to stay connected as these devices provide the ability to control homes devices from anywhere in the world through the Internet and also enable people to communicate with other people. As an example, a report commissioned by the UK government suggested that almost half (49%) the UK residents purchased an IoT device since the start of the pandemic (Stannard et al. 2020). The development initiatives mentioned above are only a few examples of many to form the foundational grounds for the adoption of IoT (Xia et al. 2012). Looking at the exponential increase and adoption of IoT across the globe, it becomes clear that IoT will be the next big challenge to address in the world of communication, IT and data analytics. Clearly, there are various areas and sectors where IoT adoption is feasible but a significant fraction of IoT devices is deployed inside buildings. These buildings can include homes, residential buildings, commercial buildings (shopping malls), retail shops, industries and so on, leading us to the smart buildings concept. A number of IoT devices can be deployed in a smart build that could range from a few sensors to thousands of sensors. While the deployment of these sensors provides many benefits including the safety, security and comfort of the resident, these IoT devices, on the other hand, pose a few challenges as well. Big data is generated in a smart building where IoT devices are communicating with each other. To improve the management, safety and efficiency of smart buildings, managing and analysing the big data generated from IoT devices is critical.

The importance of IoT can also be seen in mitigating global warming and climate change. IoT devices help us monitor the environment in real-time. By monitoring the environment through IoT devices, more dynamic and robust data can be gathered to help mitigate the risk of climate change. Climate change has become the need of the hour and IoT can play its part in ensuring that we reach the net-zero emissions goal by 2050. Another area where IoT can prove beneficial is in energy policy. In developed countries, there is a steady increase of 20% to 40% in global energy consumption in buildings, including both commercial and residential buildings (United Nations 2009), which is far more than other major sectors of energy consumption. Due to this increase in the demand for energy, there are concerns over the generation and supply of energy to meet this demand. On the other hand, the burning of fossil fuels increases greenhouse gases which increases the surface temperature of the earth and adds to the global warming effect. These policies play an important role as they define the course of action for industries and business to design and construct buildings. Moreover, it helps in maintaining and operating the building more effectively. In a building, heating ventilation and air conditioning (HVAC) systems consume a major portion of the total energy consumed by the whole building. A report from Siemens suggests that in general, HVAC consumption in a building is 60% of the total building consumption (Rathore et al. 2016). IoT can help us monitor the energy consumption of the building and can help reduce the carbonisation effect of the building. Hence, the solution to these problems lies in having IoT-enabled smart buildings. These are only a few of the many advantages of transforming buildings into smart buildings and effectively utilising the data generated by IoT devices, which is the main focus of this research. The next section discusses the definition, advantages and capabilities of smart buildings.

1.1.3. Smart Buildings

Smart buildings and other smart environments which make use of IoT technology and analytics solutions provide numerous opportunities in industry and academia. The importance of smart buildings primarily lies in the value it brings to users and organisations. Smart buildings are different to digital buildings. Smart buildings provide a safer, better and more comfortable environment for its residents and its managers. A more technical definition of smart buildings is:

"Smart buildings are buildings which integrate and account for intelligence, enterprise, control, and materials and construction as an entire building system, with adaptability, not reactivity, at the core, in order to meet the drivers for building progression: energy and efficiency, longevity, and comfort and satisfaction. The increased amount of information available from this wider range of sources will allow these systems to become adaptable, and enable a smart building to prepare itself for context and change over all timescales"(Buckman, Mayfield & Beck 2014).

In simple words, smart buildings are ‘intelligent’ buildings that provide a safe, productive and comfortable environment for its residents without sacrificing operational efficiency. The intelligence is provided by sensors which are deployed inside the smart building to monitor various environmental and non-environmental parameters of the building. This is achieved by autonomous systems which are activated and de-activated based on the data received from the sensors.

With the advent of IoT sensors, the task of transforming already developed buildings and new buildings into smart buildings has become manageable. Various types of IoT sensors ranging from environmental monitoring, energy consumption monitoring, parking lots usage and many more have been tested in smart building environments and have proven to be successful. These sensors are now easily available on the market and are ready to be deployed in smart buildings. These sensors monitor the environment in which they are deployed, collect data from the environment and push the data to a centralised location where it can be analysed in real time to actuate various controls autonomously to keep the environment safe and comfortable for the residents of the building.

A smart building may have a variety of infrastructural components that efficiently maintain and manage the occupants’ comfort and safety levels. These include HVAC systems, security cameras, parking area sensors, smart meters (for electricity, gas and water), luminosity sensors, garbage detection sensors and electric vehicle charging system (Mullassery 2016). However, when considering the residents' experience, there is a need to take into account the environmental parameters inside the smart building. These environmental parameters include oxygen concentration, carbon dioxide concentration, temperature, luminosity, humidity etc. IoT sensors have been developed over the last few years to monitor these parameters inside smart buildings. These sensors are ultra-low powered and are extremely energy efficient.

This research focuses on bridging this research gap by developing a framework for the IoT BDMA for smart buildings. The framework and the associated metamodel enable the analysis of a large amount of IoT generated data with the aim of improving the residents' experience by implementing an autonomous system for actuating various controls in the smart building to keep the environment safe, healthy and comfortable for the residents. All of this relies on efficiently managing and analysing the big data generated from the IoT sensors, hence leading us to the concepts of BDMA.

1.1.4. Big Data Management and Analytics

This section explains the key BDMA terminologies and concepts required to understand the significance of this research. Big data is a challenging term to understand since it can imply different things to different individuals.

Big data is a term that refers to two major phenomena:

- the incredible speed of data generation
- enhancing the ability to store, process and analyse that data

With the innovations occurring in the information technology domain every day, the need for data collection and storage has grown tremendously in recent times. A huge amount of data is generated in our personal life and in various business sectors daily. Storing, processing and analysing the data using suitable techniques could provide new insights about our business, society and the environment, allowing us to respond quickly to emerging possibilities and changes.

However, the growth in the volume and velocity of data in today's digital world seems to outspeed the advances made in computing infrastructure. Traditional data storage and processing solutions, such as databases and data warehouses, are becoming inadequate to cope with the amount of data generated. This new challenge is known as the big data challenge. A commonly accepted definition of big data is difficult to formulate, however, some commonalities in people's understanding of big data can be combined in the form of a big data definition as follows: "big data consists of data that is impossible to collect, store and process using traditional software and hardware tools within reasonable timeframes". Based on this concept, researchers describe big data using the 5 Vs of data: volume, velocity, variety, veracity and value. Due to its importance and usefulness, BDMA has gained enormous attention in recent years. The following sections provide insights into BDMA concepts and techniques, and the tools available to manage and analyse the data.

1.1.4.1. Big data management

Big data management is about capturing, storing, organizing, and governing large volumes of both structured and unstructured data. The aim of big data management is to provide high quality data and secure accessibility to users and applications to meet business requirements. Developing, employing and incorporating an effective big data strategy in research, corporations, and both public and private sector organisations is imperative to address the ever-growing pool of data which typically involves terabytes or even petabytes of data in a variety of file formats. By having an effective data management policy or strategy, it becomes easier for organisations to standardise and govern data storage in a safe and secure manner. It also helps data consumers in locating and accessing data in different formats (unstructured, semi-structured and structured) from a variety of sources including billing data, pricing data, call logs, social media posts and many others.

With the ever-increasing volume of data that is generated from a range of sources, various platforms have emerged in the recent past to address the challenges associated with the storage, management and governance of data. The traditional way of storing data into databases and developing data warehouses still exists and plenty of use cases can be found. However, the emergence of the cloud and concepts of the data lake and delta lake have solved various problems in terms of storage space and cost. However, the challenges associated with big data are manifold. The governance of data, archiving infrequently used data and the format in which it is stored to perform the analysis quickly and efficiently are all part of the big data management strategy. It also includes strategies and processes to decide which data to keep for compliance reasons and what data to dispose of. The goal of big data management is to improve the existing business process to provide the business competitive advantage. This increase in data which is experienced by all types of organisations across the globe provides new opportunities and untapped value.

Data management is not only important for big web companies or companies who primarily deal with data but in the present world, it is a mandatory requirement for all types of businesses as almost all businesses ranging from a small retail shop to larger shopping malls generate or store data in one form or the other. When it comes to big data storage and analytics platforms, there are numerous big data platforms ranging from Hadoop upstarts to traditional database players. Some of the most frequently used include Microsoft Azure, Amazon Web Services, Google Cloud Platform, Cloudera, IBM, MAPR, Infobright, SAP, Oracle Big Data (Henschen 2014). Some of these are open-source platforms. For this research, we use Hadoop for the management and storage of big data generated by IoT sensors deployed in smart buildings.

1.1.4.2. Big Data Analytics

We discussed big data management in the previous section. But the analysis of big data is equally imperative and challenging at the same time. Analytics of big data is the process of analysing huge volumes of data generated from a variety of sources to obtain useful and eloquent insights from the data. The analysis of big data helps to reveal hidden facts, find unknown relationships, obtain market movements, find customer preferences and other meaningful evidence that can help organisations improve their business strategies.

Traditional ways of analysing data do not work efficiently when it comes to analysing big data. Traditional systems were not designed for parallel processing (Adeli & Vishnubhotla 2020) and autoscaling (Rzadca et al. 2020). Hence, if we use traditional data analysis methods for big data analysis, they fail. The advent of the latest big data analytics technologies has opened doors for data scientists, statisticians and data analysts to analyse much larger volumes of data and in a much faster processing time. This has enabled organisations to discover information from the data that was left untapped using conventional analytics tools and business intelligence (BI) programs. The latest big data tools and technologies not only enables data to be processed faster, they have also enabled us to analyse semi-structured and unstructured data for example video feeds, social media posts, documents, emails, phone call records, audio recordings, web server logs, error logs and data generated by sensors (including the IoT sensors).

Big data analytics driven by specialized hardware and software systems can lead the way to a number of business advantages, including improved customer service, new business opportunities, better decision making, improved operational performance and a competitive advantage over rivals. In summary, data analytics tools and technologies help organisations make informed decisions by analysing huge data sets and drawing insights from those data sets. Traditional BI queries can provide information about business performance and growth, but big data analytics is an advanced form of analytics which involves analysing large volumes of data for challenging applications with elements such as statistical algorithms and machine learning models powered by high-performance analytics systems.

- **Types of Big Data Analytics**

Big data analytics can be subdivided into four types. Each of these four types are briefly outlined as follows:

Prescriptive – One of the most common types of big data analytics is predictive analytics. It focuses on the steps that need to be performed to attain a given goal. Prescriptive analytics is a sort of analysis that provides results and recommendations for the next steps. In simple terms, prescriptive analytics is concerned with the present, whereas predictive analytics is concerned with the future, and descriptive analytics is concerned with the past. Prescriptive analytics has applications in various fields, including healthcare and finance.

Descriptive – The second type of big data analytics is descriptive analytics which searches back in time for answers. The fundamental aim of descriptive analytics is to find out what happened in the past or what is happening right now.

Clustering, segmentation, and statistical analysis are the approaches and instruments utilised in descriptive analytics. Descriptive analytics provides a holistic perspective of the metrics on which to base the business plan, which aids in understanding the changes that have occurred in the business.

Predictive – Predictive analytics is arguably the most critical sort of big data analytics. Predictive analytics aims to forecast future events based on previously collected and new data. It employs machine learning, data mining, and statistical modelling approaches to gain future insights. It aids in future decision-making and is critical since organisations may change their plans depending on the predictive analysis' findings. It has a wide range of uses; for example, airlines can forecast the number of passengers and change the number of flights and ticket prices accordingly.

Diagnostic – Diagnostic analytics seeks to address the question, "Why did anything happen?" It entails examining the event logs or error logs for a previously occurring incident. It employs techniques such as data mining and correlation to analyse a problem. Identifying customer trends is an example of where they can be used. On the other hand, diagnostic analytics is typically employed in more challenging situations. For instance, a system problem or failure occurred, and the cause of the issue or failure must be explored.

We use all four types of big data analytics in designing and implementing the IBDMA framework. The complexity levels of these various types of analytics vs the added value and contribution of these types is summarized in Figure 1.1.

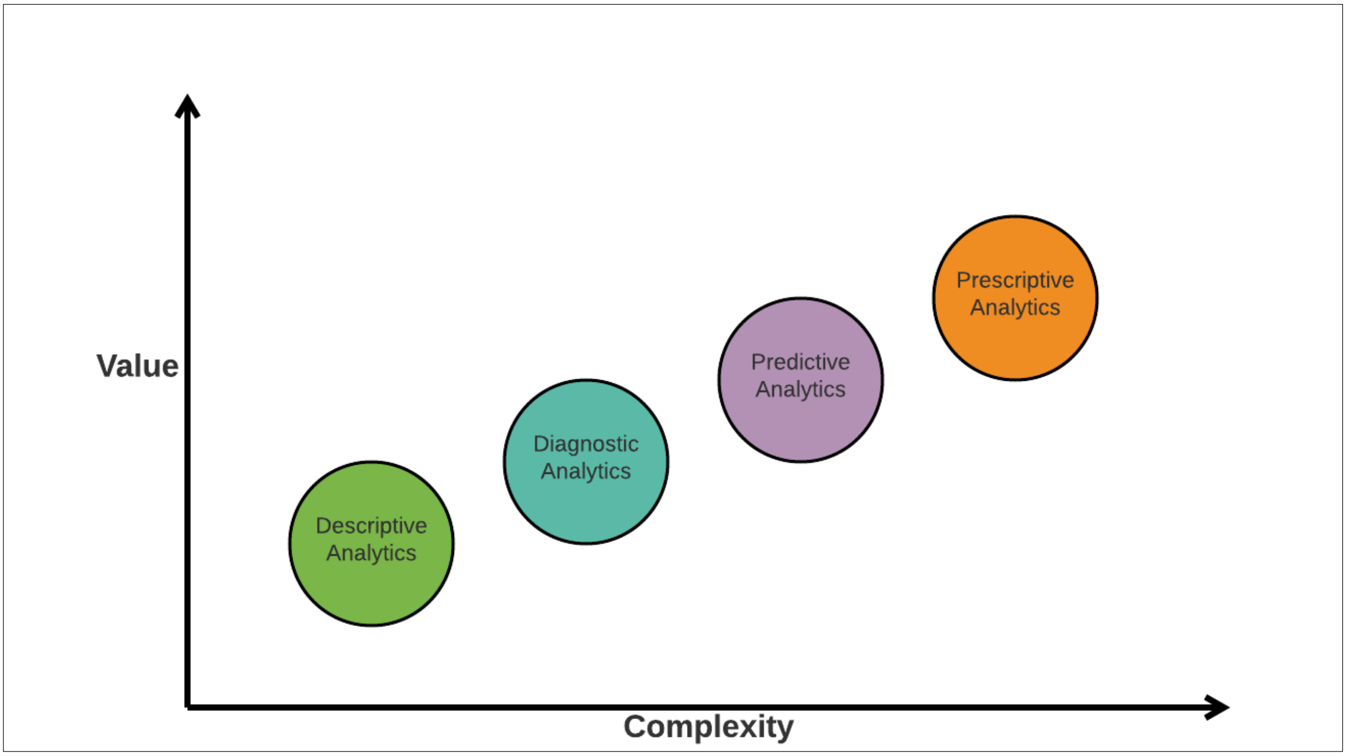


Figure 1.1 Types of Data Analytics (Walker 2014)

- **Big Data Analytics Process**

The big data analytics process involves discovering meaningful information from large sets of data stored in databases, data warehouses, data lakes or other information repositories (Chen et al. 2015). A typical data analytics process can be explained in the following three stages (see Figure 1.2).

(i) Data preparation: this consists of preparing the data for data mining process.

(ii) Data mining: this comprises applying suitable techniques and algorithms to gain knowledgeable insights.

(iii) Data presentation: this stage involves visualising and presenting data in a meaningful manner for the audience.

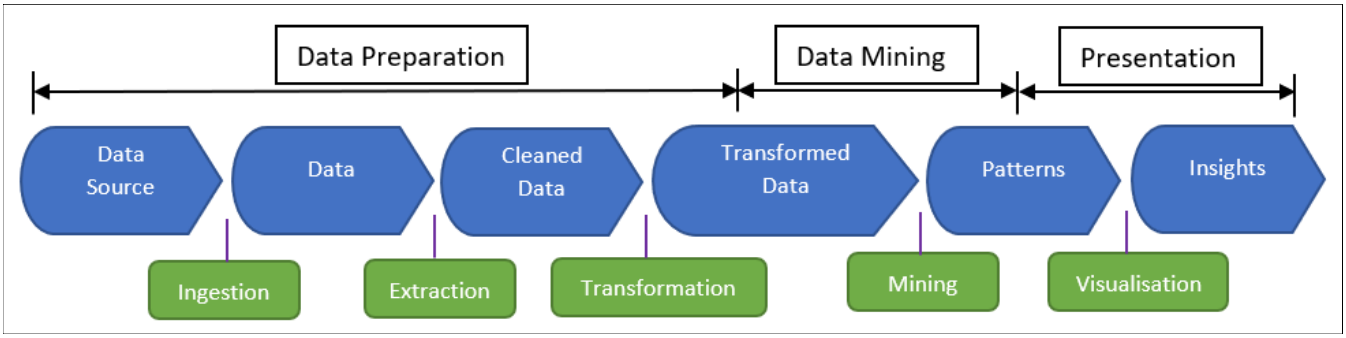


Figure 1.2 The Analytics process Overview (Chen et al. 2015)

1.2.Research Problem

This section discusses the research problem and the main research question. IoT sensors are becoming increasingly popular. With the advent of new lower power technologies and a high data transmission range in embedded sensors, data transceivers and their associated controllers, IoT is finding an increasing number of applications. With these technological advancements in the IoT domain, access to IoT-enabled smart facilities i.e., smart homes, smart buildings, smart farms, smart cities, smart devices for the common person have become a reality.

There are a lot of commercially available IoT platforms in the market to address the management of IoT sensors and the challenge of IoT big data. These include open source platforms like TTN (The Things Network) (TheThingsNetwork 2021), ThingsBoard (ThingsBoard 2021), Nodered (Node-RED 2021) etc. These also include propriety services including cloud-based IoT platform providers including AWS (Amazon Web Services) IoT, GCP (Google Cloud Platform) IoT Core and Microsoft Azure IoT Hub. Hadoop and its available use cases for IoT sensor data management and analytics is also a popular choice for many commercial organizations dealing with IoT and big data. However, each of these platforms has certain limitations as well as advantages over the others on what can be implemented on these. These platforms are intended for the implementation of data pipelines, data storage and data analysis, and we will be using one of these platforms for the implementation of some of the components of the proposed framework. None of these platforms provide the ability to design the reference architecture and a metamodel, hence, we will be using a combination of tools as explained in Chapter 4.

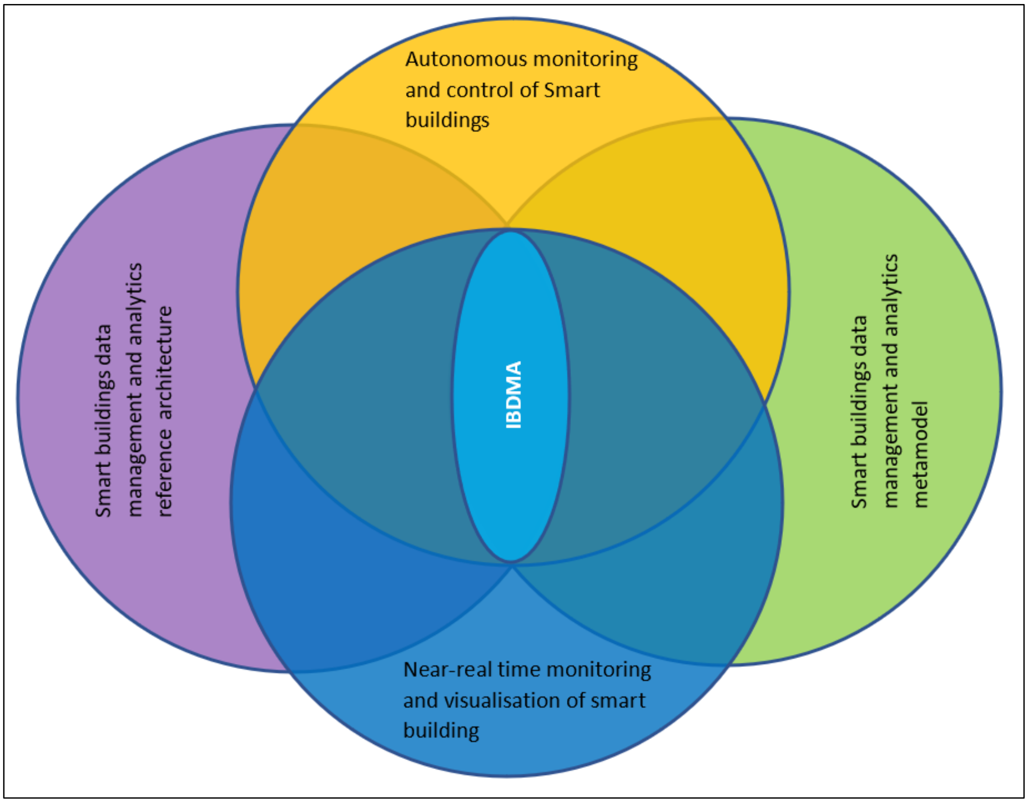


Figure 1.3 Research Gap and Scope

When developing the IoT data management and analytics architecture, ease of use, scalability and cost are the major factors to consider. The availability of tools and services for IoT management and IoT BDMA has made it easier to ingest, store and analyse data. However, there is no coherent framework which addresses the human centricity for the BDMA of IoT-enabled smart buildings while considering the autonomous control of smart facilities inside smart buildings.

One of the major aims of a smart building is user comfort, security and safety by storing and analysing data generated from the IoT sensors. In practice, it is impossible to monitor smart building environmental parameters without having these sensors. Thus, the deployment of IoT sensors in a smart building assists in monitoring the smart building's environment. The data collected from these sensors when analysed and visualized enables the administration of smart buildings to improve the environment and safety of the building. It is important to store and analyse the data generated from these sensors efficiently as the data generated from these sensors can grow exponentially and can become overly complex to manage and analyse.

To deal with the challenges of real-time big data management and analytics in the smart building context, a coherent framework which incorporates a metamodel and a reference architecture is needed.

This research is aimed at bridging this gap in the literature by proposing a framework which consists of a reference architecture and a metamodel, both linked to each other with the core conceptual elements. The reference architecture provides an end-to-end blueprint to enable real-time management and analytics of huge amounts of IoT data coming from various IoT sensors. It also intends to provide autonomous near real-time control of smart buildings by analysing, monitoring, and controlling various facilities within the smart buildings. The metamodel on the other hand provides the list of key elements in a smart building environment and how these elements interact with each other. The reference architecture and the metamodel are linked to each other through key conceptual elements. The research gap is illustrated in Figure 1.3 where it is evident that no other framework combines the capabilities of real-time monitoring, real-time autonomous control of smart buildings, a smart building reference architecture and a smart building big data management and analytics metamodel as proposed by the IBDMA framework. Hence, without the IBDMA framework, it is not possible to obtain a holistic view of all the elements required in a smart building ecosystem and missing some elements out will not result in an effective system for the big data management and analytics of smart buildings. Based on the research gap identified, the main research question is:

RQ. How to manage and analyse IoT big data for the effective management of smart buildings?

Since the above RQ is quite broad, the RQ can be subdivided into two research questions as follows:

RQ1: How to effectively manage and analyse IoT big data in a smart building context?

RQ2: How to holistically identify all the elements and the relationship between these elements to effectively manage and analyse big data in IoT-enabled smart buildings?

The first research question is focussed on developing a reference architecture, while the second research question focuses on the metamodel development as it will be further elaborated in the next section. The development of the reference architecture before the metamodel helps define the key components of the framework and to define the scope of the research.

1.3. Aims, Objectives, Significance and Scope

Based on the research problem and the research questions identified in the previous section, the aims, objectives, significance, and scope of this research are highlighted in this section. First, we define the research aims based on the research questions identified in the previous section. The research objectives are then identified based on the research questions and research aims. Table 1.1 presents the research questions, aims and objectives in a tabular form.

1.3.1. Aims

The aims of this research are:

1. to develop an IBDMA reference architecture for IoT-enabled smart buildings.
2. to develop an IBDMA metamodel for IoT-enabled smart buildings.

1.3.2. Objectives

The objectives of this research are:

1. to analyse and manage IoT sensor data efficiently for the autonomous control of facilities in smart buildings.
2. to identify all the elements and the relationship between these elements for the effective deployment of infrastructure to enable efficient BDMA in IoT-enabled smart buildings.

1.3.3. Significance/Contribution to knowledge

This research focuses on the development of a people-centric IBDMA framework which focuses on improving the comfort and safety of people by autonomously activating various controls in the smart building environment. The framework will assist researchers and IoT practitioners to store, manage and analyse IoT-generated big data in near real-time to improve residents' comfort and the energy efficiency of the smart building. The framework is implemented with the two aforementioned objectives in mind. The first objective is to analyse and manage IoT sensor data efficiently for the autonomous control of various facilities in smart buildings. The second objective is to identify all the elements and the relationship between these elements for the effective deployment of infrastructure to enable efficient BDMA in IoT-enabled smart buildings. BDMA has been the focus of many researchers in recent times, however, a coherent framework comprising a holistic reference architecture and a metamodel to address the BDMA challenges in IoT-enabled smart buildings is missing. This research aims at bridging this gap in the literature and will help researchers and practitioners to effectively manage smart environments, improve residents' comfort and improve the energy efficiency of smart environments.

Table 1.1 Research Questions, Research Aims and Research Objectives

Research Question	Research Aim	Research Objective	Contribution
How to efficiently manage and analyse IoT big data in smart buildings?	To develop an IBDMA framework reference architecture for IoT-enabled smart buildings.	Analysing and managing IoT sensor data efficiently for the autonomously control of facilities in smart buildings.	IBDMA reference architecture

How to holistically identify all the elements and the relationship between these elements to effectively manage and analyse big data in IoT-enabled smart buildings?	To develop an IBDMA metamodel for IoT-enabled smart buildings.	Identify all the elements and the relationship between these elements for the effective deployment of infrastructure to enable efficient BDMA in IoT-enabled smart buildings.	IBDMA metamodel

1.3.4. Scope

This research focuses on the development and implementation of the IBDMA framework. The research also focuses on the development and implementation of a BDMA metamodel based on the IBDMA framework.

The framework is implemented for a smart building application scenario with 1000 sensors deployed in the smart building. The evaluation results demonstrate that the IBDMA framework can be used for the autonomous monitoring and control of smart building operations.

We present visualisations for the real-time data generated by the smart building IoT sensors and make predictions about the future values of the data generated from a particular sensor. However, the predictions are made using a simplistic time series model. However, the choice of the model, its efficiency and completeness are beyond the scope of this research.

The framework also does not consider the privacy and governance aspect of data generated by IoT sensors deployed in smart buildings. This can be incorporated as an extension to the IBDMA framework. Another area where the IBDMA framework can be further extended is the consideration of using plug and play IoT devices in smart buildings. The research context and its scope are illustrated in Figure 1.4.

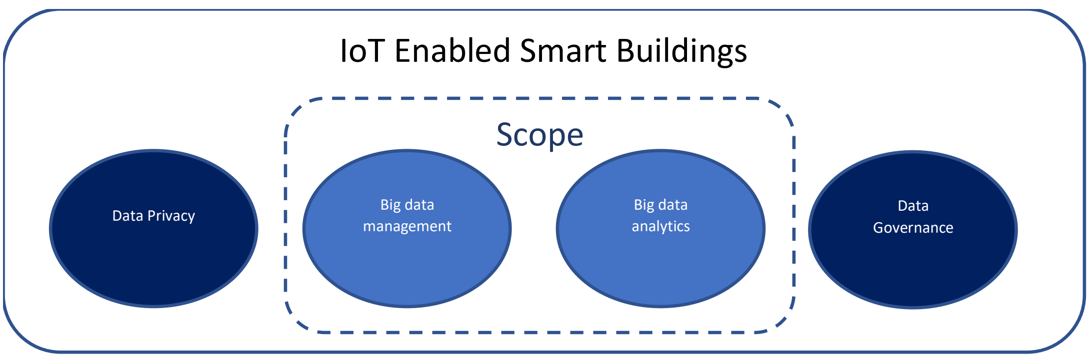


Figure 1.4 Research context and scope

1.4.Research Strategy

The design science research (DSR) method is used in this study (Prat, Comyn-Wattiau & Akoka 2014).The systematic literature review (SLR), as shown in Figure 1.5, is the first step in the DSR approach. The relevant literature is evaluated and reviewed to identify the research gaps in the domain. We conducted the SLR to evaluate 92 of the most significant studies in the literature. The SLR process involves taking data from various studies and filtering, analysing, and synthesising it. There are five stages to the SLR filtration process: 1) inclusion and exclusion criteria for selection; 2) data source selection and search strategies; 3) citation and inclusion decision management; 4) final selection and quality assessment; and 5) data extraction and synthesis from the final collection of studies. The SLR results are utilised as the starting point for the DSR approach.

The research gap aids in determining the research problem. The second stage of the DSR is to become aware of the research problem. The third stage of DSR begins with the definition of the research problem, followed by the proposal of a solution and the definition of the solution's objectives. The IBDMA artefacts were created using the DSR process. Artefact development may entail an examination of current theories and knowledge to create artefacts that are appropriate for their intended use. The fourth step in DSR is to design and develop the IBDMA reference architecture and metamodel for this study.

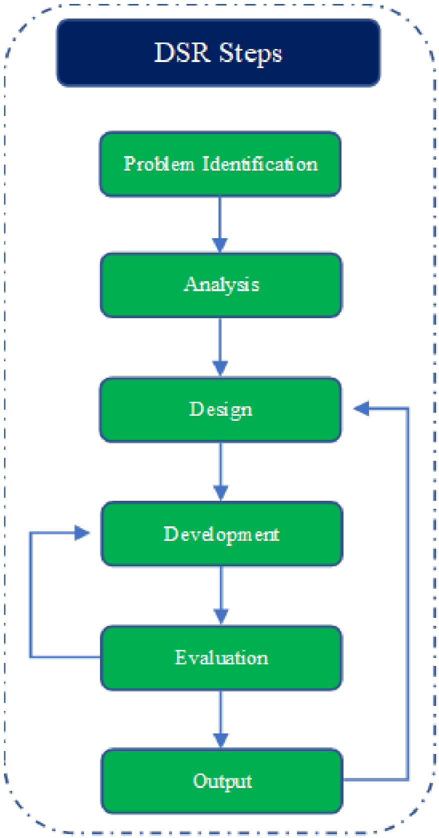


Figure 1.5 Research Strategy

The IBDMA framework is evaluated using real-world use cases and feedback from industry experts against thirteen evaluation criteria extracted from the literature. The framework was improved based on the evaluation results in an iterative way. The fourth step of DSR is to develop the proposed solution. The developed solution is evaluated in the fifth step against the research objectives and evaluation criteria. The findings are documented, discussed, and published to the community in the sixth and final step of DSR.

1.5. The Framework

The IBDMA framework, which serves as the foundation for this study, is briefly discussed in this section. The IBDMA framework was developed to create a unified and cohesive framework for big data management and analysis in IoT-enabled smart buildings. There are five important conceptual elements in the IBDMA framework: people, process, technology, information, and facility.

As shown in Figure 1.6, the framework comprises a reference architecture and a metamodel. The developed framework is being used in a smart building application scenario. It is then evaluated and tested against thirteen evaluation criteria derived from the literature to ensure that it is thorough and effective. The metamodel is also tested in smart building application scenarios. The evaluation results show that both the framework and the metamodel are suitable for the intended purpose.

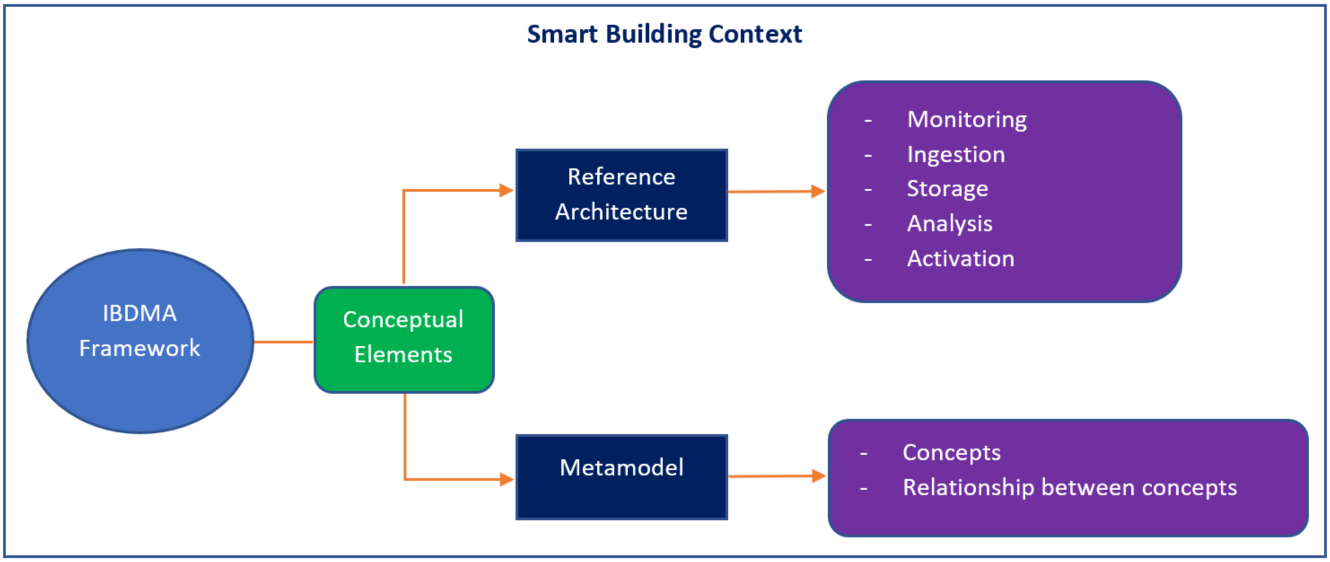


Figure 1.6 IBDMA framework high-level overview

1.6. Applications and Users

In this section, the IBDMA framework's applications and users are discussed. The IBDMA framework is useful for academics and practitioners alike, and it can be used in situations other than smart buildings. The IBDMA framework is designed to be scalable, which means it can be scaled up or down to address data management concerns for smart homes, smart cities, and a range of smart buildings ranging in size from small to large. Users of the IBDMA framework are listed in Table 1.1, along with detailed use cases for each category of user.

The IBDMA reference architecture can be used by big data managers and developers to build and develop data ingestion and data engineering pipelines for ingesting, storing, and analysing huge volumes of data. Because the IBDMA reference architecture is a generic reference architecture that is not tied to a specific use case, they can utilise it in various scenarios. The IBDMA framework can be used by IoT, 5G, and telecommunications specialists and professionals to plan, create, and develop IoT infrastructure deployment in smart buildings. The IBDMA metamodel will assist them throughout the design phase by offering a comprehensive perspective of all the aspects required for the effective design and deployment of IoT infrastructure in smart buildings. The metamodel will also aid them in visualising the relationships between all the major components and how they interact in the smart building ecosystem. This enables them to ensure that the infrastructure they develop and install incorporates all components of IoT infrastructure in smart buildings. The IBDMA reference architecture will aid IoT, 5G, and telecommunications experts in developing data pipelines and analysing IoT data to derive actionable insights. The information will enable inhabitants to have a better experience and manage the different facilities in smart buildings more effectively and efficiently.

The IBDMA framework will assist building developers in determining how a traditional building may be turned into a smart building and how new buildings can be constructed as smart buildings.

Researchers and academics can utilise the IBDMA framework as a foundation for addressing BDMA concerns in smart buildings. They can extend it to incorporate other features such as data privacy, data governance, and plug-and-play IoT devices.

Table 1.2 Applications and users of IBDMA framework

Users	Applications and use cases
Big data managers and developers	Design of data ingestion and data engineering pipelines for big data storage and analysis.
IoT, 5G and telecommunication experts and professionals	Planning, designing and installing big data infrastructure in smart buildings. Deriving meaningful insights from data to improve occupants' comfort and satisfaction. Manage various facilities in smart buildings more effectively.
Building developers	Designing, planning and building new smart buildings. Transforming existing buildings into smart buildings.
Researchers and academics	Addressing BDMA challenges in smart buildings. Extending this framework to incorporate new elements into the framework.

1.7. Thesis Organisation

The remainder of thesis is organised into six chapters. Chapter 2 presents a detailed review of the relevant literature. This review is undertaken to establish the current state of knowledge to specify a specific study area in the relevant context. Chapter 3 presents the research methodology for this research study. Chapter 4 discusses the design and development of the IBDMA framework in detail. Chapter 5 presents the evaluation details and the results obtained. Finally, chapter 6 presents the conclusions with identified opportunities for further research. The rest of the thesis includes bibliography and appendices. The appendices include the research data, publications, ethics approval and consent forms, use case details, and a questionnaire for industry experts. Figure 1.7 provides an overview of the thesis organisation.

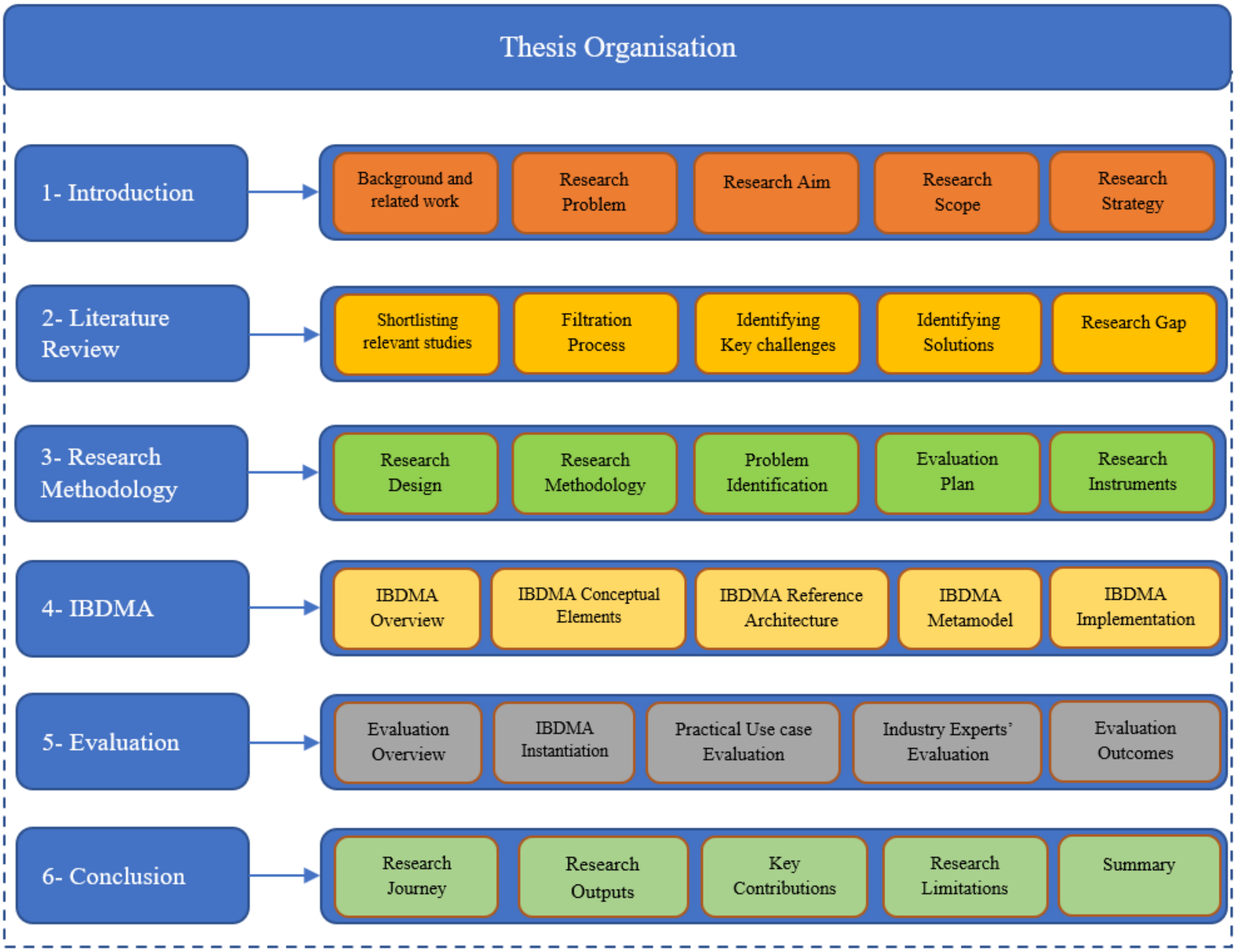


Figure 1.7 Thesis Organisation

1.8.Summary

Chapter 1 provides an introduction to the research. It discussed the terminologies and concepts needed to understand the thesis. The research problem, the aims, objectives, significance, and the scope of the research were presented in this chapter. This chapter briefly discusses the research strategy, and then introduces the framework. Then this chapter discusses the application and users of this research. Finally, the chapter concludes by providing detail on the organisation of the thesis. The next chapter discusses the literature review in detail.

Chapter 2 : Literature Review

The last few years have seen a growing interest in the areas of IoT and smart buildings. However, there are serious challenges associated with this domain, including the effective management and analysis of the data generated by the IoT sensors. Hence, a systematic literature review (SLR) to identify the challenges associated with these areas is needed and to subsequently propose solutions for these challenges.

This chapter details the SLR and is organised in five sections. The first section provides a brief background and introduction to the literature review and the work other researchers are doing in this domain. The second section discusses the research method used to perform the literature review. The third section provides the results of the literature review, highlighting key challenges and the solution to these challenges. The fourth section provides details on identifying the research gap in the literature and how this research addresses the gaps in the literature. The last section concludes the chapter by providing a summary of the findings of the SLR.

2.1. Introduction

The rate at which devices are being connected to the Internet over the last few years is exponential, leading to the enrichment of the digital world (Cisco 2011). In a variety of sectors and domains, IoT is simplifying and improving human life and productivity. Although IoT is also being deployed in industries for industrial applications, this is for specific applications and reasons. The exponential rise observed in the use of IoT devices is primarily due to the advent and usability of IoT devices for the public. Smart watches, smart TVs, smart phones, and many other smart devices are available now for use in homes to transform a home into smart home. Similar to smart homes, there has been a significant increase in interest and demand to make buildings smarter and more efficient by using IoT devices (Rathore et al. 2016). Not only is the demand to build new buildings as smart buildings, the focus is also on transforming existing buildings into smart buildings. There are several reasons for this increase in interest in smart buildings.

First and foremost is that buildings consume a significant percentage of the total energy consumption. Because of this energy consumption, buildings contribute greatly towards climate change and global warming (Cozzi et al. 2020). Energy efficiency is critical to the design and construction of buildings and hence, the primary aim of a smart building is to monitor, regulate and reduce energy consumption without impacting operational efficiency and resident comfort (Jia, Srinivasan & Raheem 2017).

Heating, ventilation and air conditioning (HVAC) systems in a building take a substantial percentage of the total building energy usage (Fan et al. 2016). In addition to the HVAC system, lighting, cooking, controlled loads (swimming pools etc.) and plug loads also contribute to energy consumption. The use of IoT sensors in HVAC and other physical facilities within the buildings makes these facilities more adaptive, efficient and intelligent, and also improves energy efficiency.

Because of the availability of a range of various types of IoT sensors and the number of sensors deployed in buildings, large volumes of data are generated within smart buildings. The massive amount of data generated by the IoT sensors deployed in a smart building can be ingested, extracted, filtered, stored and analysed to obtain meaningful information about the smart building (Bashir & Gill 2016). For instance, the analysis of data from the IoT sensors deployed in smart buildings can be used to improve residents' comfort and the energy efficiency of the smart building (Horban 2016). There are many other applications where big data analytics can be used to improve occupants' comfort, safety, and security as well as building management.

It is evident from a high-level review of the literature that the interest in smart buildings and the management and analysis of big data is growing. However, the use of IoT sensors in smart buildings poses multiple challenges. Hence, the need to identify these challenges and their respective solutions is dire to address the challenges faced in this domain. Our research aims at addressing some of these challenges by developing a framework using a model-driven architecture and design research approach (Duffy & O'Donnel 1998; Dybå & Dingsøyr 2008). In developing such a framework, it is imperative to explicitly identify the prevailing challenges in this domain. As previously mentioned, this research will follow the design research approach. In this chapter, we identify the SLR scope and present the SLR, which is the first step in DSR to identify the research gaps in the domain. The SLR will assist and guide us and other researchers to clearly understand and articulate the underpinning challenges in this vital field of research.

2.2. SLR Scope

The SLR aims at finding the research gaps to address the RQs outlined in Section 1.2 by following the guidelines presented by Kitchenham et al. (2010). The SLR is conducted by selecting most relevant studies from the literature which are labelled from S1, S2, S3… S98 and are presented in Appendix A of the thesis. This process generates rigorous raw data, which are reviewed in the SLR results section. The scope of the SLR is limited to reviewing manuscripts from well-known databases and studies which focus on IoT, smart environments, big data management, big data analytics and metamodels. The following section discusses the research method in detail which was used to conduct the SLR.

2.3. Research Method for the Literature Review

The design, development and evaluation of a BDMA framework for IoT-enabled smart buildings is the primary goal of this research, therefore, the DSR approach is adopted for this research (Prat, Comyn-Wattiau & Akoka 2014). The DSR methodology deals with the design, development and evaluation of a product, a system or an artefact in response to a real or perceived problem. The DSR methodology can be broadly classified into following three main steps: literature review, design and evaluation. As the literature review is the first step in the DSR methodology, the primary focus of this chapter is the SLR (Kitchenham et al. 2010). This is the first step in the systematic identification of the challenges and solutions for the management and analytics of big data generated by IoT-enabled smart buildings. A high-level view of the DSR framework is presented in Figure 2.1. The SLR approach provides an organised and methodical framework to identify, select and synthesise the literature relevant to the research question in hand (Kitchenham et al. 2010). This study complements the original SLR approach by maintaining the quality and level of the literature review by following the requirements for citation and assessment procedures (Kitchenham et al. 2010). Five stages are utilised in performing the SLR to identify the gaps in the research relevant to the research questions. These stages are: (1) inclusion and exclusion criteria for selection; (2) selection and search strategies for data sources; (3) citation and inclusion decision management; (4) final selection and quality assessment; and (5) data extraction and synthesis.

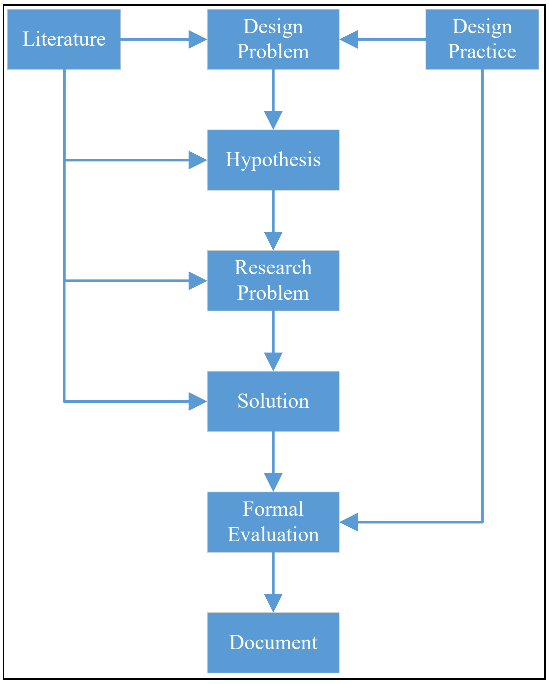


Figure 2.1 Design Science Research (Dybå & Dingsøyr 2008)

2.3.1. Inclusion and exclusion criteria for selection

This research aims to address the two research questions identified in Section 1.2. It is imperative to identify and understand the foundational concepts which are important in addressing the research questions. Table 2.1 provides a list of all the foundational concepts pertaining to each research question. On reviewing the literature, we found that smart building challenges and concepts in the big data analytics and management scope were overlapping with other smart environments. Primarily, the difference was the scalability and the use-cases between smart buildings and other smart environments, but the high-level big data management and analytics concepts and challenges were overlapping. Hence, we included other smart environments such as smart grids and smart homes to expand our knowledge base for the literature review.

The literature review is conducted to review the research articles whose focus is on big data analytics for IoT-enabled smart buildings with the aim of finding the research gaps in this area. Based on the research questions (RQ1 and RQ2) and the concepts highlighted in Table 2.1, we defined four search categories and keywords corresponding to each of the search categories as presented in Table 2.2. We followed a four-stage filtration process to come up with the final shortlist of manuscripts. This four-stage process and the assessment criteria are presented in Table 2.3. We also defined quality criteria for the selection of each research article, as presented in Table 2.4. Only research articles which focussed on the management and analysis of big data generated by IoT-enabled smart buildings and articles which focussed on developing a metamodel for IoT big data were included in the literature review. Only articles which passed the minimum selection quality criteria outlined in Table 2.4 were shortlisted. Articles which did not discuss IoT or big data management or big data analytics or smart environments or metamodeling were excluded from the selection process of the SLR. Research articles produced during January 2013 to July 2021 were included in the selection process. Articles written in a language other than English were also excluded from the shortlisting processing. Studies that focussed on smart environments other than smart buildings such as smart homes, smart cities, smart grids were also included in the selection as the basic concepts and challenges faced in the data management and analysis in these smart environments are similar in nature to smart building environments.

2.3.2. Data source selection and search strategies

The top four most distinguished online databases which were used for shortlisting of the studies for the literature review are listed in alphabetical order as follows.

- ACM Digital Library (www.portal.acm.org/dl.cfm)
- Elsevier ScienceDirect (www.sceincedirect.com/)
- IEEE Xplore (www.ieeexplore.ieee.org/Xplore/)
- SpringerLink (www.springerlink.com/)

A few webpages were also considered for the literature review on top of the aforementioned academic research databases. The research articles selected from the aforementioned electronic databases and the selected webpages provided adequate exposure to cover all aspects of the research topic. A list of these studies is presented in Appendix A.

Table 2.1 Foundational Concepts

Research Question	Concepts
RQ1: How to effectively manage and analyse IoT big data in a smart building context?	IoT, smart buildings, big data management and analytics.
RQ2: How to holistically identify all the elements and the relationship between these elements to effectively manage and analyse big data in IoT-enabled smart buildings?	Metamodel, IoT, smart buildings, big data management and analytics.

The selection of the research articles for the purpose of performing the SLR were selected from the qualitative and quantitative studies conducted both in industry as well as in academia. Figure 2.2 shows four stages for the selection and shortlisting of research articles for the SLR and the number of research articles filtered during each stage of the selection process. Table 2.2 shows the four different search categories we defined based on the research questions and the foundational concepts as illustrated in Table 2.1. This helped in identifying the search terms against each of the search categories. As the part of the first stage of the SLR for the selection of articles, each term from each search category in Table 2.2 was concatenated with each term from the other search categories. As an example, the search term from search category 1 was combined with each search term from search categories 2, 3 and 4. Similarly, each search term from search category 2 was combined with each search term from search categories 3 and 4. In the same way, each search term from search category 3 was combined with each search term from search category 4. The first step in the selection process included ORing the search terms from different search categories. During the first selection stage, the research papers that included any of the two keywords were shortlisted (keyword from search category 1 OR keyword from search category 2, keyword from search category 1 OR keyword from search category 3, keyword from search category 1 AND keyword from search category 4 and so on). The shortlisted research papers were consolidated using Microsoft Excel to keep a database of all selected articles. The articles comprising or addressing reviews, comments, correspondences, reader's letters, discussions, workshops, poster sessions and panels were excluded from the selected list of articles. Because there were a number of keywords combination, the outcome of stage 1 for the selection of the research articles produced a total of 23027 research articles, as shown in Figure 2.2.

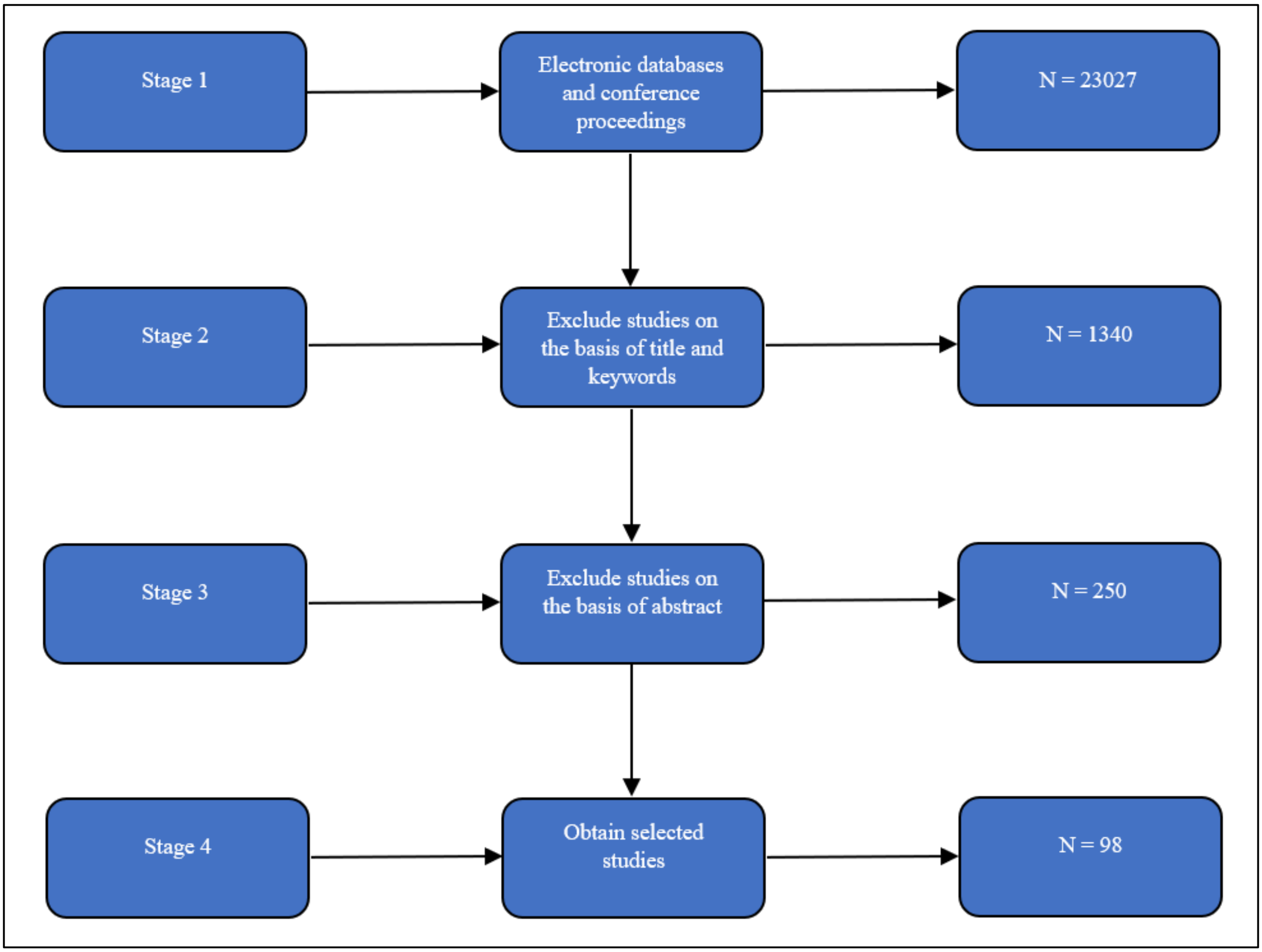


Figure 2.2 Study selection process (Alzoubi, Gill & Al-Ani 2016)

Table 2.2 Search items

Search Category	Keywords
IoT	Internet of Things, Sensor Driven
Smart Buildings	Smart Buildings, HVAC, Smart Homes, Smart Grids, Smart Cities, Energy Efficiency, HVAC, Occupancy Monitoring.
Big Data Management and Analytics	Data Management, Data Analytics, Big Data
Metamodel	Metamodel, Ontology.

Table 2.3 Assessment method (Alzoubi, Gill & Al-Ani 2016)

Filtration stage	Method	Assessment criteria
First stage filtration	Search online databases for the keywords	Keywords

Second stage filtration	Search the terms in the title and exclude studies if search terms are not found in the title	Title = search term(s) Yes = accepted No = rejected
Third stage filtration	Read the abstract and exclude articles if abstract is not relevant	Abstract = relevant Yes = accepted No = rejected
Fourth and final stage filtration	Gather shortlisted papers and critically evaluate studies	Discusses data management or data analytics for smart environments Yes = accepted No = rejected

2.3.3. Citation and inclusion decision management

For the purpose of the SLR, the citation procedure presented by Dyba˚ and Dingsøyr (Dybå & Dingsøyr 2008) was adopted. Following this citation procedure, in the second stage, all 23027 paper titles that were archived in a spreadsheet from four academic research databases, as mentioned in Stage 1, were examined by searching both the search terms in the title and the keywords of the academic manuscript. Generally speaking, the AND operation was performed while conducting this filtration process. If a search term in search category 1 and a search term in search category 2 were both found in the title AND the keywords of the academic manuscript, the manuscript was shortlisted otherwise it was not shortlisted for the next filtration stage. The number of articles at the end of this filtration stage reduced to 1340.

2.3.4. Final selection and quality assessment

The relevance of the selected manuscripts was determined by reviewing the abstracts obtained during the previous filtration stage. If after reviewing the abstract of the selected article, we found the abstract to be relevant to our field of study, the article was shortlisted for the next stage. If, however, the abstract was found to be irrelevant to our focus area, the article was not chosen for the next screening stage. The number of selected manuscripts was reduced to 250 at the completion of this filtration process. The assessment criteria used for the selection process to shortlist the research manuscripts across different filtration stages is presented in Table 2.3.

2.3.5. Data extraction and synthesis

During the final extraction and synthesis stage, the filtered and shortlisted papers were reviewed and studied in detail. During the review process the quality of each of the selected papers (out of 250) was assessed based on the assessment criteria proposed by Dyba˚ and Dingsøyr (Dybå & Dingsøyr 2008) as presented in Table 2.4. The seven quality assessment criteria proposed by (Dybå & Dingsøyr 2008) were used to ensure that the selected articles satisfied the quality, credibility and relevance to the research topic.

Table 2.4 Quality criteria (Dybå & Dingsøyr 2008)

Quality criteria	
1	Is the paper based on research?
2	Are the aims of the research paper explained clearly?
3	Is the context described accurately in the paper?
4	Was the method chosen for research suitable for addressing the research aims?
5	Was the data analysis done comprehensively?
6	Is the conclusion presented clearly?
7	Is the study of value for research or practice?

When the quality criteria highlighted in Table 2.4 was applied to 250 articles from the previous filtration stage, the number of shortlisted articles was reduced to 98 (out of which 25 were specifically on smart buildings), excluding 152 articles as part of this filtration stage.

The number of manuscripts selected from the aforementioned research databases at each filtration stage is presented in Table 2.5. The table also lists the percentage of selected articles chosen from each database. Most of the selected papers (40.8%) are from IEEE, followed by Elsevier (28.5%) and SpringerLink (20.4%). The least number of papers (10.2%) were shortlisted from ACM.

Table 2.5 Summary of Selection Process

Databases	**1st filtration**	**2nd filtration**	**3rd filtration**	**Studies selected**	**Percent selected (%)**
IEEE	8928	897	125	40	40.8
ACM digital library	2201	154	32	10	10.2
SpringerLink	5129	86	44	20	20.4
Elsevier	6769	203	49	28	28.5
Total	23027	1340	250	98	100

2.4. Literature Review

The final selection of 98 papers (S1-S98) as presented in the previous section is provided in Appendix A. These manuscripts were reviewed in detail, and it was found that all the shortlisted articles encompassed the research questions (**RQ1 and RQ2**: challenges AND solutions) comprehensively and focussed on metamodeling and reference architecture for BDMA for IoT-enabled smart environments.

The selected 98 papers were evaluated against each of the selection quality criteria detailed in Table 2.4 and the results of the evaluation are presented in Table 2.6. Each selected paper was assigned a score if it satisfied a quality criterion. It can be seen from the table that all the papers scored either a 6 or a 7 for the quality assessment.

Table 2.6 Study Assessment Criteria

Study	1 Research	2 Aim	3 Context	4 R design	5 Data analysis	6 Findings	7 Value	Total
S1	1	1	1	1	0	1	1	6
S2	1	1	1	1	0	1	1	6
S3	1	1	1	0	1	1	1	6
S4	1	1	1	1	1	1	1	7
S5	1	1	1	1	1	1	1	7
S6	1	1	1	0	1	1	1	6
S7	1	1	1	1	0	1	1	6
S8	1	1	1	1	1	1	1	7
S9	1	1	1	0	1	1	1	6
S10	1	1	1	1	1	1	1	7
S11	1	1	1	0	1	1	1	6
S12	1	1	1	0	1	1	1	6
S13	1	1	1	0	1	1	1	6
S14	1	1	1	1	1	1	1	7
S15	1	1	1	1	1	1	1	7
S16	1	1	1	1	1	1	1	7
S17	1	1	1	1	1	1	1	7
S18	1	1	1	0	1	1	1	6
S19	1	1	1	1	1	1	1	7
S20	1	1	1	1	1	1	1	7
S21	1	1	1	1	1	1	1	7
S22	1	1	1	0	1	1	1	6
S23	1	1	1	1	1	1	1	7
S24	1	1	1	0	1	1	1	6
S25	1	1	1	1	0	1	1	6
S26	1	1	1	0	1	1	1	6
S27	1	1	1	1	0	1	1	6
S28	1	1	1	1	1	1	1	7
S29	1	1	1	0	1	1	1	6
S30	1	1	1	1	0	1	1	6
S31	1	1	1	0	1	1	1	6
S32	1	1	1	0	1	1	1	6
S33	1	1	1	1	0	1	1	6

S34	1	1	1	1	1	1	1	7
S35	1	1	1	1	0	1	1	6
S36	1	1	1	0	1	1	1	6
S37	1	1	1	1	0	1	1	6
S38	1	1	1	1	0	1	1	6
S39	1	1	1	0	1	1	1	6
S40	1	1	1	1	1	1	1	7
S41	1	1	1	1	0	1	1	6
S42	1	1	1	0	1	1	1	6
S43	1	1	1	1	0	1	1	6
S44	1	1	1	1	0	1	1	6
S45	1	1	1	0	1	1	1	6
S46	1	1	1	1	1	1	1	7
S47	1	1	1	1	0	1	1	6
S48	1	1	1	1	0	1	1	6
S49	1	1	1	0	1	1	1	6
S50	1	1	1	1	0	1	1	6
S51	1	1	1	0	1	1	1	6
S52	1	1	1	1	0	1	1	6
S53	1	1	1	1	0	1	1	6
S54	1	1	1	0	1	1	1	6
S55	1	1	1	1	0	1	1	6
S56	1	1	1	1	0	1	1	6
S57	1	1	1	0	1	1	1	6
S58	1	1	1	1	0	1	1	6
S59	1	1	1	1	0	1	1	6
S60	1	1	1	1	1	1	1	7
S61	1	1	1	1	0	1	1	6
S62	1	1	1	0	1	1	1	6
S63	1	1	1	1	0	1	1	6
S64	1	1	1	1	0	1	1	6
S65	1	1	1	0	1	1	1	6
S66	1	1	1	1	0	1	1	6
S67	1	1	1	1	0	1	1	6
S68	1	1	1	1	1	1	1	7
S69	1	1	1	1	0	1	1	6
S70	1	1	1	1	1	1	1	7
S71	1	1	1	1	0	1	1	6
S72	1	1	1	1	0	1	1	6
S73	1	1	1	0	1	1	1	6
S74	1	1	1	1	0	1	1	6
S75	1	1	1	1	1	1	1	7
S76	1	1	1	1	0	1	1	6
S77	1	1	1	1	0	1	1	6
S78	1	1	1	0	1	1	1	6
S79	1	1	1	1	0	1	1	6

S80	1	1	1	1	1	1	1	7
S81	1	1	0	1	0	1	1	6
S82	1	1	1	1	0	1	1	6
S83	1	1	1	1	1	1	1	7
S84	1	1	1	1	0	1	1	6
S85	1	1	1	0	1	1	1	6
S86	1	1	1	0	1	1	1	6
S87	1	1	1	1	1	1	1	7
S88	1	1	1	0	1	1	1	6
S89	1	1	1	1	0	1	1	6
S90	1	1	1	1	1	1	1	7
S91	1	1	1	0	1	1	1	6
S92	1	1	1	0	1	1	1	6
S93	1	1	1	1	1	1	1	7
S94	1	1	1	1	1	1	1	7
S95	1	1	1	1	1	1	1	7
S96	1	1	1	1	1	1	1	7
S97	1	1	1	1	1	1	1	7
S98	1	1	1	1	1	1	1	7
Total	98	98	97	68	60	98	98	

The BDMA challenges and the solution to these challenges in the context of IoT-enabled smart buildings were identified after reviewing the shortlisted 98 papers. The aim of reviewing and analysing these selected manuscripts was to come up with the body of knowledge to identify the research gap pertaining to BDMA for IoT-enabled smart buildings. The research gap provides a comprehensive list of challenges faced by both industry and academia related to data management and analytics.
On analysing the selected studies for challenges and solutions in the field of research, several common themes or categories emerged. We list all the major challenges and solutions identified as an outcome of the SLR process in the subsequent sections.

2.4.1. Challenges

A comprehensive review of the selected 98 studies revealed five major categories of challenges in the BDMA domains for smart buildings and other smart environments, like smart cities and smart grids. The five categories identified are then sub-divided into seventeen subcategories of challenges for the BDMA in smart environments. These challenges help us identify the research gap and formally formulate the research questions. A consolidated list of BDMA challenges for smart ecosystems enabled by IoT is provided in Table 2.7.

2.4.1.1. Big data management challenges

The challenges associated with big data management that were identified from the final selection of 98 papers are presented in this section. The identified challenges pertaining to big data management that are listed below are applicable to various smart environments including smart homes, smart buildings, smart cities, smart grids etc.

- **Collection and governance**

One key challenge in the management of big data is the governance and collection of data itself. Most organisations do not follow standardised procedures to collect and govern their data assets which ultimately results in the creation of serious issues at the data reporting level. Data is growing faster than ever before, and the collection and secure storage of structured, unstructured, and semi-structured data bring new challenges to organisations. Study S68 addresses this key challenge. Studies S1, S4, S51 and S52 propose that future work needs to be done to address this particular challenge.

- **Integration and sharing**

Another critical challenge faced by various organisations in the area of big data management is the integration and sharing of big data. Traditional databases and data warehouses need to be integrated with modern data lakes and delta lakes to obtain meaningful insights. This, in most cases, requires data to be migrated as well as transformed for it be useful. In addition to integrating different data storage facilities and databases, the sharing of big data among both external and internal users is also a challenge since it requires not only supporting infrastructure but also the security of data at rest as well as in transit which needs to be looked into carefully. While these challenges have been identified as a proposed future area of research in studies S1 and S51, this challenge has been identified and addressed in studies S4, S26, S73, S90 and S91.

- **Consolidating heterogenous big data**

Data in most organisations is stored in various source systems and it can be varied in nature. To obtain insights from this heterogeneous data, it needs to be consolidated. Hence, the consolidation of data is critical to the performance of business. If the data in various formats cannot be consolidated effectively and efficiently, it can potentially cause negative impacts on the business. Fortunately, in today's world, we have the tools available to accomplish this. Studies S4, S55 and S67 identify and address this challenge. It has also been discussed in study S2 as an area of future research work.

- **Security and privacy**

The security and privacy of big data is arguably the most critical challenge in this domain. There are different tools and techniques available now to collect data in various formats from a multitude of sources. However, the privacy and security aspect of data poses new and unexplored challenges. Security and privacy breaches can have a huge impact on businesses and customers and can make millions of users vulnerable to fraud and identity thefts. This important challenge is raised and addressed in studies S5, S25, S59, S76, S77, S79 and S82. It has also been listed as a future focus area in studies S1, S4, S51, S52, S54 and S56.

- **Asset and power generation side management**

Asset and power generation data management is one of the most difficult aspects of smart grid big data management. Studies S1 and S3 address this important and relevant challenge. Some of the data sources for this type of data include greenhouse gas emission data, technology cost and performance data, and power generation cost data. S1 begins by outlining the attributes and sources of big data in the energy sector. The author then presents a smart energy management process paradigm based on big data. Lastly, the paper outlines a comprehensive assessment of big data analytics for smart energy management, using a smart grid as a research foundation.

- **Demand side management (DSM)**

Studies S1 and S2 identified demand-side big data management as a challenge. Some of the most typical sources of this sort of data include historical load data, weather data, electricity consumption data, and consumer behaviour data. Demand-side big data management is also mentioned as an area of future research work in study S3.

2.4.1.2. Big data analytics challenges

Big data analytics challenges for smart environments are discussed in this section. We also include the smart city data analytics challenges as the related challenges which also apply to smart buildings, which is the focus of our research.

- **Real -time, iterative and stream data analytics**

Real-time, iterative, and stream processing of data generated by IoT sensors is a key challenge for the analysis of big data. Batch processing can be used in some applications, but in others, such as catastrophe management, the real-time or stream processing of massive data is required. In studies S3, S12, S24, S27, S29, S34, S38, S39, S44, S56, S58, S62, S64, S70, S71, S80 and S81 this was identified and addressed a relevant challenge. Study S2 identifies this challenge as a future area of research.

- **Machine learning and data mining for healthcare and people comfort**

To obtain meaningful insights from big data and to create business value, machine learning and deep learning has been identified as a challenge. Machine learning and deep learning can also help in healthcare and improve people's lives. This has been identified and addressed as a challenge in studies S23, S40, S41, S43, S47, S49, S50, S57, S63, S66, S69, S72, S74, S75, S78, S82 and S86.

- **Energy efficiency in miscellaneous systems and environments**

Another challenge related to big data analytics is to effectively improve the energy efficiency of various systems and environments including smart homes and smart cities. Two studies S30 and S36 identify and address this challenge.

- **Risk management and fraud detection**

Risk management in various smart environments and fraud detection in the financial sector has been identified and addressed as a key challenge in studies S32 and S53. Big data can help manage and mitigate risk in various industrial sectors. The analysis of big data can also help banks to identify fraud and to mitigate the risk of unauthorised transaction.

- **Improving reliability, security, efficiency, and accuracy of systems**

Big data analytics can be employed to improve the reliability, security, efficiency and accuracy of various systems. This has been addressed as a key challenge in six studies (S35, S37, S42, S45, S46 and S48). Since IoT sensors continuously monitor various environments in which they are deployed, the data analytics for IoT sensor data can improve the reliability of systems.

- **Energy efficiency**

Improving the energy efficiency of smart buildings is one of the primary difficulties linked with IoT-enabled smart buildings. Because energy is both limited and expensive, designing energy efficient buildings is a major challenge in big data analytics. In studies S13, S14, S15, S16, S17, S18, S19, S60 and S87, this was identified and addressed as a challenge. In studies S5 and S21, this problem was also indicated as a future research topic.

- **User centric services**

Another problem linked with smart building big data analytics is providing recent-focussed services to make the building occupants feel safer and more secure. The aim of user-centric services is to ensure buildings are more energy efficient and provide users with a comfortable environment. In studies S21, S22, S83, S84, S87, S88, S92, S93, S94, S95 and S98, this was identified and addressed as a challenge. In studies S5 and S13, this was also mentioned as a potential future research topic.

- **Occupancy monitoring**

Another important problem when performing real-time data analytics in smart buildings is keeping track of the building's occupancy. Because occupancy monitoring cannot be evaluated in isolation, we must look at both the energy statistics and the user profile. In smart building operations, this poses a possible difficulty. In studies S6, S13, S20, and S21, this was recognised and addressed as a challenge.

- **Predicting building operations**

Predicting smart building operations through data analytics by analysing the IoT sensor data is another challenge that has been identified and addressed in study 95. Predicting building operations aids in lowering energy costs and identifying any severe problems that could result in the building's routine operations failing. This particular topic has been marked as a future research area in study S13.

- **Planning and improving smart ecosystems**

Planning and improving smart cities by analysing data from IoT sensors installed in smart cities is one of the primary challenges related with big data analytics in smart cities. Weather monitoring, traffic monitoring, pollution monitoring and parking area monitoring are examples of such challenges. In studies S7, S8, S9, S10, S11, S28, S31, S33, S54, S61, S65, S85, S89, S96 and S97, this has been identified and addressed as a challenge.

- **Plug and play IoT devices**

Using plug-and-play IoT devices in smart environments is one of the issues that has been identified as a future research area in study S10. These devices can be replaced with other IoT devices, allowing for the collection of different parameters from many places with fewer sensors.

Table 2.7 summarises the BDMA challenges in the context of smart ecosystems that have been discussed above.

Table 2.7 Big Data Management and Challenges

Challenges	**Challenge ID**	**Studied that identified the challenge**	**Identified as a Future work/research gap**	**Frequency**
IoT Big Data Management	C1			
1. Collection and governance.	C1.1	S68	S1, S4, S51, S52	1
2. Integration and sharing.	C1.2	S4, S26, S73, S90-S91	S1, S51	5
3. Bringing heterogeneous big data together.	C1.3	S4, S55, S67	S2	3
4. Security and privacy	C1.4	S5, S25, S59, S76-S77, S79, S82	S1, S4, S51, S52, S54, S56	7
Smart Grid/Meter Big Data Management	C2			
1- Asset and Power Generation Side Management	C2.1	S1, S3		2
2- Demand side management (DSM)	C2.2	S1-S2	S3	2
Big Data Analytics	C3			
1. Real Time, iterative and stream Analytics	C3.1	S3, S12, S24, S27, S29, S34, S38-S39, S44, S56, S58, S62, S64, S70-S71, S80-S81	S2	17
2. Machine learning and data mining for healthcare and people comfort	C3.2	S23, S40, S41, S43, S47, S49, S50, S57, S63, S66, S69, S72, S74-S75, S78, S82, S86		17
3. Energy efficiency in miscellaneous systems and environments	C3.3	S30, S36		2
4. Risk management and fraud detection	C3.4	S32, S53		2
5. Improving reliability, security, efficiency, and accuracy of systems	C3.5	S35, S37, S42, S45-S46, S48		6
Big Data Analytics- Smart Buildings	C4			
1. Energy Efficiency	C4.1	S13-S19, S60, S87	S5, S21	9
2. User Centric Services	C4.2	S21-S22, S83-S84, S87-S88, S92-S95, S98	S5, S13	11
3. Occupancy monitoring	C4.3	S6, S13, S20, S21		4
4. Predicting Building operations	C4.4	S95	S13	1
Big Data Analytics-Smart Cities	C5			
1. Planning and improving smart cities	C5.1	S7-S11, S28, S31, S33, S54, S61, S65, S85, S89, S96-S97		15
2. Plug and Play IoT devices	C5.2		S10	0

On reviewing the literature and identifying the challenges that have been raised, it is evident that a coherent framework which combines a reference architecture and a metamodel for the BDMA of IoT-enabled smart buildings is missing in the literature.

2.4.2. Solutions

The answers to the issues raised in the preceding section are identified in this section. Table 2.8 lists all of the solutions to the challenges outlined in the previous section that have been found in the literature.

2.4.2.1. Big data management solutions

The challenges in the BDMA of IoT-enabled smart buildings helped us identify the research gaps and formulate the research questions for the thesis. But we also gathered information on what the proposed solutions are to those challenges that are discussed in the literature. The solutions to the BDMA challenges outlined in studies S1-S98 are presented in this section. Based on these proposed solutions, we define our research aims based on the research questions as identified in Sections 1.2 and 1.3.

- **IT infrastructure**

Having a robust IT architecture that accounts for big data acquisition from numerous data sources is one of the solutions to big data management difficulties. Data in multiple formats should be able to be integrated by the infrastructure. Having the optimal infrastructure is key for the successful implementation of a big data management platform. Another crucial feature of any IT architecture is to enable effective data protection and privacy. In studies S1-S2, S4-S5, S20, S33-S36, S38-S43, S45, S55-S57, S60-S62, S64, S67-S68, S72-S73, S77, S79-S80, S86-S88 and S96-S97, this solution was proposed.

- **Big data storage solutions**

For the management of big data, the availability of big data storage solutions has been proposed as a solution to big data management challenges. In recent years, a number of big data storage solutions have been proposed and have since been used to tackle the management of big data. These tools include but are not limited to HDFS, cloud storage, NoSQL databases etc. The usage of these tools to solve big data management issues has been proposed in studies S28, S30, S31, S37-S40, S44, S51-S53, S59, S63, S65, S67, S69, S70, S76, S79, S85, S89, S92, S94 and S98 as a solution.

- **Metamodeling and ontology development**

Due to the complexity of big data and the complexity of managing big data, metamodeling and ontology development is presented as a solution in studies S32, S91, S93. Metamodeling and ontology technique help provide a holistic view of all the components involved in a big data ecosystem. This approach also helps in identifying how various elements of the system interact with each other and the linkage between these components.

- **Software defined networking**

In studies S24, S31 and S95, software defined networking (SDN) has been identified and presented as a solution to address big data management challenges. SDN can aid in the effective management of the network to improve the performance of big data applications.

2.4.2.2. Big data analytics solutions

The solutions identified in studies S1-S98 for big data analytics challenges are presented in this section.

- **Apache Spark**

In studies S2-S3, S12, S27, S29-S30, S85, S90 and S94 Apache Spark (*Apache Spark*) has been presented as a solution for the real-time, iterative, and stream processing of huge data. It is a free and open-source cluster computing platform that is used for real-time data processing and analysis. Spark divides a big workload into multiple workloads for efficient computing and processing. Spark has gained immense popularity in recent years and programming its inherent big data capabilities can be utilised in Scala, Python and R languages.

- **Analysing smart meter data**

Studies S13-S19 and S29 propose the construction of a modelling-based strategy to improve the energy efficiency of smart buildings. The research proposes analysing data from smart metres to make informed judgments about how to enhance energy efficiency.

- **Analysing environmental conditions**

The technique for increasing the comfort of the occupants of the smart building is to analyse the environmental conditions using diverse types of IoT sensors to reach a meaningful conclusion based on the data collected from these sensors. Using this data, we can also anticipate how smart buildings will operate. S5, S13, S21, S22 S28, S80, S84 and S92 research studies have all identified and proposed this solution.

- **Analysing occupancy monitoring data**

To determine the occupancy of smart building residents, the use of wearable technology, sensor-based monitoring and camera-based monitoring are the available options. Occupancy monitoring is vital since it aids in improving energy efficiency and also helps in enhancing user comfort. Studies S6, S13, S21, S31 and S58 present this as a solution to the challenges addressed in Table 2.7.

- **Real time analytics**

Data from numerous IoT sensors should be analysed in real time for smart city planning and improvement. Because smart cities feature a plethora of sensors, such as proximity sensors, motion sensors, sound sensors, moisture sensors, and magnetic sensors, analysing data in real time is the best way to improve living conditions in a smart building. In studies S7-S11, S54, S71, S86, S92 and S98, this has been presented as a solution.

- **Develop plug and play IoT devices**

The lack of plug-and-play IoT devices, which would allow the number of devices to be reduced while new devices could be added without altering the underlying architecture, was one of the concerns mentioned in the preceding section. The solution is to create new IoT gadgets that are ready to use right away. In research S10, this was recognised as a possible solution.

- **Machine learning and data mining**

Machine learning has gained importance in recent years and has found many applications in the analysis big data. In studies S1, S3, S6, S13, S17, S23 and S25-S26 machine learning and data mining techniques have been marked as solutions to multiple challenges associated with the analysis of big data in various sectors from healthcare to transportation and many more.

- **Data visualisation tools**

The advent and arrival of sophisticated and more mature data visualisation tools has also helped the analysis of big data across industry and academia. Studies S33, S37-S40, S60, S63, S68, S70, S74, S75, S81, S85-S86 and S92 present this as a solution to address big data analysis challenges in various sectors.

- **Fog and edge computing**

For analysing big workloads, studies S29, S47, S60 and S88 mark fog and edge computing as a solution to analyse data closer to the edge. Fog and edge computing results in improved data processing efficiency since the analysis is done closer to the edge. Only the results or the critical information is sent to the centralised location, which helps in improved bandwidth across the network infrastructure.

Table 2.8 summarises BDMA solutions in the context of smart ecosystems against the challenges as outlined in Table 2.7.

Table 2.8 Big Data Management and Analytics Solutions

Solutions	Challenges addressed	Studies	Frequency of challenges addressed
Big Data Management Solutions to the challenges identified in Table 2.7			
1- IT infrastructure (real-time ingestion pipelines, cloud)	C1.1, C1.2, C1.3, C1.4, C2.1, C2.2, C3.1, C3.2, C3.3, C3.5, C4.1, C4.2, C5.1	S1-S2, S4-S5, S20, S33-S36, S38-S43, S45, S55-S57, S60-S62, S64, S67-S68, S72-S73, S77, S79-S80, S86-S88, S96-S97	13
2- Big data storage solutions (cloud, HDFS, Elasticsearch etc)	C1.1, C1.3, C1.4, C3.1, C3.2, C3.3, C3.4, C3.5, C4.2, C5.1	S28, S30-S31, S37-S40, S44, S51-S53, S59, S63, S65, S67, S69-S70, S76, S79, S85, S89, S92, S94, S98	10
3- Metamodeling and ontology development	C1.2, C3.4	S32, S91, S93	2
4- Software defined networking (SDN)	C3.1, C4.2	S24, S31, S95	2
Big Data Analytics Solutions to the challenges identified in Table 2.7			
1- Apache Spark (big data analysis tools)	C1.2, C3.1, C3.3, C4.2, C5.1	S2-S3, S12, S27, S29-S30, S85, S90, S94	5
2- Analysing smart meter data	C4.1, C4.3, C4.4	S13-S19, S29	3
3- Analysing environmental conditions	C3.1, C4.2, C4.4, C5.1	S5, S13, S21-S22, S28, S80, S84, S92	4
4- Analysing data from occupancy monitoring sensors	C4.3	S6, S13, S21, S31, S58	1
5- Real time analytics of IoT sensor data	C3.5, C4.2, C5.1	S7-S11, S54, S71, S86, S92, S98	3
6- Develop plug and play IoT devices	C3.1, C5.2	S10	2
7- Machine learning and data mining	C1.2, C1.4, C2.1, C2.2, C3.1, C3.2, C3.5, C4.4	S1, S3, S6, S13, S17, S23, S25-S26	8
8- Data visualization tools (Power BI, Tableau, Kibana)	C1.1, C3.1, C3.2, C3.5, C4.1, C4.2, C5.1	S34, S46, S48-S50, S66, S72, S74-S75, S78, S81-S83, S86, S90 S33, S37-S40, S60, S63, S68, S70, S74, S75, S81, S85-S86, S92	7
9- Fog and edge computing	C3.1, C3.2, C4.1, C4.2	S29, S47, S60, S88	4

2.5.Research Gap

The review of the literature in section 2.4 identified gaps related to the BDMA in the context of IoT-enabled smart buildings. It is evident from the literature review that a coherent framework which incorporates both the reference architecture and metamodel to identify and address BDMA challenges by clearly identifying all the elements in the smart building and the relationships between these elements for the autonomous control of IoT-enabled smart buildings is missing in the literature. There are reference architectures available, but they don't provide a holistic view of all the elements involved in the smart building ecosystem since an associated comprehensive metamodel is missing from the design, which means researchers, developers and practitioners can miss potential critical elements or the relationships between these elements while designing the BDMA solution for smart buildings. Hence, this research focuses on addressing this gap by developing the IBDMA framework. The research gaps in this thesis are derived from the research questions (Section 1.2) and research aims (Section 1.3.1), and the literature review (section 2.4). The proposed IBDMA framework aims to address the research gaps by providing a comprehensive reference architecture and metamodel that enables the big data management and analysis challenges in IoT-enabled smart buildings to be addressed. The research gaps are presented in Table 2.9 and linked to the corresponding challenges identified as a result of SLR (section 2.4).

Table 2.9 Research challenges, gaps and questions

Challenges (Thesis scope)	Research gaps	Research questions
IoT sensors deployed in smart buildings generate huge amount of data which is challenging to manage.	Management of big data requires infrastructure and tools which are different to traditional ways of storing and managing data with a holistic view of all elements in the smart building ecosystem.	RQ1 aims to address the challenge of effectively managing big data generated by IoT sensors deployed in smart buildings. RQ1 aims to also provide details on the infrastructure elements required to enable effective big data management.
Huge amount of data produced by IoT sensors deployed in smart buildings poses analytics challenges.	The analysis of big data in real-time requires stream processing and real-time data analytics ensuring all components are taken care of.	RQ1 will address the challenge of effectively analysing big data generated by IoT sensors deployed in smart buildings. RQ1 will also provide details about the tools and infrastructure required to enable effective big data analysis.
Autonomous control of smart buildings.	A BDMA framework which enables autonomous control of various facilities in smart buildings.	RQ1 will address challenge of autonomously actuating various controls of smart buildings facilities.

Identification of all the elements and the relationship between these concepts that are required in smart building ecosystem for the effective management and analysis of big data is a challenge.	A metamodel to provide a holistic view of all the elements required in the big data ecosystem.	RQ2 will address the challenge of identifying and defying all the elements and the relationship between these elements by developing a metamodel.
A scalable and robust framework that could be easily scaled up for all types of smart buildings and for environments other than smart buildings.	Hard to scale up and down the big data systems for varying environments.	RQ1 and RQ2 must address the challenge of developing a scalable framework to address BDMA challenges.

2.6.Summary

Based on the guidelines provided by Kitchenham et al. (2010), this SLR seeks to give a complete evaluation of the challenges and solutions associated with IoT-enabled smart building data management and data analytics. This is accomplished through a thorough examination of 98 relevant studies culled from four separate electronic sources. On consolidating the most relevant research articles, five primary types of IoT-enabled smart building data management and analytics challenges were found in this study. The challenges are presented in Table 2.7 and the solution to these challenges are presented in Table 2.8. The challenges identified in the literature helped us identify the research gaps and formulate the research questions for this thesis (Table 2.9). A review of the literature identifies that a coherent framework with an integrated reference architecture and metamodel to address BDMA challenges for IoT-enabled smart buildings is missing in the literature. This review identifies both issues and potential solutions, and it will serve as a resource for scholars and practitioners involved in IoT-enabled smart buildings, big data management, and analytics. This chapter focussed on the SLR which serves as the first step of DSR adopted from (Hevner 2007) and (Peffers et al. 2007). In chapter 3, we discuss the DSR approach in detail which is the methodology used to perform this research.

Chapter 3 : Research Method

This chapter outlines the research method applied in performing this research. A research methodology outlines the methods a researcher should adopt in the development, implementation and evaluation of an artefact. The selection of a research methodology is based on the research problem and its underlying objectives as specified in Chapter 1. The selection of the appropriate research methodology also depends on the outcome of the SLR as presented in Chapter 2. Based on the research question, research objectives and the SLR, numerous research methodologies could be adopted. However, the nature of the research question and its underlying objectives requires a practical iterative approach to achieve the research aims specified in Chapter 1. Thus, a DSR method based on guidelines published by (Hevner 2007) and (Peffers et al. 2007) was selected to develop and evaluate the proposed IBDMA framework as the DSR method is the most suitable approach to investigate and iteratively produce the IBDMA framework. This chapter presents a review of the research methods considered for this research and provides a justification for choosing DSR as the research method for this research. The chapter then presents the iterative approach used in the DSR to identify the research problem, analyse the SLR data, design, develop and evaluate the proposed IBDMA and outline the thesis output.

3.1. Review of Research Methods

There are numerous ways to categorise research methods. The most prominent distinction of research methods is qualitative and quantitative research methods. Qualitative research may be appropriate for focusing on a detailed description of the problem and the development of a new hypothesis. This research focuses on addressing the challenges of managing and analysing big data generated by IoT sensors deployed in smart buildings by developing IBDMA framework. To comprehend and explain the topic, the most fundamental qualitative research includes the gathering and use of qualitative data such as interviews, publications, conversations, observations, empirical case studies, visual text, and introspection. This section examines some of the qualitative research methods that could be used in this research.

3.1.1. Case study

Case study research is the most often utilised research method in information systems (Smith 1990; Zainal 2007). This is considered a reliable research method, particularly when a comprehensive analysis of the issue at hand is required. The case study research method is characterised in a variety of ways, nevertheless (Yin 2012) defines the case study as an experimental exploration that investigates an novel research area inside its real-world environment, particularly when the segregation between a research area and context are not quite evident. The case study research method finds many applications in the areas of social science, information systems and logistic research (Caplinskas & Vasilecas 2004; Orlikowski & Baroudi 1991; Toomer & Bowen 1993). Despite the case study's importance in IS research, the case study research technique essentially lacks the ability to produce generalised conclusions (s) as well as the well-known lack of statistical reliability and validity (Gummesson 2000).

3.1.2. Action design research

Action design research (ADR) is defined in (Sein et al. 2011) as “a research method for generating prescriptive design knowledge through building and evaluating ensemble IT artefacts in an organizational setting”. It was created to aid IS practitioners by allowing them to apply what they learned in real-world situations while also contributing to the body of knowledge (Rogerson, Scott & Education 2014). After (Sein et al. 2011), numerous research endeavours (Haj-Bolouri, Bernhardsson & Rossi 2016; Hilpert et al. 2013; Huhtamäki 2016; Keijzer-Broers, Florez-Atehortua & De Reuver 2016; Lempinen, Rossi & Tuunainen 2012) began or shifted to ADR as their primary research strategy. ADR is a kind of design research (Hevner et al. 2004; March & Smith 1995) that incorporates industry feedback early in the design and development of the artefact, emphasising development-implementation-evaluation iterations as a replacement for the stage-gate method, allowing researchers to shape the research outcome in collaboration with industry practitioners throughout the research period. ADR was developed (Sein et al. 2011) and identified by the design scholars as a kind of design research (Gregor & Hevner 2013; Iivari 2015) because it requires the research outcomes to find a solution to a specific real-world problem using specific artefacts (Keijzer-Broers, Florez-Atehortua & De Reuver 2016) or refined design knowledge such as in (Lempinen, Rossi & Tuunainen 2012).

3.1.3. Design science research

Design research and design science research are two approaches to solving a problem in research. Design research is more suited to practice-based research (Österle et al. 2011), while design science research incorporates scientific and theoretical contributions (Winter 2008). For the purpose of this research work and for simplicity’s sake, we refer to design science research as design research. This form of research aims to develop knowledge for artefact design as well as create and assess artefacts in order to solve a real-world problem (Dresch, Lacerda & Antunes 2015). The use of design science research, with its emphasis on problem-solving, may be able to bridge the current gap between theory and practice (Romme 2003; Van Aken 2005). Because of the balance it provides between study relevance and rigour, it is commonly used in IS research. (Benbasat, Goldstein & Mead 1987; Deng & Ji 2018; Hevner & Chatterjee 2010). As a result, the application of design science research to IS research could be beneficial (Arnott & Pervan 2008; Goes 2014). The DSR approach has been used successfully in a number of research projects and publications including the areas of configuration information system architecture (Gill & Chew 2019), IoT-enabled digital information systems (Dasgupta, Gill & Hussain 2019), and the areas of BDMA (Litchfield & Althouse 2014).

The DSR method was selected as the most appropriate for the objectives defined in this research. The next section provides the rationale for choosing DSR as well as details of the DSR research method employed in this thesis.

3.1.4. Rationale for choosing DSR

The reasons for choosing DSR for this research are:

- Firstly, DSR appears suitable for the research problem at hand as it involves an incremental development of IBDMA and its evaluation.
- Secondly, DSR is effective in solving real-world problems due to its emphasis on problem-solving. Our research focuses on addressing the challenge of BDMA for IoT-enabled smart buildings through the design and development of IBDMA.
- Thirdly, the implementation of design science research can possibly reduce the current gap between theory and practice. This research aims to address the research gap in the area of BDMA for IoT-enabled smart buildings.

A high-level overview of the DSR method is presented in Figure 3.1.

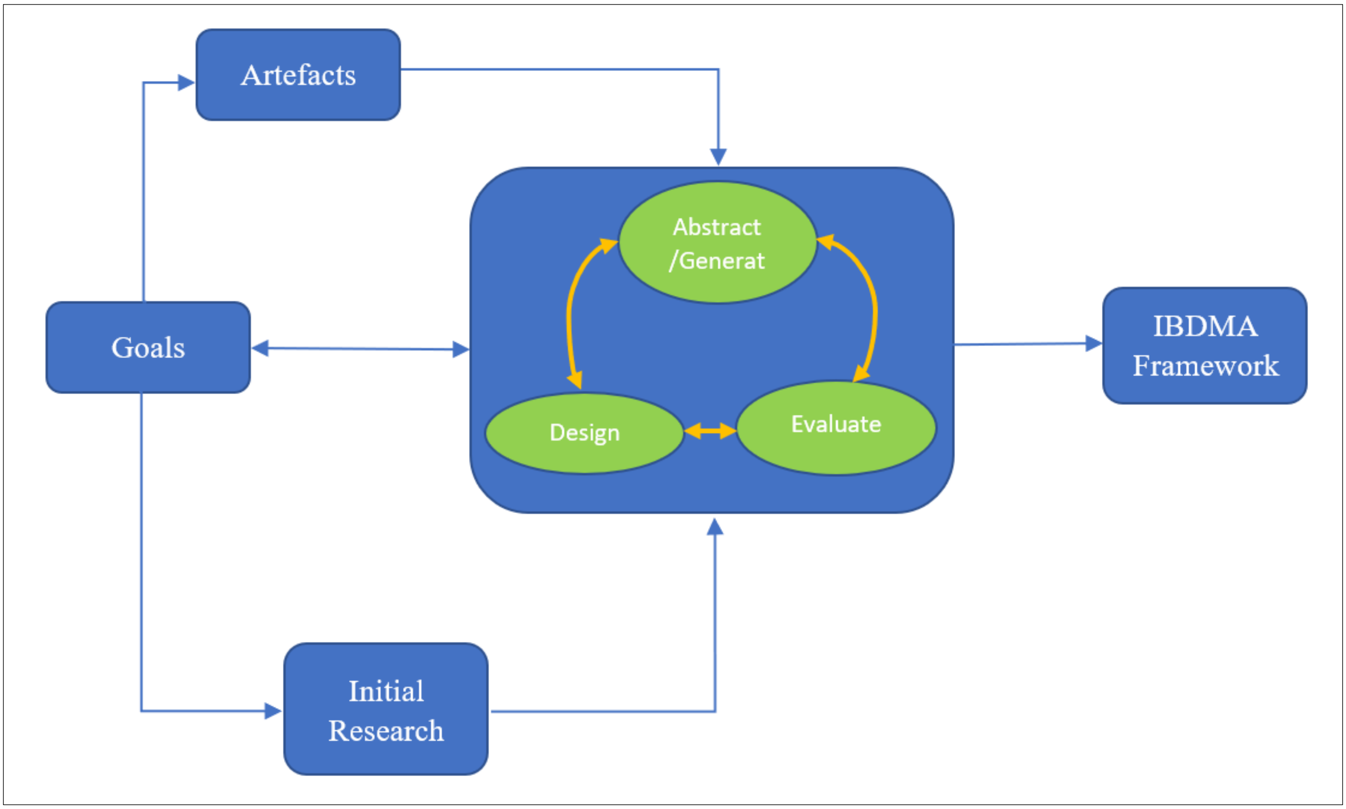


Figure 3.1 Design Research Approach (Hevner & Chatterjee 2010)

A description of each of the high-level steps involved in DSR is presented in Table 3.1.

Table 3.1 DSR Steps

Steps	Description
Purpose and Scope	The DSR's goal is to produce a transparent, systematic approach that can be utilised to develop the IBDMA artefact. The DSR systematic approach is a well-known approach that relies on Gregor and Hevner (2013) and Peffers et al. (2007)'s guidelines. The DSR method is described as follows: The DSR technique uses the rigorous findings from the initial study to identify the research challenge (see Chapter 1, research background and related work and see Chapter 2, SLR results). The DSR's objective is to accomplish the research objectives and produce an IBDMA artefact that meets the thesis scope (see Chapter 1), and research aims (see Chapter1). The DSR approach includes a method for determining whether IBDMA is appropriate for its intended use (addressing the research gaps explained in Chapter 2).

Knowledge of initial research	DSR leverages the preliminary research and results from the SLR analysis to assess what is known about the research topics and the research gaps that indicate the research problem. Recent articles that address the research question are also included in the initial research (e.g., Bashir et al. 2020, Bashir & Gill 2016, Bashir & Gill 2017).
Abstraction and generalisation	The goal of DSR is to create a comprehensive reference architecture based on shared terminologies. To overcome the research gaps, the reference architecture can be utilised to simulate the IBDMA design models in any scenario. The general characteristics in the reference architecture are based on BDMA concepts and terminologies in the SLR results in chapter 2.
Design and development of artefacts	DSR aims to create a comprehensive architectural design artefact that enables the effective management and analytics of big data for IoT-enabled smart buildings and address the challenges and research gaps explained in Chapter 2. The DSR design and development results in a new IBDMA framework with generic characteristics, an abstract design, and reusable framework components. The IBDMA framework is not tied to any specific instance. Instead, depending on the development environment, the suggested IBDMA framework can be used in a variety of situations. The IBDMA framework provides new knowledge about the BDMA challenges for IoT-enabled smart buildings.
Evaluation	The DSR contains an evaluation method (empirical evaluation) to evaluate the proposed IBDMA framework and determine if it is suitable for its purpose (e.g., answering the research question and fulfilling the thesis goals within the research scope). The goal of the IBDMA framework evaluation is to see if it addresses the research gaps indicated in Chapter 2.
Justificatory knowledge and output	The IBDMA framework's effectiveness and applicability are assessed in Chapter 5, utilising instruments that employ indicative measuring approaches to analyse evaluation data. The project outputs (IBDMA artefacts) include the proposed IBDMA framework restrictions and future research suggestions.

3.2. DSR Method

This research uses the well-known DSR approach (Gregor & Hevner 2013) (Peffers et al. 2007) which is a set of principles, practises, and procedures for developing and disseminating high-quality research artefacts in a specific field of study. The DSR tries to make provable contributions through the design, development, and assessment of an object. The artefact development process may include reviewing current ideas and information to create a solution or object for the intended purpose and audiences. In the DSR, there are three primary process stages (see Figure 3.2):

- Stage 1— Main DSR flows. This stage has two distinct flows:
    - Preliminary research and SLR outcomes: Research background, related work analysis, SLR analysis and outcomes.
    - IBDMA Framework: The IBDMA framework development (create the IBDMA framework reference architecture and metamodel).
- Stage 2— Steps in the DSR process. In this thesis, the DSR method is divided into six steps.:
    - Identification of problem: The research context and associated work analysis provided in Chapter 1 aided in the identification of the research problem, research question, and underlying aims. This preliminary research established the project's scope and assisted in identifying the research aims.
    - Analysis: The SLR conducted in Chapter 2 provided contextual information about BDMA, smart buildings and IoT, and how they are related to each other. The SLR results identified the key challenges in the management and analysis of big data generated by IoT sensors in the smart building context (Table 2.6). The solution to some of the challenges that have been addressed in the research were consolidated and presented in Table 2.7. The challenges and solutions assisted in identifying research gaps and, as a result, in determining the scope of this study.
    - Design: This step aims to create a comprehensive framework design model for the proposed IBDMA framework. The main conceptual level elements of the framework are proposed and defined based on the characteristic terminologies derived from the IoT, smart buildings, big data management and big data analysis concepts. The IBDMA framework aims to address the research gaps (see Chapter 2) within the boundaries of the research scope (see Chapter 1, Section 1.2.4).
    - Development: The IBDMA framework components are developed based on the conceptual model. The new comprehensive IBDMA framework aims to address the

research gaps (see Chapter 2). The IBDMA framework (reference architecture and metamodel) is not applicable to one particular situation or environment but can be applied to numerous instances in the context of smart environments.

  - Evaluation: The comprehensive IBDMA framework is evaluated to determine if the IBDMA framework is fit for purpose (addressing the research gaps within the project scope boundaries). The evaluation step involves evaluating the framework components using practical use cases as well as being evaluated by the industry experts.
  - Output: The research journey and output, the project's key contributions and publications, the project's limitations, and future research are compiled.

- Stage 3—DSR outcomes. There are four outputs from the DSR approach in this research:

  A. Design artefact (see Chapter 4).

  B. Development artefact (see Chapter 4).

  C. Empirical evaluation (see Chapter 5).

  D. Thesis results and discussion (see Chapter 6).

Since DSR consists of six steps with the first two steps being problem identification and analysis respectively, there is an outcome for each of these DSR steps. These outcomes are presented in Figure 3.2.

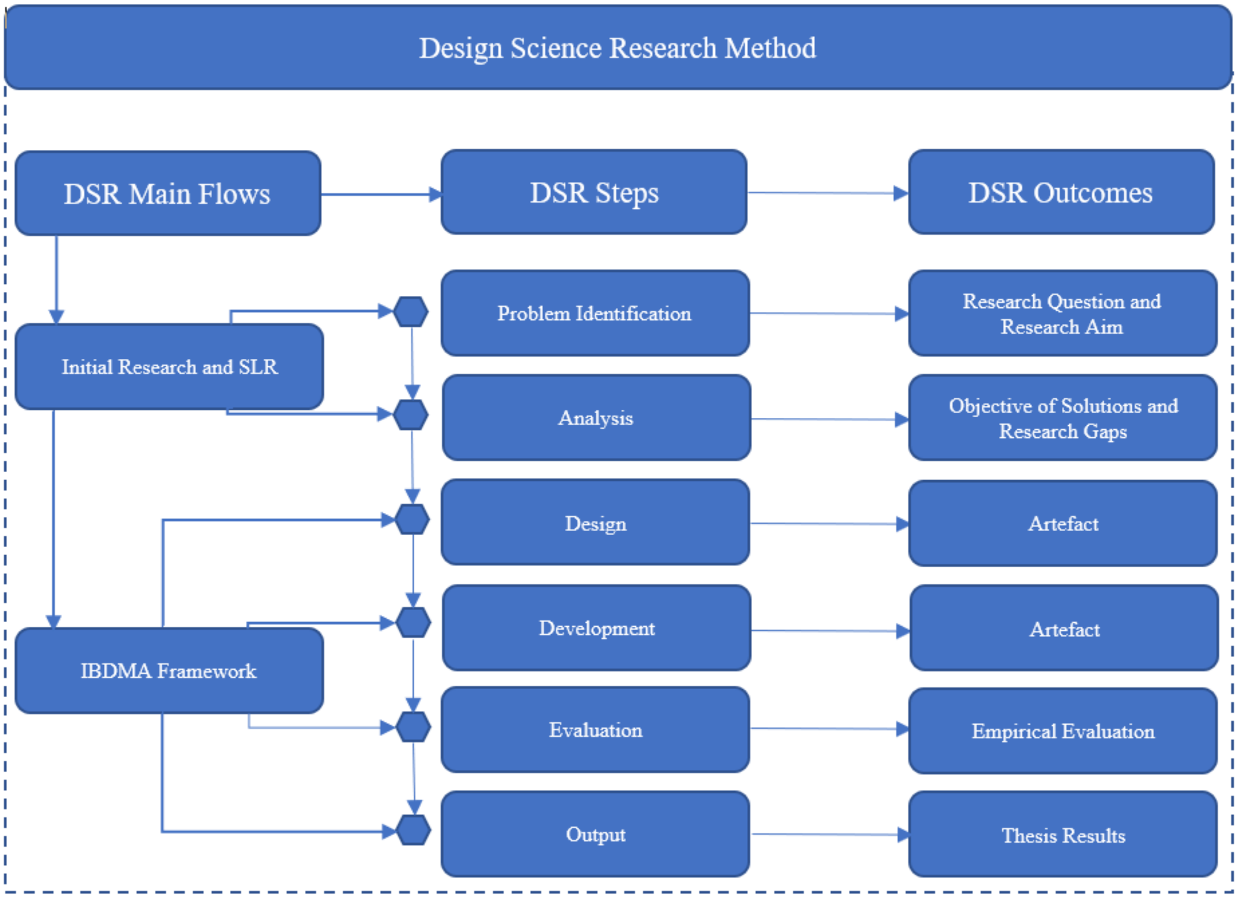


Figure 3.2 DSR Process

3.2.1. Problem identification

The main research question is: How to manage and analyse IoT big data for the efficient and effective management of smart buildings? as outlined in Section 1.1.4.

The background studies and related work analysis reported in Chapter 1, and the SLR reported in Chapter 2 comprise the first step of the DSR. The research problem was identified primarily due to the analysis reported in Chapter 1. The research question (as well as its underlying objectives) was developed in response to the issues raised in the section on the research gaps (Chapter 2). The research gaps discussed in Chapter 2 identify the challenges in the management and analysis of big data generated by IoT-enabled smart buildings. The research question was subdivided into RQ1 and RQ2:

RQ1: How to effectively manage and analyse IoT big data in the context of smart buildings?

RQ2: How to holistically identify all the elements and the relationship between these elements to effectively manage and analyse big data in IoT-enabled smart buildings?

Figure 3.3 Problem Identification

3.2.2. Analysis

The DSR method's analysis step examines and synthesizes the available data and resources on big data, IoT, and smart buildings (see Figure 3.4). The SLR in Chapter 2 presented a plethora of information about big data and its relationship to research subjects like IoT and smart buildings. The SLR results also revealed critical research gaps (Chapter 2). The research gaps constitute the challenges for the research aim (Chapter 1) concerning the BDMA challenges for IoT-enabled smart buildings. The following suggestions are in line with the research aims (Chapter 1) and are intended to provide practical answers to the research gaps (Chapter 2):

- Define IBDMA conceptual elements: IBDMA conceptual elements are foundational elements used to provide a foundation for the development of the IBDMA framework.
- Develop an IBDMA reference architecture: The IBDMA reference architecture provides a generic architectural design model that is not fixed to a particular instance but can be applied to numerous instances.
- Develop an IBDMA metamodel: The IBDMA metamodel provides a generic metamodel listing the foundational elements and the relationship between these elements.

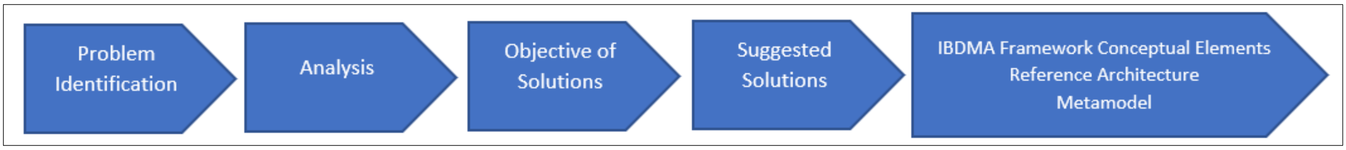


Figure 3.4 Analysis

3.2.3. Design

In this stage, the IBDMA framework's conceptual elements, reference architecture, and metamodel are developed to efficiently handle and analyse the massive data generated by IoT sensors deployed in smart buildings. The proposed IBDMA reference architecture and metamodel are based on conceptual elements of the IBDMA framework. The IBDMA reference architecture is a generic architecture model that is not tied to a particular instance but can be applied to numerous development scenarios. To avoid vendor lock-in, the IBDMA reference architecture was mostly built with open-source technologies. The

IBDMA metamodel provides a generic metamodel listing the foundational elements and the relationship between these elements.

The IBDMA framework elements are presented in Figure 3.5:

- IBDMA conceptual elements: A list of the foundational elements of the IBDMA framework for the smart building context.
- IBDMA reference architecture: Represents the physical-level architecture based on the IBDMA conceptual elements.
- IBDMA metamodel: Represents the elements and the relationships between those elements for the effective management and analysis of big data generated by IoT-enabled smart buildings.

The IBDMA reference architecture and metamodel provide template solutions that can be used to address the research gaps (see Chapter 2). The objective of the development of the reference architecture and the metamodel is that they can be instantiated for any development context and scenario. The IBDMA reference architecture provides a practical method to address the BDMA challenges for IoT-enabled smart buildings. The reference architecture and the metamodel provide the answer to research questions RQ1 and RQ2.

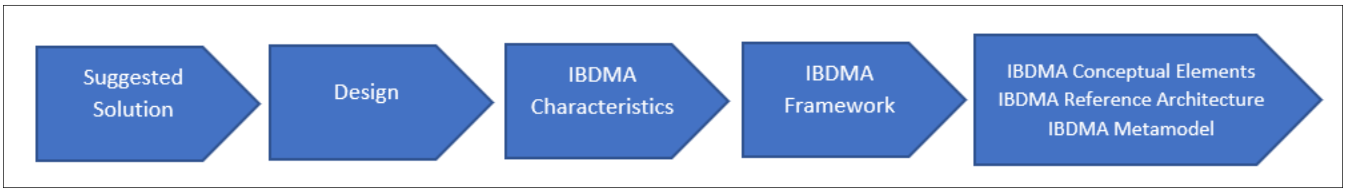


Figure 3.5 Design

3.2.4. Development

The proposed IBDMA framework components are developed in this step. The reference architecture and the metamodel are instantiated using practical use cases within the smart building scenario.

The IBDMA framework provides a new comprehensive artefact development in this step of the DSR method to provide practical solutions to the research gaps (Chapter 2) and address the research problem highlighted in the research questions (RQ1 and RQ2).

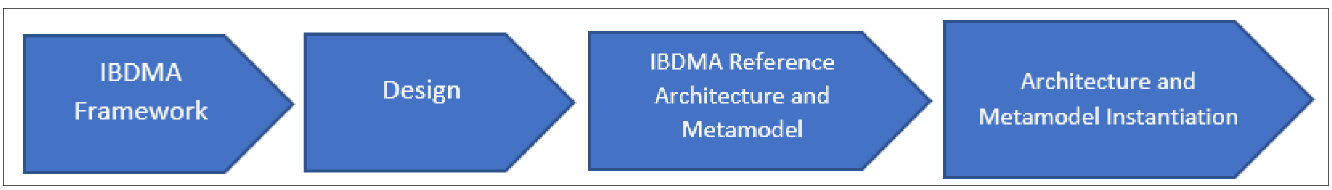


Figure 3.6 Development

3.2.5. Evaluation

The reference architecture and metamodel of the IBDMA framework is evaluated using an empirical evaluation process (see Figure 3.7). The evaluation process can be sub-divided into two categories:

1. Evaluation using practical use cases against the selected evaluation criteria found in the literature.

The DSR is an iterative approach to artefact development. Each iteration of the IBDMA framework is evaluated against the thirteen evaluation criteria outlined in Figure 3.8 as adopted from (Prat, Comyn-Wattiau & Akoka 2014). The thirteen evaluation criteria are divided into five groups namely: *goal, environment, structure, activity* and *evolution*.

Under *goal*, there are three evaluation criteria namely: *efficacy* (EC1), *validity* (EC2) and *generality* (EC3). *Environment* comprises two evaluation criteria: *consistency with people* (EC4) and *consistency with technology* (EC5). The third group *structure* comprises three evaluation criteria: *completeness* (EC6), *simplicity* (EC7) and *clarity* (EC8). The fourth group *activity* comprises three evaluation criteria: *accuracy* (EC9), *performance* (EC10) and *efficiency* (EC11). The fifth group *evolution* comprises two evaluation criteria: *robustness* (EC12) and *scalability* (EC13).

It took five iterations for us to bring the IBDMA framework to a stage where it satisfied all thirteen evaluation criteria. Each iteration of the IBDMA framework was evaluated against the thirteen evaluation criteria using practical use cases. The details of each test case (context, problem, solution, test metrics and consequence) are presented in Appendix G. The evaluation results for the first four iterations suggested that they did not satisfy all the evaluation criteria and hence a new iteration was required. Iteration five, which is the final version of IBDMA framework, satisfied all thirteen evaluation criteria.

2. Evaluation by industry experts

The evaluation by the experts involved a presentation and demo that was presented to the industry experts. The industry experts, who have extensive experience in the fields of IoT or big data, were requested to participate to provide their feedback by evaluating the IBDMA framework against the thirteen evaluation criteria detailed in Figure 3.8. The industry experts were presented the fifth and final iteration of the framework which satisfied the 13 evaluation criteria. The outcome of the industry experts' evaluation proved that the final framework satisfied the research questions and the 13 evaluation criteria. The industry experts' evaluation design and analysis are presented in section

3.2.5.2. The industry experts pointed out the future research directions which are included in the future work in Chapter 6 (Section 6.3).

Figure 3.7 Evaluation

3.2.5.1. Design of use cases

The use case methodology is an empirical research method that can be used to test, generate or describe a theory or phenomenon (Runeson & Höst 2009) which is widely used in software engineering applications. A case study represents any contemporary phenomenon in a real-life context (Aberdeen 2013; Jedlitschka & Pfahl 2005). For the evaluation, we choose 13 evaluation criteria found in the literature (Prat, Comyn-Wattiau & Akoka 2014) against which each use case will be tested and evaluated. The 13 evaluation criteria are categorised into 5 groups: *goal*, *environment*, *structure*, *activity* and *evolution* as presented in Figure 3.8. If the use cases satisfy the criteria, it will be marked as 'satisfied', otherwise it will be marked as 'failed'. The following steps are involved in the use case evaluation methodology.

1. Design: The use cases will be chosen so that they encompass all possible practical scenarios that can occur in a smart building in a real-life environment. While defining the scope of use cases, special attention is given to the fact that the use cases test and evaluate all the important elements of the framework to prove their completeness and effectiveness. The results of each use case are recorded against the 13 evaluation criteria. For each use case, we define the context, the problem statement, the solution to the problem, the test metrices, a description of the issue and the consequences of the use case (see Chapter 5 (Section 5.3) for further details).
2. Preparation of the data collection: The methodology for data collection needs to be defined for the use cases. Since this research focuses on the data generated by IoT sensors deployed inside smart buildings, we use the data generated by IoT sensors to evaluate the IBDMA framework.
3. Collecting data. The data for the evaluation is collected using the reference architecture instance by developing a pipeline which ingests data from IoT sensors and stores it in Hadoop.

4. Data analysis: The data generated by the IoT sensors once stored in Hadoop is analysed using Spark code to make useful and autonomous decisions on controlling smart buildings.
5. Reporting. The findings of each use case are reported against the 13 evaluation criteria to check if each use case satisfied the criteria or not as illustrated in Figure 3.8.

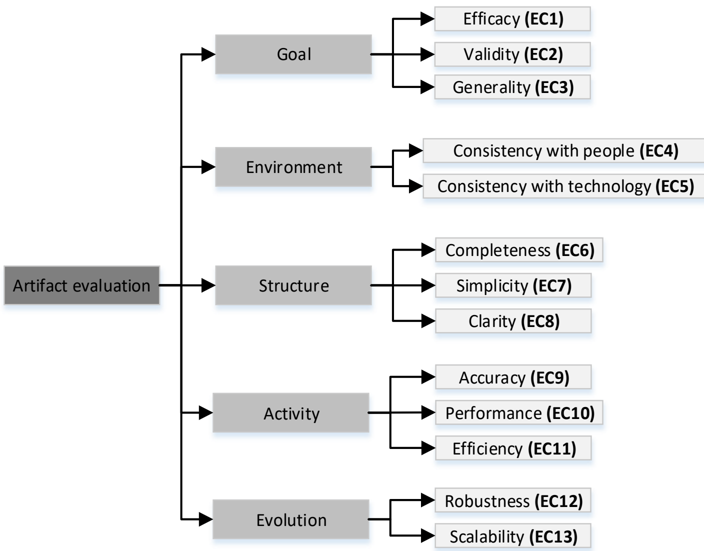


Figure 3.8 Evaluation Criteria for IBDMA Evaluation

3.2.5.2. Industry expert evaluation design

This section details the steps performed in the industry experts' evaluation of the IBDMA framework. The evaluation followed a five-step process.

1. Design: The participants were contacted through LinkedIn by searching for people based in Australia who have been working in big data or IoT domains, who had extensive experience in either of these two domains and who were at the manager level or above. The participants were requested to participate in the research by providing their feedback. The participants who agreed were then given the Participant Information Sheet and the Consent Form (Appendix B).
2. Preparation of data collection: The participant information sheet gave the participants preliminary details including a brief about the research work being conducted and what their participation would involve. The participants were informed that their participation was voluntary, and they could withdraw their participation at any stage if they so desired.
3. Data collection: On the day of the evaluation, the participants were given a presentation on the IBDMA framework. At the conclusion of the presentation, a demo was presented to the

participants. All questions raised by the participants were answered. At the end of the session, a questionnaire comprising five questions was given to all the participants for them to provide written feedback regarding the IBDMA framework. The sample questionnaire is presented in Appendix G. The five questions were:

a. Please provide comments about the IBDMA reference architecture. Does it fulfil its intended purpose?
b. Please provide comments about the IBDMA metamodel. Does it fulfil its intended purpose?
c. Please provide comments on the tools used to implement the IBDMA framework.
d. Please provide feedback on the setup and implementation process of the IBDMA framework.
e. Please provide your overall feedback on the research (including any areas of improvement).

4. Data analysis: The experts' feedback collected during the experiment is analysed in Table 5.20. The data was analysed using the cross-examination method between participants' feedback and the thirteen evaluation criteria in Figure 3.8. This analysis aims to connect or relate the hypotheses (evaluation criteria) to the experts' feedback. The output of the analysis is organised into two columns: *participant feedback*, which is the experts' feedback on the evaluation of the IBDMA framework, and *IBDMA evaluation criteria categories*, which is the relationship between the feedback and the evaluation criteria.
5. Reporting: The outcome of the industry expert evaluation is presented in this section. The report aims to draw a conclusion from the industry experts' point of view and feedback about the IBDMA. Table 5.21 presents the systematic testing procedure and outcome of the industry expert evaluation.

3.2.6. Output

The output of this research is an IBDMA framework for IoT-enabled smart buildings. The framework comprises two elements:

1. Reference architecture: This represents the physical level architecture based on the IBDMA conceptual elements.

2. Metamodel: This represents the elements and the relationships between these elements for the effective management and analysis of big data generated by IoT-enabled smart buildings.

In this final step of the research, the research limitations and suggestions for possible future research work are also discussed.

3.3. Research Instruments

This section presents the instruments used in the design and development of the proposed IBDMA framework. The research instruments implemented in this thesis are resources and research ethics.

3.3.1. Resources

The resources used for the design, development and evaluation of the proposed IBDMA are outlined in Table 3.2. The resources used to develop and evaluate the IBDMA framework are a combination of IoT sensors, software development, data ingestion and storage tools, data analysis and visualization tools, actuators, smart building facilities, practical use cases and industry experts' evaluation. The IBDMA framework development process aims at addressing the research question and providing a practical working solution to the research gaps in the management and analysis of IoT data in smart buildings, as evidenced by the integration of BDMA tools and techniques with IoT sensors and smart building facilities.

Table 3.2 Resources for IBDMA

Resources	Description	Reference
Data	- Background and research problem - SLR review and analysis - SLR results	- Section 1.1 - Section 2.2 - Section 2.3
Method	DSR method for IBDMA design, development and evaluation	- Section 3.2
BDMA challenges	- 17 BDMA challenges	- Table 2.6
Software development	- Programming languages (Python, R) - Development IDE (PyCharm)	- Table 5.1 - Table 5.1
IoT sensors	- Luminosity, Oxygen, Smoke Detection, Parking, Garbage detection.	- Section 4.3
Data ingestion tool	- Apache Flume	- Table 5.1
Data Storage tool	- Hadoop distributed file system	- Table 5.1

	- Elasticsearch	- Table 5.1
Data analysis tool	- Apache Spark	- Table 5.1
Data visualization tool	- Microsoft Power BI - Tableau - Kibana	- Table 5.1 - Table 5.1 - Table 5.1
Data forecasting	- Programming Language R - ARIMA time series forecasting used - Power BI for data visualization	- Table 5.1 - Section 4.3.6 - Table 5.1
Evaluation Criteria	- 13 evaluation criteria adopted from the literature	- Section 3.2
Evaluation through use cases	- Practical use cases to evaluate the IBDMA framework	- Appendix G
Evaluation by industry experts	- Questionnaire for industry experts' feedback	- Appendix H

3.3.2. Research Ethics

In accordance with the University of Technology Sydney's research ethics procedures, formal approval was acquired from the UTS Research Ethics Committee. Appendix B contains the approval document. There were no ethical concerns raised by the study. Each participant received a formal consent letter (see Appendix C). The participants had the option to leave the study at any moment and could contact the supervisor or the institution to do so. Along with the consent form, willing participants received additional documents with information about what their involvement would entail and about the IBDMA framework. The forms provided clear information about the project, the feedback questionnaire, data collecting anonymity, and storage.

3.4. Summary

This study was carried out to create a new framework, the IBDMA framework, for big data management and analytics in IoT-enabled smart buildings. The DSR in this thesis was created using the guidelines presented in (Gregor & Hevner 2013) and (Peffers et al. 2007). The research method and resources used to perform the research were outlined in this chapter. The evaluation techniques for the

IBDMA framework reference architecture and metamodel were also discussed in this chapter. In the next chapter, the framework development is presented in detail as an output of the DSR method.

Chapter 4 : IBDMA Framework

This chapter presents the IBDMA framework, which is the primary contribution of this research. The IBDMA framework provides a practical solution to the research questions identified in chapter 1. The IBDMA framework was developed using the DSR method discussed in chapter 3. The IBDMA framework is designed to address the issues of managing and analysing big data produced by IoT sensors installed in smart buildings, as well as the research gaps identified in Chapter 2. The framework consists of two components: 1) reference architecture; and 2) metamodel. In this chapter, first the key conceptual elements of the IBDMA framework are presented, then the reference architecture development process is presented and finally the IBDMA metamodel is presented. The IBDMA reference architecture and the metamodel are evaluated in Chapter 5.

4.1. IBDMA Framework

The IBDMA framework aims at addressing the BDMA challenges for smart building by providing an integrated and coherent framework comprising of a reference architecture and a metamodel. Hence, it consists of two components, 1) reference architecture; and 2) metamodel as shown in Figure 4.1. The reference architecture addresses the first research question (**RQ1**: How to effectively manage and analyse IoT big data in a smart building context?) by providing a comprehensive reference architecture which allows the effective management and analysis of IoT big data in smart buildings. And the metamodel addresses the second research question (**RQ2**: How to holistically identify all the elements and the relationship between these elements to effectively manage and analyse big data in IoT-enabled smart buildings?) by providing a holistic view of all the elements and the relationship between those elements that are used in the effective management and analysis of IoT big data in smart buildings.

The development process of the reference architecture and the metamodel is discussed separately in the following sections. The research work on the metamodel development has been published in IEEE Access (Bashir et al. 2020). The IBDMA framework will assist professionals and researchers working in big data and IoT domains to address the BDMA challenges.

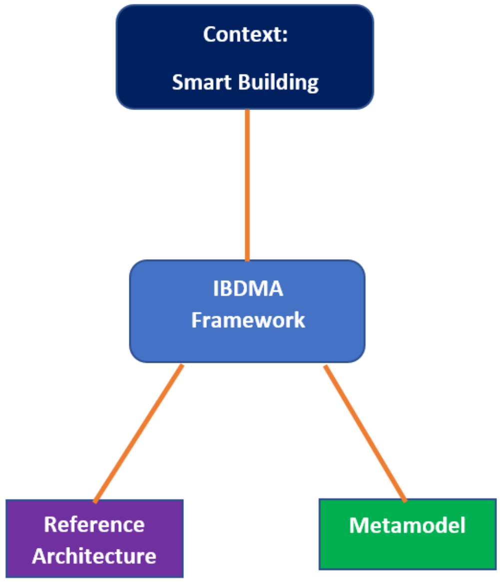


Figure 4.1 IBDMA framework

The IBDMA framework is designed to provide a scalable and adaptive architecture as advocated in (Gill 2015a, 2015b) with the aim of addressing the BDMA challenges in smart buildings. This means that the framework can be extended to other smart environments (e.g., smart cities). There is no coherent framework with an integrated reference architecture and metamodel to address this challenging and critical research area. The framework has five key conceptual level elements as shown in Figure 4.2.

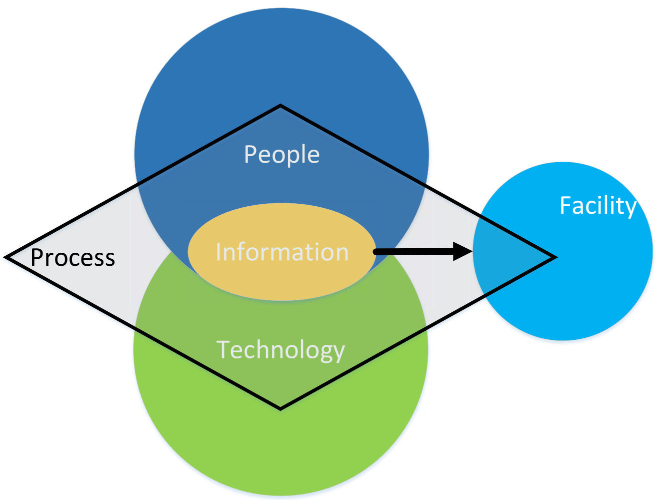


Figure 4.2 IBDMA framework – conceptual elements

The *people* element is the core of the IBDMA framework. This comprises the residents, policy-makers and developers of smart buildings. The policymakers put together policies which support and regulate smart buildings systems. The developers build smart buildings that adhere to the policies developed by the policymakers. The residents may include students, staff, homeowners, shop owners etc. The developers and policymakers make policies which assist in identifying the process element of IBDMA. The process element encompasses all the processes which are required for the effective management and analysis of smart building data. The technology element comprises of the technology stack that supports the processes as outlined by the process element. The overlap of these elements results in useful information which makes up the fourth element of IBDMA known as information. The information is then autonomously utilised to regulate numerous facilities within smart buildings, which fall under the facility element of IBDMA. The process element links all other elements, as shown in Figure 4.2. The five core elements of the framework are generic and can be extended for other use cases that are not covered in this research. For instance, the same framework can be extended for monitoring people movement; this holds true for both the reference architecture and the metamodel. The way these different elements are linked and interact with each other is explained in the following sections.

4.1.1. People

The people element is the first element in IBDMA. It consists of policy makers, developers and building residents, as demonstrated in Figure 4.3. These can be broken down into two groups, one consisting of policy makers and developers, and the other consisting of building residents. The policy makers define the policies which govern the building. Key requirements from the stakeholders are defined by the stakeholders through these policies which in-turn helps in proposing optimum solutions to meet stakeholders' expectations.

The developers comprise of the building developers who develop the smart building in light of the policies and regulations. Their role is to safeguard the residents' safety and security. Smart building residents may include students, staff, tenants, homeowners, shopkeeper etc. depending on the nature of the smart building. They are considered the users of the smart buildings.

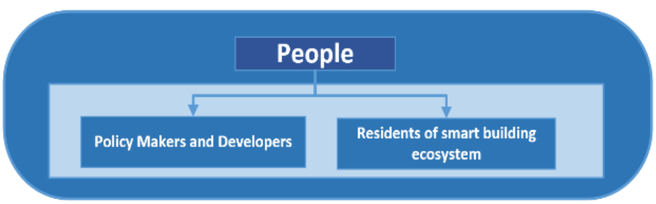


Figure 4.3 The first element of IBDMA – People

IBDMA suggests that the processes necessary for the efficient execution of these policies be developed based on the policies established by the policy makers. These processes aid in determining the technology stack needed for these processes to be carried out correctly. This includes software applications and tools required for the execution of the processes e.g. Microsoft Power BI (Microsoft 2020) and Tableau (Tableau 2020) for data visualization, Apache Flume (*Apache Flume* 2018) for data ingestion, Apache Spark (*Apache Spark*) for data analysis etc. The applicability and usability of each of these tools and the process elements is presented in detail in the following sections.

Building developers and policy makers (people) clearly articulate the requirements for the smart building. After this, the processes are identified and executed. This includes ingestion of IoT data, storing it and then analysing it. Hence, the *process* element is the second element that follows when applying the framework.

4.1.2. Process

The second element of the IBDMA framework is *process*. It plays a crucial part in formulating the plan for putting the IBDMA framework into practise. Processes describe the steps involved in an operation and how they should be coupled to produce a clear and practical solution. Thus, the IBDMA processes outlined should be transparent and streamlined to have an efficient solution.

Goals, requirements, and guidelines defined by people serve as inputs for defining processes and form the basis for process selection and implementation. As this research focuses on the collection, management, and analysis of data generated by IoT sensors deployed in smart buildings, the IBDMA framework is a smart solution for IoT data acquisition, data collection, and storage. We propose the following process for implementation, including building environment monitoring. Analysis of data at a central location, analysis of data in near real time, decision making, visualization of data in near real time, and near real time autonomy of various intelligent devices in smart buildings, as shown in Figure 4.4. control. Utilising the proposed framework, various smart building facilities can be autonomously controlled. However, to ensure a realistic scope for this research, we consider five facilities, oxygen level control, luminosity level control, garbage management, parking management and fire management.

In all IoT systems, the first process in implementing a BDMA infrastructure is monitoring the environment in which the IoT sensors are deployed, which for this research is the smart building. Various types of IoT sensors are available today and can be used to monitor different parameters and attributes of smart buildings, depending on the use case and requirements of residents and stakeholders. These sensors generate data while monitoring the environment in which they operate. The output of these sensors can be binary or continuous, depending on the type and type of IoT sensor. Data generated by these sensors is ingested into a central repository using an ingestion pipeline. Data for cleansing, editing and further processing are stored in a central repository. Once the data is in a central location, it is ready for analysis. The nature of this analysis will depend on the specific use case or stakeholder requirements. The analytics process is the process of gaining useful insights about smart buildings. The output of the analysis helps us in decision making and autonomously controlling the smart building. Using the analysis outcomes, the IBDMA framework enables us to control and manage various facilities in the smart building with an aim to improve residents' comfort and safety. The process element of the IBDMA, which carries out the primary function of integrating all the elements of the IBDMA as shown in Figure 4.2, encompasses all of these various processes, from monitoring to ingestion, from storage to analysis, and from decision-making to autonomous control of the smart building. The subsequent sections provide a more thorough implementation, where it will be clearer how various IBDMA components work together to create a practical and effective solution.

The *technology* stack is needed to implement the processes once they have been specified in accordance with the requirements acquired by the *people*. The success of an effective solution depends on selecting the appropriate tools and software solutions. The third aspect to be considered in IBDMA is the technological element, which includes these tools, technologies, and software packages.

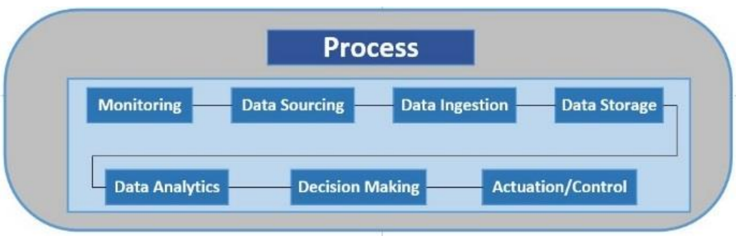


Figure 4.4 The second element of IBDMA – Process

4.1.3. Technology

The third component of the framework is *technology*. It is crucial to the successful implementation of the infrastructure and strategy for big data management. So, picking the appropriate technological stack is crucial. As depicted in Figure 4.5, technology includes software programmes and tools used to efficiently develop and execute IBDMA.

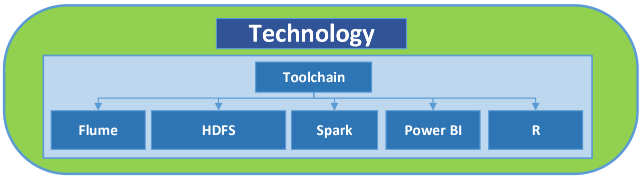


Figure 4.5 The third element of IBDMA – Technology

For this study, we use PyCharm (a Python IDE) to build a virtual IoT sensor-based application in the Python programming language. This virtual program modelled the generation of data from IoT sensors installed in smart buildings. Data from IoT sensors is ingested and loaded into HDFS (Hadoop Distributed File System), a high-performance distributed file system that offers Hadoop clusters dependable data access.

In order to reliably transmit the huge amounts of batched and streaming data, such as logs, IoT data, financial data, and other types of data to a central location, the data created by the sensors is ingested into HDFS using data pipelines that were developed using Apache Flume. Flume is a fault-tolerant data ingestion and transportation tool that uses a basic extensible data model to allow for real-time analytical applications.

The data produced by the IoT sensors is analysed using Apache Spark. For a variety of streaming and batch applications, Apache Spark, an in-memory data processing engine, offers quick data analysis and processing capabilities. Resilient Distributed Dataset (RDD), on which its architecture is built, offers a fault-tolerant method of keeping large datasets by distributing them over a cluster of computers. For the sake of this study, we write Spark code in Python. The system is able to control and maintain smart buildings and their numerous facilities autonomously owing to the analysis, which aids in decision-making. The purpose of this autonomy is to provide residents of smart buildings with comfort and security.

The visualisation is done using Microsoft Power BI. Predictive analytics may be carried out within the Power BI environment using code written in R and Python, and Power BI includes a built-in adapter to connect to HDFS. It was therefore an obvious choice as the data visualisation tool for this study. Power BI, like any other tool, has some drawbacks, and it might be challenging to create a near-real time dashboard in Power BI. We need to be able to provide IoT sensor data in a near-real time dashboard because we are dealing with IoT sensor data in this project. This will give us a better understanding of how to monitor the smart building environment so that any alerts may be addressed in near real-time. We pick Elasticsearch (*Elasticsearch* 2018) and Kibana (*Kibana* 2018) to have this near real-time visualisation capability. Elasticsearch is an open-source application that offers a distributed search and analytics engine and was developed on Apache Lucene (*Apache Lucene* 2018). Using an API or an ingestion tool like Logstash (*Logstash* 2018), new incoming data is ingested in Elasticsearch as documents. This adds a searchable reference to the data (document) in the cluster's index and receives and saves incoming data. The Elasticsearch API can then be used to search for and retrieve these documents. When reading documents on an Elasticsearch cluster, the open-source data visualisation plugin Kibana from Elasticsearch offers near real-time graphing features.

4.1.4. Information

IBDMA's fourth component is *information*. It originates from the overlap of the first three elements shown in Figure 4.2. *People* specify the regulations and specifications for smart buildings, as was covered in the preceding sections. These policies serve as the foundation for identifying the *processes* that will define the *technology* stack for their implementation. When the *processes* and the *technology* infrastructure are successfully installed, *information* is produced from the data produced by the IoT sensors. The data produced by IoT sensors can be used to provide a variety of *information* in different formats and types, allowing us to control various building elements and variables to increase comfort and safety for occupants.

Because of the information that is produced, several facilities within the smart building could be autonomously operated. These facilities could involve the operation of vending machines, elevators, HVAC systems, smart parking based on parking sensors, and smart lighting based on brightness levels. IBDMA proposed the autonomous monitoring and control of five facilities, namely: HVAC systems, brightness levels, parking management, garbage management, and fire event management, for this research and to have a constrained scope. This information, according to the IBDMA, will allow the luminosity levels of the room to improve by turning on the lights in that spot, like in the example scenario where the luminosity sensor indicates that brightness levels in a specific location of a building are below a specified level. Generally defined, the IBDMA suggests that the information element also comprises the analysis output produced by Apache Spark, as well as the IoT data visualisation carried out in Power BI and Kibana, as shown in Figure 4.6. Data visualisation, data analysis, and decision-making are examples of processes that couple the information element with the other aspects of the IBDMA, as previously described. These processes are shown in Figure 4.2.

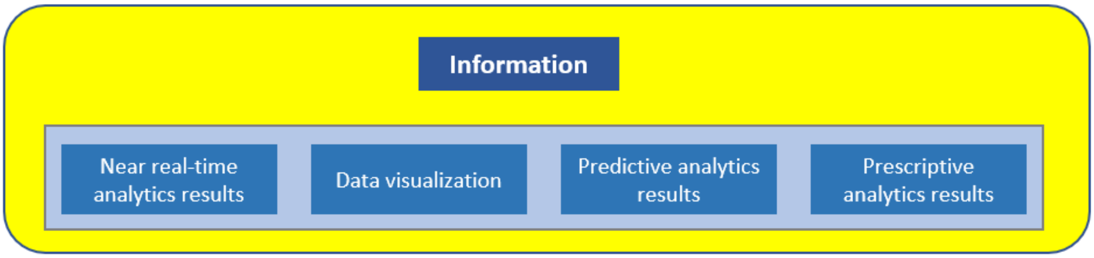


Figure 4.6 The fourth element of IBDMA – Information

4.1.5. Facility

The framework's final element is *facility*. It features a variety of smart building amenities designed to improve occupants' comfort, security, and safety. As shown in Figure 4.2, the knowledge derived from IoT data enables the autonomous control of these facilities. The facilities include garbage management, HVAC systems, and elevator maintenance and management, among others. For the purposes of this study, we focus on the five features of smart buildings shown in Figure 4.7: smart HVAC systems, smart parking, smart garbage monitoring, smart lighting, and smart fire management. When creating the regulations and requirements of the smart building ecosystem, it is important to identify and take into account the target facilities that need to be regulated in the smart building. This lessens the likelihood of running across any major impediments when implementing the big data management infrastructure for IoT-enabled smart buildings.

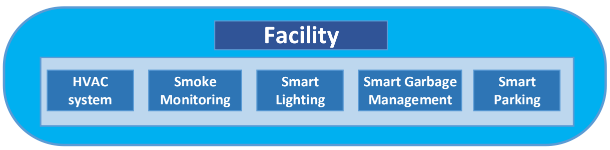


Figure 4.7 The fifth element of IBDMA – Facility

4.2. IBDMA Reference Architecture

As outlined in the previous section, the IBDMA framework has two components i.e., the reference architecture and metamodel, as shown in Figure 4.1. In this section, we focus on the reference architecture which is the first component of the IBDMA framework. The IBDMA reference architecture was developed over the course of five revisions before reaching its culmination. This section also explains how the physical-level elements of IBDMA link to the contextual-level elements, as shown in iterations 4 and 5 in the sections that follow.

In this section, we discuss the final version of the reference architecture. The detail on each of the previous iterations is presented in Appendix I.

The IBDMA reference architecture is presented in Figure 4.12. The reference architecture takes into account both batched data as well as real-time data ingestion. For batched data source, we selected open research data as part of the reference architecture implementation. R programming scripts are used to scrape this open data (this may also be done with Python) and then it is added to HDFS.

The virtual IoT sensors provide data from streaming data sources to two sinks: HDFS and Elasticsearch. The Apache Spark algorithm analyses the IoT sensor data as it enters HDFS in near real-time, enabling decision-making for the efficient administration and control of the five facilities previously mentioned within the smart building. After being stored in HDFS, this data is batch-visualized using Power BI. We utilised R scripts to create an ARIMA model in Power BI for predictive analytics. The data is indexed as it gets ingested into Elasticsearch which acts as our second data source. IoT data may be visualised in near real-time using Elasticsearch's Kibana data visualisation plug-in. It demonstrates the link between the five physical components of the framework and the five conceptual components shown in Figure 4.2. People are at the top-most level, which represent the stakeholders of the smart building environment, such as building developers, building management, IT professionals and residents of the building. The data-driven processes that are pertinent to smart buildings are defined by the process element. This includes monitoring via sensors, gathering data, ingesting it, storing it, analysing it, visualising it, making decisions, and eventually acting on the results. The technical component covers the technological stack, which includes Power BI, Kibana, Spark, Elasticsearch, Flume, and R. The data contains the result of the decision-making process using Spark, the near real-time data visualisations in Kibana, and Power BI dashboards for the IoT data visualisation. The facility element, which also represents the smart building's amenities like HVAC systems, fire alarms, lighting, parking spaces, and waste places, is the last one.

The information on several big data tools utilised in the creation of the IBDMA is summarised in Table 4.1. It lists the processes in which each of these tools is utilised and what each tool's function is in the implementation of the IBDMA reference architecture.

Table 4.1: Elements in the IBDMA architecture and their purpose.

Sr. No	Element	Process	Purpose
1	Flume	Ingestion	For ingesting streaming IoT data in Elasticsearch and HDFS.
2	Elasticsearch	Storage	Indexing streaming data to be visualized in Kibana.

3	Kibana	Visualization	Visualizing streaming IoT data in near real-time.
4	HDFS	Storage	Storing both batched and streaming data.
5	Spark	Analysis/Decision Making	Analysing IoT data in near real-time to enable decision making and actuation of smart facility.
6	Power BI	Visualization	Visualizing batched and forecasted data.
7	R	Ingestion/Decision Making	Web scrapping and predictive analytics.

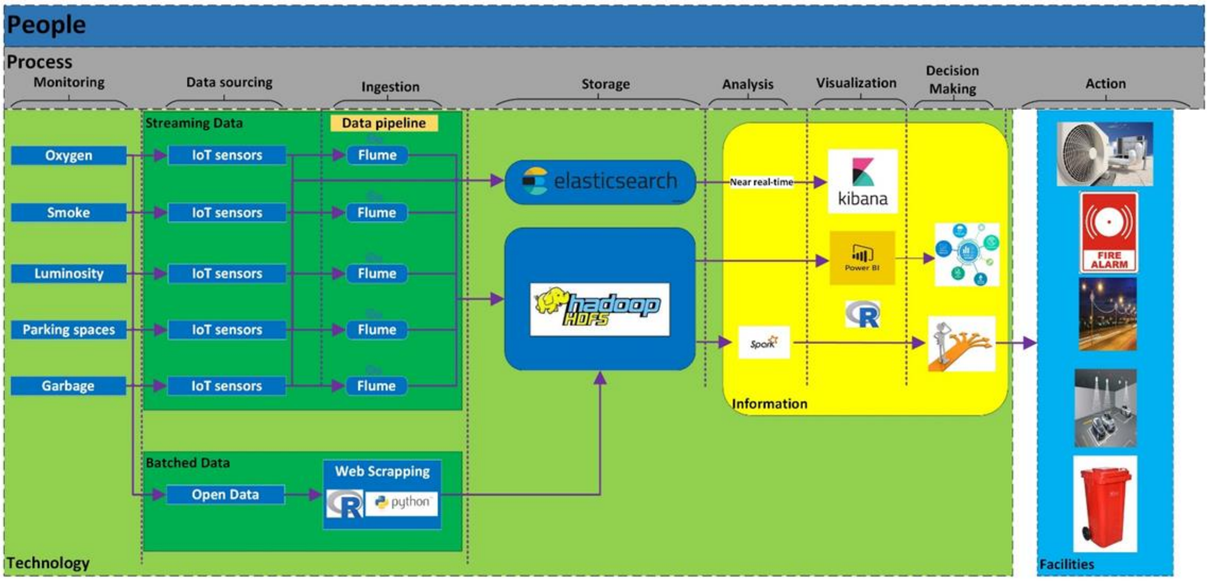


Figure 4.8 Iteration 5 – Updated and improved near real-time data analysis and actuation architecture

4.2.1. Reference Architecture Implementation

The IBDMA architecture is implemented in accordance with the design suggested in Figure 4.12 for a smart building application scenario. We took into account both static and streaming data, which is evident in Figure 4.12. We developed 1000 simulated IoT sensors for streaming data. This is seen in Figure 4.12's section titled Streaming Data. The software programme used to implement these virtual sensors was created in Python. We take into account five main types of sensors for the IoT. We model 200 oxygen sensors, 200 smoke detectors, 200 luminosity sensors, 200 parking space sensors, and 200 rubbish detection sensors out of the total 1000 sensors. These 1,000 sensors are supposed to be placed in 200 various places around the smart building, including rooms and levels. We use the Cloudera VM (Virtual Machine) Hadoop distribution to implement the example scenario, and we use Python to build a virtual sensor application that produces IoT data. The majority of the big data tools (Apache Flume, Apache Spark, HDFS, and Hive) necessary for implementing the IBDMA architecture are included in the virtual machine. On the VM, the other software applications (Pycharm IDE, Elasticsearch, and Kibana) were installed. The IoT data is pushed to two locations by the virtual sensor application. Elasticsearch is the first place the data is sent after it has been indexed and stored, allowing Kibana to visualise it. Several Flume agents ingesting data into HDFS is the second destination. We set up ten Flume agents, each of which serves 100 sensors. PySpark (Python Apache Spark API) [20] is used to analyse the data after it is loaded into HDFS in almost real-time.

We used R to import the publicly accessible smart building open data from the University of Technology Sydney (UTS) into HDFS as the static data. This can be found in Figure 4.12 in Static Data section.

All the other elements of the IBDMA framework are integrated by the *process* component. The processes, as depicted in Figures 4.2 and 4.12, must be implemented before the IBDMA architecture may be implemented. As a result, we go into further depth about the processes shown in Figure 4.4 to illustrate how the IBDMA reference architecture was implemented.

4.2.1.1. Monitoring

Monitoring is the initial step in the IoT-enabled smart building process, as depicted in Figure 4.12. It involves the monitoring of different IoT-enabled smart building ecosystem variables. We select oxygen and gas detection sensors, which track the oxygen and gas levels at UTS Building 11, respectively, for static data.

In order to monitor streaming data, a Python programme that replicates the monitoring of oxygen levels, temperature levels (for fire detection), brightness levels, waste levels, and parking places in the smart building was implemented.

4.2.1.2. Data Sourcing

The second step in the IBDMA implementation is *sourcing*, as shown in Figure 4.12. We take into account both static and real-time streaming data as our data sources, as was previously discussed. Open data from the UTS smart building sensors was retrieved from the web using R and stored in HDFS as the static data source. This information consists of two types of historical sensor data for a particular level or floor of UTS Building 11. Oxygen sensors and gas detection sensors are examples of these sensor types.

A Python virtual sensor software application was used to create real-time streaming data for 1000 virtual IoT smart building sensors. We used big data technologies to complete the data ingestion operation, as described in more detail in the following section, and this data was then ingested and stored in HDFS. Each virtual sensor is identifiable by a different sensor ID. In our research, we classified the first 200 virtual sensors with sensor IDs ranging from 1 to 200 as oxygen sensors, the next 200 sensors with sensor IDs ranging from 201 to 400 as smoke detection sensors, the next 200 sensors with sensor IDs ranging from 401 to 600 as parking space sensors, the next 200 sensors with sensor IDs ranging from 601 to 800 as luminosity sensors, and the next 200 sensors with sensor IDs ranging from 801 to 1000 as garbage detection sensors. PyCharm IDE community edition (*Apache Pig* ; *Pycharm* 2018) is used to create the Python virtual sensor application for data generation. Figure 4.13 depicts a screenshot of the Python virtual sensor application's data generation part.

Figure 4.13 shows that the socket and Elasticsearch modules are imported first, along with the other necessary components. The port and TCP IP are specified. The sensor class is established, and the initialisation of the class's objects—including the sensor ID, sensor value, and sensor placement (a smart building room or floor)—is done. The sensor ID and sensor location are then increased by 1 in a while loop. For each type of sensor, a sensor value is generated at random from a predetermined range of values. For instance, Figure 4.13 displays a snapshot of the oxygen sensors where the percentage oxygen concentration in the air is represented by random values between 8 and 21.

For this research, we configured data generation so that ten sensors would produce data simultaneously, which ten Flume agents would then serve concurrently. Given that the data is generated at one-second intervals, a realistic latency for reporting the ambient conditions of a smart building is 100 seconds between two consecutive data readings from a single sensor. This time period can be shortened by slightly altering the Python programme.

Data is sent to two locations via the virtual sensor application: 1) to Elasticsearch to provide Kibana-based near real-time data visualisation; 2) to Flume agents to enable HDFS-based near real-time data import. We utilised this Python API to save and index data into Elasticsearch. Elasticsearch provides a Python API for this purpose. An Elasticsearch document is defined in the virtual sensor application and contains the sensor ID, sensor value, sensor location, and data production time. As seen in Figure 4.13, this document is subsequently submitted to the "iot" Elasticsearch index.

The data generation code for the first 100 oxygen sensors is shown in Figure 4.13. The numbers for oxygen sensors, which represent the percentage concentration of oxygen in the air, are randomly selected between 8 and 21. The oxygen levels are regarded as normal if the values are more than 14. The same range of numbers is generated at random for smoke detection devices, with values greater than 14 indicating a potential fire scenario. When parking space sensors are used, each sensor emits a high (1) or a low (0) signal indicating whether a specific parking place is occupied or vacant. Random values between the ranges of 8 and 21 are generated for the luminosity sensors, with values over 14 regarded as normal and below 14 indicating luminosity levels below normal. Garbage detection sensors produce a high (1) or low (0) signal to indicate whether a trash can is full or empty.

The sensor ID is reset in the final section of the code once it jumps from 1 to 100. As a result, sensor 1's value is produced, and the cycle is then repeated. Finally, when the programme is finished, the TCP connection and the connection to the Elasticsearch cluster are both terminated.

```
import socket
import time
import random
import datetime

from elasticsearch import Elasticsearch     # import elasticsearch module
es = Elasticsearch()

TCP_IP = '127.0.0.1'        # Define the TCP IP
TCP_PORT = 5005             # Define the TCP Port
BUFFER_SIZE = 500

s = socket.socket(socket.AF_INET, socket.SOCK_STREAM)          # Create an INET streaming socket
s.connect((TCP_IP, TCP_PORT))                                  # Connect to the server on TCP_PORT

# define the Sensor class
class Oxygen:
    id = 1                              # Initialize sensor id
    value = 1.00                        # Initialize sensor value
    def description(self):
        desc_str = "The value of sensor %d is %.2f" % (self.id, self.value)
        return desc_str

# Initialize sensor class
sensor = Oxygen()

# Define and initialize variables
sensor.value = 0
sensor.id = 1

# Start generating random values for the sensor class
while(1):
# sensors 1-200 are Oxygen sensors
# sensors 200- 400 are Smoke detection sensors
# sensors 400-600 are Parking space sensors
# sensors 600- 800 are Luminosity sensors
# sensors 800-1000 are Garbage sensors
      sensor.value = (random.randint(8,21))         # Generate the values for Oxygen in the range from 8 to 21
      t = str(sensor.id).encode()                   # Convert the id to string so it can be sent to the TCP connection
      print (sensor.id)                             # Print the sensor id
      s.send(t)                                     # Send the id on the TCP connection
      s.send(",")                                   # Send a comma to separate id and value
      m = str(sensor.value).encode()                # Convert sensor value to string so it can be sent to TCP connection
      s.send(m)                                     # Send the id on the TCP connection
      s.send(",")                                   # Send a comma to separate value and postcode
      s.send("\n")                                  # Send a newline character so the next set of values be printed on next line
      sensor.id = sensor.id+1                       # Increment the sensor id by 1
      time.sleep(10)                                 # Wait for 1 second to generate next value
      # Create a document to be sent to elasticsearch
      doc = {
          'sensorid': sensor.id,
          'value': sensor.value,
          'timestamp': datetime.datetime.now(),
      }
      res = es.index(index="iot", doc_type='smart_building', body=doc)   # Index the document in elasticsearch
      print(res['created'])                                              # Print if indexed successfully
      if(sensor.id==101):                           # Reset the sensor id and postcode when it reaches 100 sensors
            sensor.id = 1                           # Reset sensor id
s.close()
```

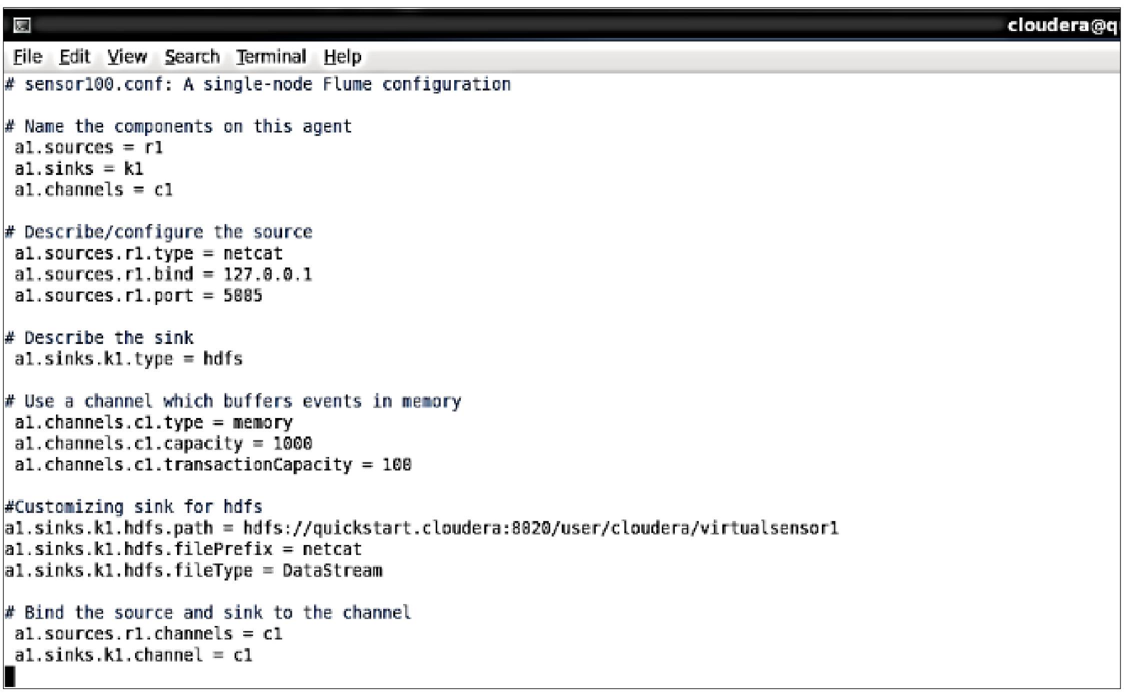
Figure 4.9 Sensor Data Generation Code

4.2.1.3. Data Ingestion

The third process is data ingestion, as depicted in Figure 4.12. For static data, the UTS building data is imported into HDFS after being extracted in the.csv format.

We use Apache Flume to feed IoT sensor data into HDFS for streaming data. Ten Apache Flume agents are deployed for data ingestion. To decrease time latency, boost data throughput, and prevent data loss, these agents are set up to listen to ten different TCP ports as defined in the virtual sensor application. The virtual IoT data generated by the Python virtual sensor programme acts as the source, and HDFS acts as the sink, storing the data as soon as it is received from the virtual sensors; this is achieved using Apache Flume agents.

One of the Flume configuration files out of 10 configuration files is shown in Figure 4.14 by its contents. Source, Sink, and Channel are the three main components of the configuration file. This means that each Flume agent will roll over files after every thirty seconds, finish writing to it, and create a new file in HDFS every thirty seconds as a.tmp file. The 'roll-over interval' for the Flume agent was left at its default setting of thirty seconds. After the thirty seconds, this.tmp file is changed into a permanent one. To have all events written to a single file instead of multiple files, the roll-over interval value should be specified as 0.




```
# sensor100.conf: A single-node Flume configuration

# Name the components on this agent
 a1.sources = r1
 a1.sinks = k1
 a1.channels = c1

# Describe/configure the source
 a1.sources.r1.type = netcat
 a1.sources.r1.bind = 127.0.0.1
 a1.sources.r1.port = 5885

# Describe the sink
 a1.sinks.k1.type = hdfs

# Use a channel which buffers events in memory
 a1.channels.c1.type = memory
 a1.channels.c1.capacity = 1000
 a1.channels.c1.transactionCapacity = 100

#Customizing sink for hdfs
a1.sinks.k1.hdfs.path = hdfs://quickstart.cloudera:8020/user/cloudera/virtualsensor1
a1.sinks.k1.hdfs.filePrefix = netcat
a1.sinks.k1.hdfs.fileType = DataStream

# Bind the source and sink to the channel
 a1.sources.r1.channels = c1
 a1.sinks.k1.channel = c1
```

Figure 4.10 Flume Configuration File

The Flume agent name specified in the configuration file is 'a1', as can be seen in Figure 4.14. A flume configuration file has three essential parts: a source, sink, and channel. Whereas sink ties to the location where the data is to be stored, source binds to the data's incoming source. As the name implies, channel offers a route for data to be sent from the source to the sink. The configuration file's source is a netcat source that binds to the TCP IP address 127.0.0.1 on port 5005. With the URL hdfs:/quickstart.cloudera:8020/user/cloudera/virtualsensor1, the sink is an HDFS sink. The channel is a memory channel having capacity and transaction capacity of 1000 and 100 transactions per second, respectively. The maximum number of events that can be stored in a channel is defined by its capacity. The value of Transactioncapacity defines the maximum number of events per transaction that the channel will ingest the source and push it to a sink.

Web scraping the open data using R or Python is the data ingestion process for batched data. The data was downloaded to our local disc using R, then the extracted data was manually put into HDFS.

4.2.1.4. Data Storage (Big Data Management)

The fourth process is data storage, as depicted in Figure 4.12. We used R to execute web-scraping to download the IoT sensor data from UTS Building 11 for the static data. The received data was first transmitted to our local disc as a single text file, then after some data transformation, it was manually uploaded to HDFS at a designated folder. The temperature, luminosity, humidity, and oxygen sensors data contributed to static data as well. The sensor IDs, the value produced by the sensors, and the timestamp are all included in the data; separated by commas.

We store data for the streaming data in both Elasticsearch and HDFS. Elasticsearch data storage is necessary to make Kibana real-time visualisation possible. The alternative storage destination for Apache Spark's near real-time analysis is HDFS.

Elasticsearch is a distributed search engine that is scalable and has robust APIs to enable lightning-fast data searches for enterprise-grade data discovery applications. By tagging each document in the cluster's index with a searchable reference, Elasticsearch saves data as indexed documents via APIs. After that, the documents can be obtained by using the Elasticsearch API. We used the Kopf plugin, which is built on JavaScript + AngularJS + jQuery + Twitter bootstrap and provides an intuitive web-based administration tool to manage the Elasticsearch cluster, to administer Elasticsearch and the data stored in Elasticsearch.

The data from the IoT sensor application is sent to Elasticsearch, where it is indexed and saved as documents. Kibana makes it possible to visualise data in near real-time. As was mentioned in the section previously, 10 Flume agents are used by the IoT sensor application to deliver data to HDFS. For storing files into the HDFS, Flume uses a default naming scheme that comprises the timestamp. As was mentioned in the earlier section, all ten Apache Flume agents had a thirty-second rollover time. Every thirty seconds, a new .tmp file was produced in the designated HDFS location for each Flume agent. Once the rollover period has elapsed, the.tmp file is automatically changed into a permanent file on HDFS.

The sensor ID, sensor value, and sensor location are all included in the data produced by the Python IoT sensor application. The data produced by the Python IoT sensor programme is stored in Elasticsearch and HDFS, respectively.

4.2.1.5. Data Analysis (Big Data Analytics)

In the IBDMA architecture shown in Figure 4.12, data analysis is the fifth process. PySpark — Spark Python API (Application Programming Interface) was employed in this research to analyse IoT sensor data in near real-time. For the streaming data, we create an algorithm in PySpark that keeps track of the incoming data as it is being ingested into HDFS in almost real-time. The created algorithm determines whether or not it is necessary to activate any controls in the smart building based on the data provided by the sensors. The PySpark algorithm recognises it and emits a text message stating that the fire alarm near the sensors has been activated, for example, if a certain smoke detection sensor gives out a value that is too high suggesting a potential fire scenario. When dealing with batched data, the PySpark algorithm scans the entire file and prints a descriptive message on the terminal for each sensor, indicating whether all of the values received were within the acceptable range or not. This helps with the detection of systemic problems and the enhancement of inhabitants' security and comfort. The data analysis PySpark code is available in the GitHub repository (Bashir 2017).

4.2.1.6. Data Visualization

The sixth process in the IBDMA architecture is data visualisation, as shown in Figure 4.12. The data visualisation process comes after the data analysis process. We evaluated both near real-time and batch visualisations for real-time streaming data. We employ Kibana, which connects with Elasticsearch and generates near real-time visualisations of the received data, for near real-time viewing. By using these visualisations instead of sifting through large data sets or applying algorithms to the data, smart building issues can be identified in almost real-time. In order to simplify the visualisation process for the batched visualisation, we combined several HDFS files created throughout the data storage process into a single file. To generate the batch visualisations, we used a variety of tools. By putting the data into a Hive database, we first decided to use the built-in visualisation tool provided by Cloudera. Cloudera's Hue tool was used to construct Hive tables, which were then imported with data from HDFS. The Hue visualisation is displayed in Figure 4.15. We are only able to visualise up to 1000 rows of data while using Hive for visualisation within Cloudera; this is due to the limitation in Hive.

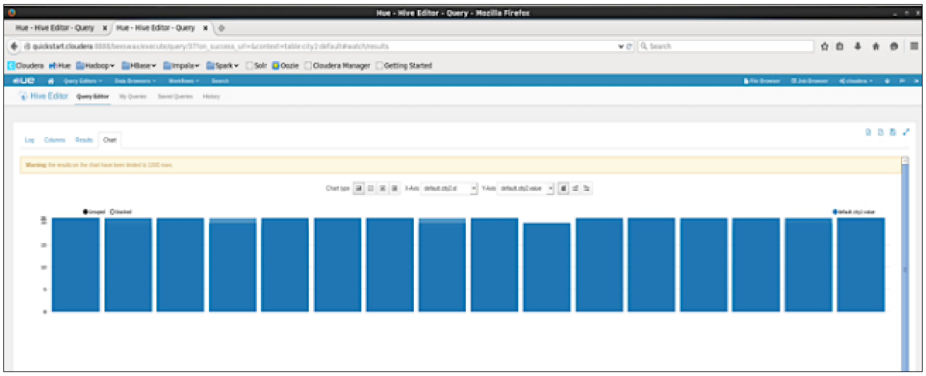


Figure 4.11 Data Visualization in Cloudera using Hive Tables

Tableau is the second tool we used for data visualisation. With the use of a connector, Tableau is able to connect to the HDFS data. Snapshots of Tableau's data visualisation are shown in Figure 4.16. The visualisation in Tableau is much more versatile than Hive data visualisation since the results can be better customised to provide representations that are more insightful. In a top-to-bottom view, Figure 4.16 displays the minimum and maximum values produced by a specific sensor, with the minimum values on top in red and the maximum values on the bottom in blue.

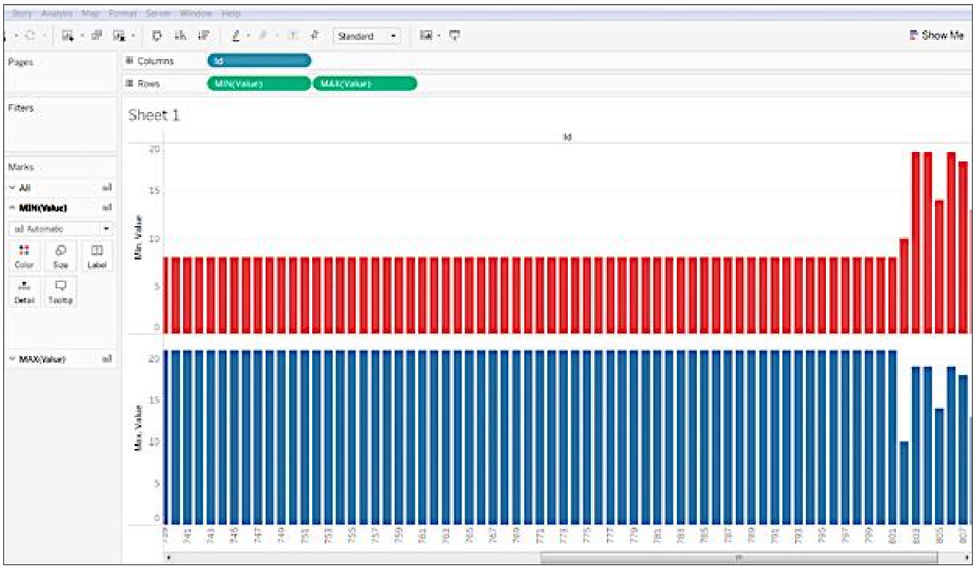


Figure 4.12 Data Visualization in Tableau (Min and Max Values displayed top and bottom)

MS Power BI was the third product we used for batch data visualisation. Because it integrates effectively with HDFS, has a user-friendly interface, and allows us to develop R programming scripts, we discovered MS Power BI to be the best tool for our research. We can create and use predictive models in Power BI by using R scripts, which is an advantage. Figure 4.17 shows the visualisation created in Power BI for open batched data for a temperature sensor during the previous 13 months.

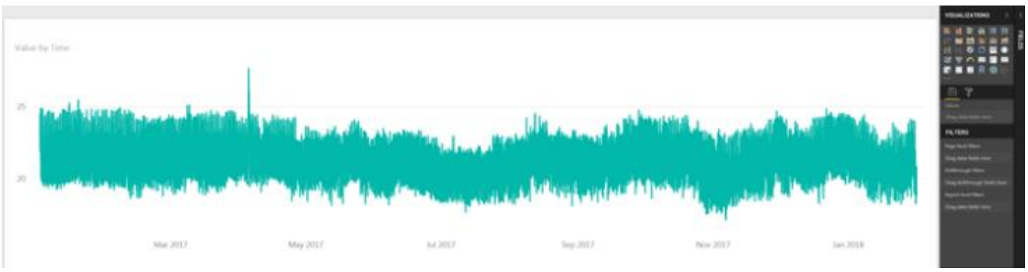


Figure 4.13 Temperature Sensor data visualization in MS Power BI

Every 7 minutes, the sensor generated new data. We created a R script to predict the subsequent 1000 temperature sensor readings, or about the following five days. To perform the prediction, we opted for an ARIMA model. Figure 4.18 displays the predicted outcomes.

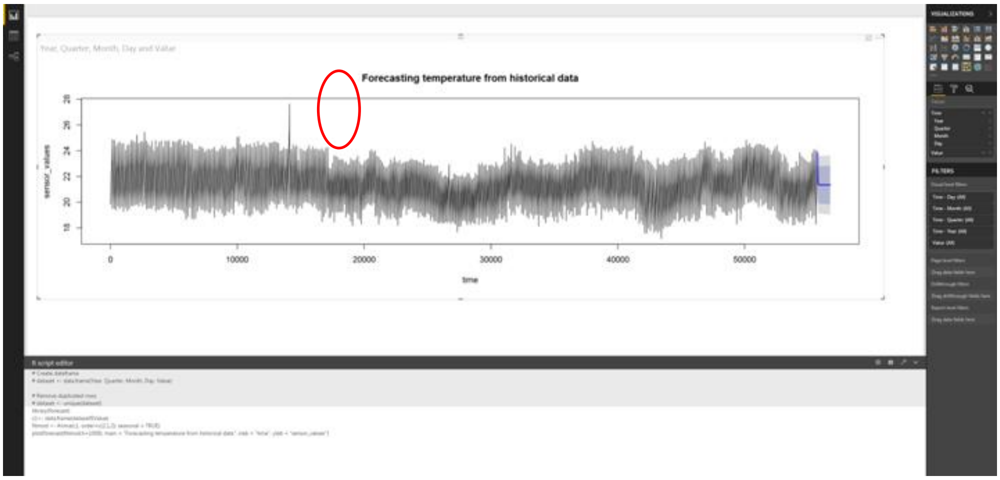


Figure 4.14 Forecasting Sensor data in Power BI using R

The discussion of suitable model selection, fine-tuning, and evaluation is outside the purview of this research. Nonetheless, Figure 4.19 shows the code we employed to create a straightforward Arima model and to predict the following 1000 values for the temperature sensor. We set the seasonality to "true" in our ARIMA model because seasonal influences affect temperature sensor results. The modelling can be expanded to detect and forecast different smart building attributes, however this discussion is outside the purview of this work.

```r
# load sensor data
dataset = read.csv('Users/Riz/temperature_sensor.csv',
                   check.names = FALSE, encoding = "UTF-8", blank.lines.skip = FALSE);

# load forecast library
library(forecast)

# convert sensor values to data frame
c1<- data.frame(dataset$Value)

# create an Arima model
fitmod <- Arima(c1, order=c(2,1,2), seasonal = TRUE)

# plot the next 1000 sensor values based on the arima model
plot(forecast(fitmod,h=1000), main = "Forecasting temperature from historical data",
     xlab = "time", ylab = "sensor_values")
```

Figure 4.15 Forecasting sensor values using ARIMA model

When it comes to displaying static or extracted data, Power BI is an exceptional visualisation tool. Nevertheless, Power BI and many other products on the market fall short when it comes to visualising IoT streaming data in close to real-time. Because of this, we decided to use Elasticsearch and Kibana to quickly display streaming data.

Kibana, an open-source data visualisation tool, is used to create near real-time visualisations and dashboards for the documents indexed into the Elasticsearch cluster once the data has been indexed in

Elasticsearch as described in section 6.4. Figure 4.20 displays the minimum and maximum values obtained from this specific IoT sensor along with the near real-time data visualisation obtained in Kibana.

Figure 4.16 Near real-time sensor data visualization in Kibana

4.2.1.7. Decision Making

According to Figure 4.12 of the IBDMA architecture, *decision-making* is the seventh process. Understanding and assessing the outcomes of the data analysis and data visualisation processes are part of the decision-making process. It entails analysing and presenting the outcomes of predictive analytics produced by fusing R predictive modelling scripts with MS PowerBI's simulation of IoT sensor data. Analysis of the PySpark algorithm's results is another step in the decision-making process. The decision-making process enables stakeholders in smart buildings to make better data-driven decisions to raise the smart building's level of efficiency, comfort, and security.

4.2.1.8. Action: Smart Building Control

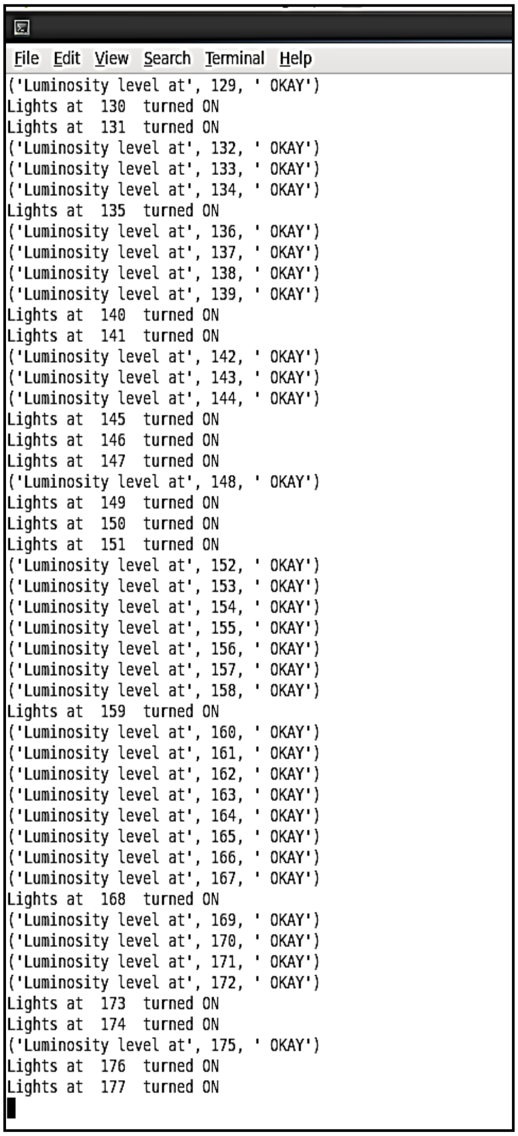
According to Figure 4.12, *action* is the final process in the IBDMA architecture. Once a decision is taken based on the results of the data analysis and data visualisation, the IBDMA manages the smart building facilities autonomously. This is done by using the information generated by the related IoT sensor to activate the appropriate controls of the facilities installed in the smart building. The system outputs the findings as textual output on the Cloudera terminal in order to imitate this behaviour.

By analysing the data gathered from the oxygen sensors installed in the smart building, the oxygen concentration is managed and maintained. The Apache Spark algorithm analyses the incoming data as soon as the data from the oxygen sensor is ingested into HDFS. If the value received by a sensor is below 14, which represents the oxygen concentration level, the Spark algorithm considers it to be a value below the necessary threshold and turns on the corresponding HVACIn order to imitate this for the purposes of our research, we display "HVAC system X turned ON" on the Cloudera terminal, where X stands for a particular location in the building. The oxygen concentration levels in that specific area will now rise as the HVAC system is left on for a while, and eventually the levels will be within the acceptable range. The HVAC system is then turned off by the system when this occurs. By showing "HVAC system X switched OFF," we imitate this behaviour for the purposes of this study. The suggested IBDMA reference architecture, on the other hand, does not initiate any triggering activities if another oxygen sensor deployed at a different position in the smart building already detects acceptable oxygen concentration levels. The PySpark method displays "Oxygen level at X OKAY" in this case, indicating that the oxygen levels measured by sensor ID X are satisfactory.

We must make sure that the smoke detection sensors are installed to keep an eye on any smoke or fire threats in order to maintain a safe and productive atmosphere for the people. The smoke detection IoT sensors continuously scan their surroundings for potential fire hazards and, in the event of a fire, convey a value that is higher than a set threshold value to indicate a fire in the smart building. When this occurs, the suggested architecture activates the associated fire alarm placement at the same spot where the fire was first identified by the sensor. In order to simulate the behaviour of turning on a fire alarm for the purposes of our research, the Spark algorithm outputs "Fire alarm X switched ON" on the Cloudera terminal, where X is the location of the building where the sensor detected smoke or fire. In contrast, the IBDMA architecture will not generate alerts in typical circumstances where no fire or smoke is detected and will continue operating normally simply printing "No fire at X" on the Cloudera console. This will be carried out by the system for each point in the smart building.

Parking sensors are installed in the smart building to monitor and efficiently maintain the parking spaces. Indicating "Parking X is occupied" when a parking space is full is the IBDMA architecture. The parking space that a car had occupied is indicated by the letter X. The architecture prints up the words "Parking X is empty" as soon as the car pulls out of the parking space. To decide how to enhance the parking places for the inhabitants, this data can be gathered and analysed.

The smart building's installed luminosity sensors measure the levels of illumination in various spaces. The suggested architecture would switch on the lights in that location if the value supplied by the sensor is below a threshold luminance level after the data generated by these sensors has been ingested into HDFS and has been processed by the Spark algorithm. In order to imitate this for our research, we output "Lights at X turned ON" on the Cloudera terminal, where X stands for the sensor ID or location of the smart building where low light levels were identified. The proposed architecture, on the other hand, shows "Luminosity level at X OKAY" to indicate that the luminosity levels at that specific spot are OK and no more action is needed if they are already within the allowed range in a particular room. The suggested architecture can, however, handle situations where the lights need to be turned OFF at a specific area and simulates this by showing "Lights at X turned OFF" on the Cloudera terminal.



```
File Edit View Search Terminal Help
('Luminosity level at', 129, ' OKAY')
Lights at  130  turned ON
Lights at  131  turned ON
('Luminosity level at', 132, ' OKAY')
('Luminosity level at', 133, ' OKAY')
('Luminosity level at', 134, ' OKAY')
Lights at  135  turned ON
('Luminosity level at', 136, ' OKAY')
('Luminosity level at', 137, ' OKAY')
('Luminosity level at', 138, ' OKAY')
('Luminosity level at', 139, ' OKAY')
Lights at  140  turned ON
Lights at  141  turned ON
('Luminosity level at', 142, ' OKAY')
('Luminosity level at', 143, ' OKAY')
('Luminosity level at', 144, ' OKAY')
Lights at  145  turned ON
Lights at  146  turned ON
Lights at  147  turned ON
('Luminosity level at', 148, ' OKAY')
Lights at  149  turned ON
Lights at  150  turned ON
Lights at  151  turned ON
('Luminosity level at', 152, ' OKAY')
('Luminosity level at', 153, ' OKAY')
('Luminosity level at', 154, ' OKAY')
('Luminosity level at', 155, ' OKAY')
('Luminosity level at', 156, ' OKAY')
('Luminosity level at', 157, ' OKAY')
('Luminosity level at', 158, ' OKAY')
Lights at  159  turned ON
('Luminosity level at', 160, ' OKAY')
('Luminosity level at', 161, ' OKAY')
('Luminosity level at', 162, ' OKAY')
('Luminosity level at', 163, ' OKAY')
('Luminosity level at', 164, ' OKAY')
('Luminosity level at', 165, ' OKAY')
('Luminosity level at', 166, ' OKAY')
('Luminosity level at', 167, ' OKAY')
Lights at  168  turned ON
('Luminosity level at', 169, ' OKAY')
('Luminosity level at', 170, ' OKAY')
('Luminosity level at', 171, ' OKAY')
('Luminosity level at', 172, ' OKAY')
Lights at  173  turned ON
Lights at  174  turned ON
('Luminosity level at', 175, ' OKAY')
Lights at  176  turned ON
Lights at  177  turned ON
```

Figure 4.17 Smart building Control Messages

Garbage detection sensors are used to monitor the levels of the bins installed in the smart building's rubbish regions in order to properly and efficiently handle the garbage there. The suggested design generates a warning when a garbage can is full by displaying, "Garbage at X is Full," where X is the position of the garbage detection sensor that noticed the can was full. The proposed architecture continuously monitors the garbage levels and prints out "Garbage at X has space" for all the trash cans in the building that have room for more trash. This makes it possible for the smart building's rubbish to be managed more effectively and efficiently. By examining the trend of the data produced by garbage detection sensors, additional decisions about the installation of additional trash cans and garbage collection schedules can also be altered.

Figure 4.21 shows the results of the reference architecture's analysis of the incoming data from the IoT sensors as they are shown on the Cloudera interface. A simplified flow chart of the actuation process is shown in Figure 4.22.

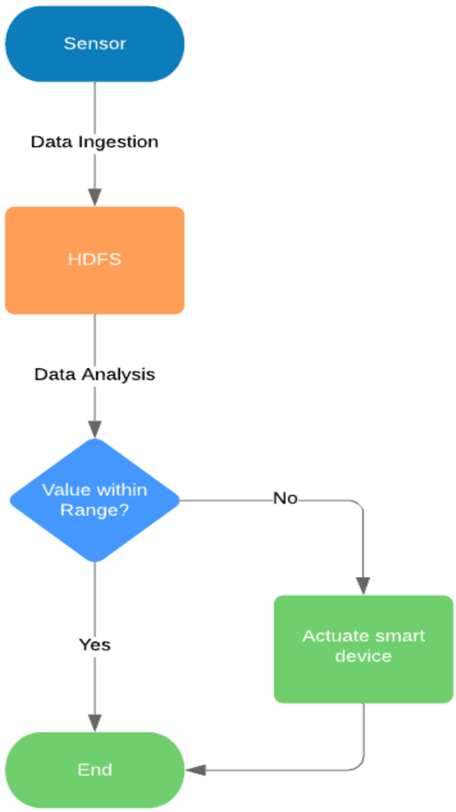


Figure 4.18 Flow chart of Actuation Process

4.3.IBDMA Metamodel

As presented in Figure 4.1, the IBDMA metamodel is the second component of the IBDMA framework. In this section and the next, we discuss the metamodel development and implementation processes. To initiate the metamodel development process, a set of appropriate metamodels and architectures were first chosen. The principles of the IBDMA, as well as the relationships between these principles, are based on the existing literature. We used a six-step metamodeling creation process adopted from (Othman & Beydoun 2010) and (Beydoun et al. 2009) to create the IBDMA metamodel. The six-step process is as follows:

- Step 1: Defining the IBDMA metamodel concepts and relationships from the IBDMA framework reference architecture and the use case
- Step 2: Mapping similar concepts and relationships onto relevant domain metamodels and architectures
- Step 3: Reconciliation of definitions
- Step 4: Designation of concepts
- Step 5: Identification of relationships and resultant IBDMA metamodel
- Step 6: Evaluation of metamodel

4.3.1. Defining metamodel concepts and relationships from the IBDMA framework reference architecture and the use case

In the first step, we defined the concepts and relationships employed in smart environments in general and IoT-enabled smart buildings in particular after reviewing both the IBDMA framework conceptual elements and its reference architecture (Figure 4.1 and Figure 4.2). On considering the framework conceptual elements and the reference architecture development and implement process as presented in sections 4.2 and 4.3, we gathered various concepts and instances of the concepts. Table 4.2 summarises the concepts based on our analysis of the metamodel development process for the IBDMA framework.

Table 4.2IBDMA Concepts

IBDMA Concepts	Concepts Definition
People	Denotes a person with the capability of playing a function in a smart building.
Users/residents	Represents the smart building's occupants/residents. A smart building in a university, for example, will house students, academics, and administrative employees.

Building developers	Denotes the smart building's developers and architects.
Physical Entity	Represents a distinct physical object that occupants or residents are interested in to achieve their objectives. Anything in the environment, from humans to computers; from electrical gadgets to everyday goods, can be considered a physical entity.
Virtual Entity	Denotes the digital representation of the physical entity. Database records and virtual sensors are a few examples.
Digital Artefact	The properties of a virtual entity.
Active Digital Artefact	An operating software application or service that can access other services or resources.
Passive Digital Artefact	Passive software elements are represented by a passive digital artefact. These include database entries which is a digital representation of the physical entity.
Augmented Entity	Represents the concept that allows everyday objects to participate in digital operations, and hence can be considered the "thing" in the Internet of Things.
Facility	A physical facility or building.
Environment	A room or any location inside the smart building.
Resource	Software components that provide data from physical entities or are employed in their actuation.
On-device resource	The resources that are hosted on devices and are linked to the physical entity. These not only include code that can be executed for retrieving, analysing, and storing sensor data, but also the control code for actuators.
Information	The data analysis results obtained after processing IoT data.
Data	The data produced by IoT sensors.

Data Process	Numerous steps in the data analytics process.
Device	A physical object with the ability to generate or act on data. This covers sensors and actuators for the Internet of Things.
Actuators	Devices that can be used to activate controls in a smart building.
Sensors	IoT sensors that monitor and produce data about the smart building environment.
Agent	An entity that comprises the groups of all types of devices i.e. sensors and actuators.

After analysing the IBDMA framework conceptual elements and reference architecture described in sections 4.2 and 4.3, we extracted the metamodel relationships, as shown in Table 4.3.

Table 4.3 IBDMA metamodel relationships

IBDMA relationships
Originates
Processes
Generates
Interfaces with
Installs
Helps maintain
Is composed of
Is monitored by
Has information on/acts on
Is associated with
Interacts with
Contains
Represents
Hosts

In the second step, we ensured the concepts and the relationships we developed could be mapped onto the concepts and relationships from other relevant architectures and metamodels.

4.3.2. Step 2: Mapping similar concepts and relationships onto relevant domain metamodels and architectures

This stage serves as a first check to confirm the IBDMA metamodel is semantically sound and capable of generating relevant concepts and relationships on top of the appropriate metamodels and architectures. For this reason, we combed through the literature and selected the seven most relevant metamodels and architectures.

1- ArchiMate
2- FAML
3- Adaptive Architecture metamodel
4- TOGAF
5- ISO/IEC/IEE 42010
6- IoT reference model (Bassi et al. 2013)
7- BIM (Building Information Model) (Borrmann et al. 2015)

We mapped the IBDMA metamodel concepts developed in the preceding stage to equivalent concepts discovered in the selected architectures and metamodels after thoroughly reviewing and analysing them. Table 4.4 shows the IBDMA metamodel concepts mapped to the concepts present in the aforementioned metamodels and architectures. Our research concludes that some of the IBDMA metamodel concepts are present in the short-listed metamodels and architectures, as shown in Table 4.4.

Similarly, to map the IBDMA metamodel relationships onto the relationships observed in the aforementioned metamodels and architectures, we used the same seven metamodels and architectures. Table 4.5 shows the generated relationship mapping table.

As shown in Tables 4.4 and 4.5, a number of IBDMA metamodel concepts and relationships we defined for the IBDMA metamodel have a one-to-one mapping with the concepts and relationships of some of the shortlisted metamodels. None of the selected metamodels, on the other hand, covers the concepts and relationships that we described in a single unified metamodel. Furthermore, as shown in Table 4.5, there are a few relationships that are not present in any other metamodel. This supports the assertion that the IBDMA metamodel thoroughly covers all concepts and their relationships to handle BDMA issues in smart building systems.

4.3.3. Step 3: Reconciliation of Definitions

This stage reconciles the inconsistencies in the definitions of the concepts. When picking or reconciling a common concept definition, concept definitions chosen in the previous section are taken into account. Concept definitions are derived from a variety of architectures and models and were created by individuals with diverse viewpoints and expertise. If two or more sources employ concept definitions in contradictory ways, we require a procedure to harmonise and fit the definitions into the metamodel. Some architectures and models do not explicitly specify a few of their concepts. In these instances, the reconciliation procedure is absent from the models. For instance, *people* as a concept may be described in a different way in the seven selected metamodels and architectures outlined in Step 2 than in the IBDMA metamodel shown in Table 4.2. However, we discovered that the majority of the concepts were specified in the same way as the IBDMA metamodel (see Table 4.2).

4.3.4. Step 4: Designation of Concepts

The concepts are labelled and placed into metamodeling layers (M2, M1 and M0) (Atkinson & Kuhne 2003) once they have been finalised and reconciled. M2 layer concepts are applicable to all smart settings, including smart farms, smart buildings, smart cities, and so on. The M1 layer concepts are unique to smart buildings, whereas the M0 layer concepts are instances of the concepts in M1.

People, process, technology, information, and facility are the five key conceptual level elements of the IBDMA framework and are represented by five concepts in M2. These concepts apply to a wide range of smart environments, including smart buildings, smart offices, and many others. In the M1 layer, these five M2 layer concepts are split into their instances. M1 represents smart building concepts, while M0 represents the instances of M1 concepts. The naming and organisation of IBDMA metamodel concepts into the metamodel layers is depicted in Figure 4.23.

Table 4.4 IBDMA metamodel concepts mapped with model relationships found in the literature

IBDMA Concepts	**ArchiMate Concepts**	**FAML Concept**	**Adaptive Architecture Meta Model Concept**	**TOGAF Concept**	**ISO/IEC/IEE 42010 Concept**	**IoT reference model (Book: Enabling Things to Talk. Chapter 7 IoT reference model) Springer Open**	**BIM (Springer book)**
People			Human		Stakeholders	User	People
Users/residents	Business role	Role		Actor/role		Human user	Users
Building developers							Developer
Physical Entity						Physical Entity	Entity
Augmented Entity						Augmented Entity	
Digital Artefact	Artifact	Facet				Digital Artefact	Artifacts

Active Digital Artefact						Active Digital Artefact	
Passive Digital Artefact						Passive Digital Artefact	
Virtual Entity						Virtual Entity	
Facility	Facility		Facility				Facility
Environment		Environment	Environment		Environment		Environment
Resource		Resource				Resource	Resource
On-device resource						On-device resource	
Information			Information				Information
Data				Data Entity			Data
Data Process	Technology Process		Technology				Process
Device	Device					Device	Devices
Actuators						Actuators	Actuators
Sensors						Sensors	Sensors
Agent		Agent					

4.3.5. Step 5: Identification of Relationships and resultant IBDMA Metamodel

In this stage, the relationships between the IBDMA metamodel concepts are organised into different metamodel layers. The symbols (→), (➔), and (−◇) are used to signify *association*, *generalization*, and *aggregation* relationships, respectively, as shown in Figure 4.24. As an example, the relationship *Helps Maintain* between the *Information* and *Facility* concepts indicates that information aids in the maintenance of all facility elements. As an example of generalisation, the *Actuators* concept is generalised by *Fire Extinguisher, Fire Alarm, HVAC System, Parking Space Switches,* and *Garbage Detection Switches*. The relation 'hosts' connect *On-Device Resource* and *Device* as an aggregation example. Table 4.6 contains further examples of binary relationships. Hardware is depicted in blue, software in green, animate (humans/animals) in yellow, and concepts that fall into many or none of the categories are depicted in pink. The IBDMA reference design for an IoT-enabled smart building, as stated in the preceding section, is used to determine the relationship between the concepts. Table 4.5 contains the relationships between the concepts defined in Figure 4.24. Table 4.6 represents the relationships between the metamodel concepts. The final metamodel is shown in Figure 4.24 by considering the IBDMA reference architecture as provided in Sections 4.2 and 4.3 and merging the metamodel concepts and relationships between these concepts.

4.3.6. Step 6: Evaluation of Metamodel

In the final step, the metamodel is evaluated. The detailed evaluation results are presented in Chapter 5

Table 4.5 IBDMA metamodel relationships mapped with the metamodel relationships found in literature

IBDMA relationships	ArchiMate	FAML	Adaptive Architecture metamodel	TOGAF	ISO/IEC 42010 relationships	IoT reference model (Book: Enabling Things to Talk. Chapter 7 IoT reference model) Springer Open	BIM (Springer book)
originates		generated from					
processes							
generates							generates
interfaces with							
connects to							
installs							
helps maintain							
is composed of	composition	composition	composed of			composition	composition
is monitored by					has	monitors	
has information on/acts on						has information on/acts on	
is associated with	association	association	associated with		association	is associated with	association
interacts with						interacts with	
contains						contains	
represents						represents	
hosts						hosts	

Table 4.6 IBDMA metamodel relationship between concepts

Concept 1	Relationship	Concept 2
Building Developers	Association – *'Installs'*	Sensors

Building Developers	Association – *'Installs'*	Actuators
Device	Association – *'Generates'*	Data
Device	Association – *'Interfaces with'*	Data Process
Data Process	Association – *'Processes'*	Data
Information	Association – *'Originates'*	Data
Information	Association – *'Helps Maintain'*	Facility
Environment	Association – *'Is Composed Of'*	Facility
Environment	Association – *'Is Monitored By'*	Device
Device	Association – *'Hosts'*	On-Device Resources
Resource	Association – *'Has Information On/Acts On'*	Physical Entity
Facility	Association – *'Contains'*	Physical Entity
Virtual Entity	Association – *'Is Associated With'*	Resource
User/residents	Association – *'Interacts With'*	Physical Entity
Virtual Entity	Association – *'Represents'*	Physical Entity
Physical Entity	Association – *'Is Monitored By'*	Building Developers

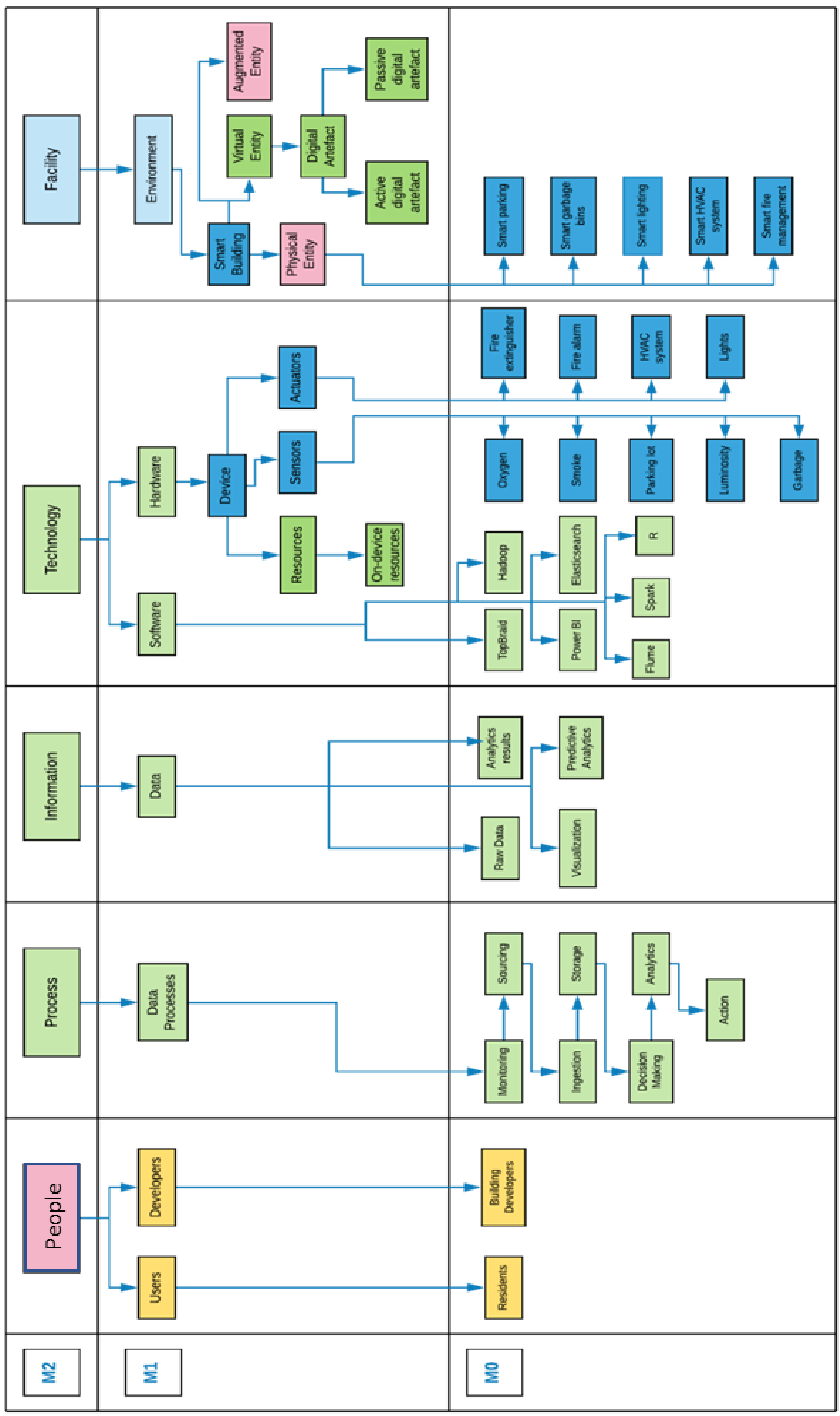


Figure 4.19 Designation of IBDMA concepts into metamodel layers

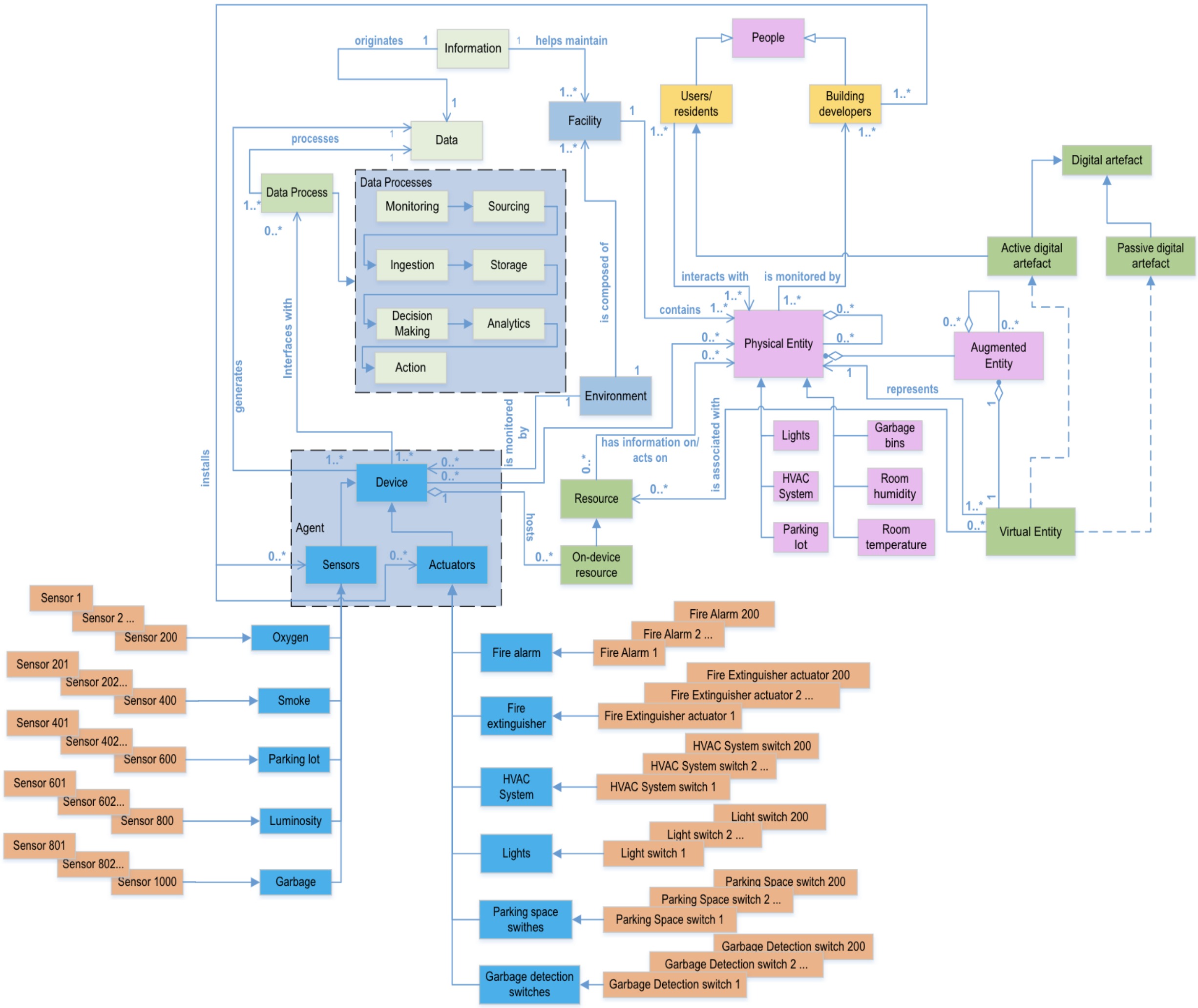


Figure 4.20 IBDMA metamodel

4.4. Summary

The main components of the IBDMA framework were discussed in this chapter. The framework consists of two components i.e., reference architecture and metamodel which are scalable and can be extended for other smart environments. For instance, the same reference architecture and metmodel can be extended for practical use cases in smart cities where temperature monitoring or garbage management are required. The development process as well as the implementation of the metamodel and the reference architecture are presented in detail. In the next chapter, the IBDMA framework is evaluated. Both the reference architecture and the metamodel are evaluated in detail. The evaluation results are detailed in Chapter 6.

Chapter 5 : Evaluation

Chapter 4 presented the IBDMA framework to address the main research question (see Chapter 1). The validity of the IBDMA framework is confirmed in this chapter using an empirical evaluation. The research questions presented in Section 1.2 and the research aims presented in Section 1.3.1 focused on effectively managing and analysing IoT big data in smart buildings, and how to holistically identify all elements involved and relationships between them by developing a reference architecture and metamodel. The reference architecture and the metamodel were presented in Chapter 4. This chapter discusses the evaluation of both the reference architecture and the metamodel. Two types of evaluation methods are used. Firstly, both the reference architecture and the metamodel are evaluated using practical use cases. Secondly, the reference architecture and the metamodel are evaluated by industry experts. The next section presents a high-level overview of the overall evaluation process.

5.1. IBDMA Framework Evaluation Overview

This section provides a high-level overview of the evaluation methodology for the IBDMA framework. Since the IBDMA framework has two components i) reference architecture and ii) metamodel, the evaluation of these two components is done separately using the data generated from the IoT sensors. However, for both the reference architecture and the metamodel, two empirical evaluation methodologies are used: i) practical use case evaluation; and ii) industry expert evaluation. Thus, the evaluation of the comprises the following:

1. Evaluation using practical use cases
    a. Evaluation of reference architecture
    b. Evaluation of metamodel
2. Evaluation by industry expert

The steps involved in IBDMA empirical evaluation are presented in Figure 5.1. Evaluation steps of specific framework components and deployed criteria is presented in Table 5.1.

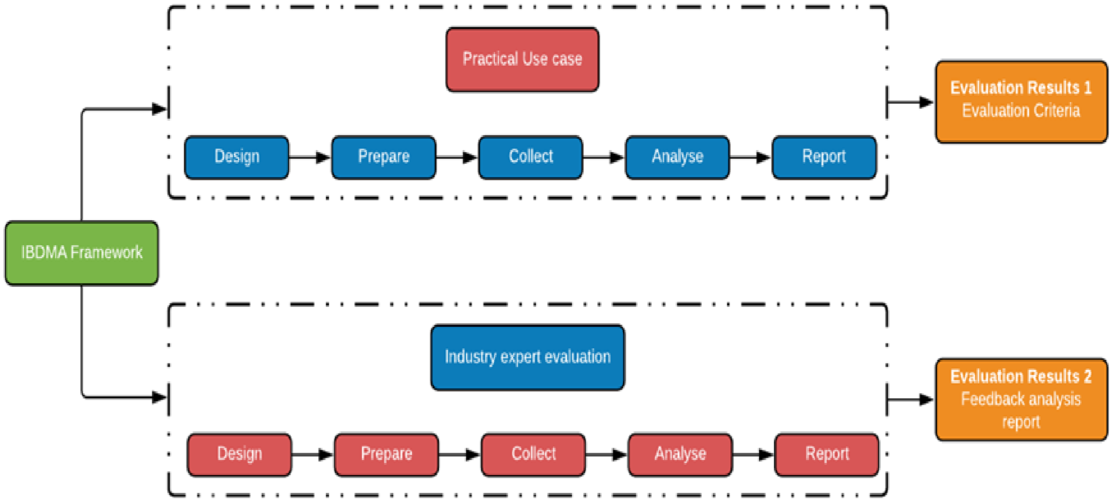


Figure 5.1 IBDMA empirical evaluation overview

To evaluate the IBDMA framework using practical use cases, the IBDMA framework needs to be instantiated so that both the reference architecture and the metamodel can be evaluated using practical use cases. The next section discusses the instantiation process to create instances of the IBDMA framework for various practical use cases used for the framework evaluation.

Table 5.1 IBDMA evaluation method and evaluation criteria

Framework component	Evaluation method	Evaluation Criteria
Reference Architecture	Use case evaluation	EC1 (Efficacy), EC2 (Validity), EC3 (Generality), EC4 (Consistency with People), EC5 (Consistency with Technology), EC6 (Completeness), EC7 (Simplicity), EC8 (Clarity), EC9 (Accuracy), EC10 (Performance), EC11 (Efficiency), EC12 (Robustness), EC13 (Scalability)
Metamodel	Use case evaluation	EC1 (Efficacy), EC2 (Validity), EC3 (Generality), EC4 (Consistency with People), EC5 (Consistency with Technology), EC6 (Completeness), EC7 (Simplicity), EC8 (Clarity), EC9 (Accuracy), EC10 (Performance), EC11 (Efficiency), EC12 (Robustness), EC13 (Scalability)
Reference architecture and metamodel	Industry experts' feedback	EC1 (Efficacy), EC2 (Validity), EC3 (Generality), EC4 (Consistency with People), EC5 (Consistency with

		Technology), EC6 (Completeness), EC7 (Simplicity), EC8 (Clarity), EC9 (Accuracy), EC10 (Performance), EC11 (Efficiency), EC12 (Robustness), EC13 (Scalability)

5.2. IBDMA Instantiation

This section presents the IBDMA framework instantiation process used to create instances for the empirical evaluation using practical use cases. The empirical evaluation considers the implementation of both the reference architecture and the metamodel. The IBDMA framework evaluation is outlined in detail in section 5.3; section 5.3.1 presents the evaluation of the reference architecture utilising use case scenarios, 5.3.2 presents the evaluation of the metamodel against use case scenarios; and 5.3.3 presents the outcomes for the industry expert evaluation.

The IBDMA instantiation process is explained in the following tables:

- Table 5.2 presents the list of tools that were used to create instances of the reference architecture. A list of features and a description for each tool is provided.
- Table 5.3 identifies the step-by-step instructions to create the reference architecture instances using the tool list in Table 5.1. These steps are used to create the use cases for the evaluation of the reference architecture.
- Table 5.4 identifies the list of tools that were used to create instance of the metamodel. A list of features and a description for each tool is provided.
- Table 5.5 identifies the step-by-step instructions to create the metamodel instances using the tools list listed in Table 5.3. These steps are used to create the use cases for the evaluation of the metamodel.

To create instances of the IBDMA reference architecture, the tools required to create instances for each use case need to be first identified. The required tools are listed in Table 5.2.

Table 5.2 IBDMA reference architecture tools description

Tool	Description
Cloudera VM	Cloudera VM (virtual machine) provides a virtual machine for the Cloudera platform to enable BDMA. The

	virtual machine comes built in with most of the big data tools including Flume, HDFS and Hive Metastore.
Apache Flume	Flume is a distributed and reliable service for gathering, aggregating, and transporting huge amounts of log data. Its architecture is simple and adaptable, based on streaming data flows. It is resilient and fault tolerant, with multiple failover and recovery techniques that can be tuned. It employs a straightforward extendable data model that enables online analytic applications.
Elasticsearch	Elasticsearch is a search engine that uses the Lucene library as its foundation. It is a distributed, multitenant full-text search engine with an HTTP web interface and schema-free JSON documents. Elasticsearch is Java-based and is dual-licensed under the source-available Server-Side Public License and the Elastic licence, with the proprietary (source-available) Elastic License covering the rest.
Kibana	Kibana serves an open-source data visualisation dashboard for Elasticsearch. It enables visualisation features for the Elasticsearch indexed documents. Utilising massive amounts of data, users can develop bar, line, and scatter plots, as well as pie charts and maps.
HDFS	The Hadoop Distributed File System (HDFS) is a commodity hardware-based distributed file system. HDFS is fault-tolerant and built to run on low-cost hardware. HDFS is a file system that enables high-throughput access to application data, making it ideal for applications with large data sets. HDFS relaxes a few POSIX rules to offer streaming access to file system data.
Apache Spark	Apache Spark is a large-scale data processing unified analytics engine. Its architectural base is the resilient

	distributed dataset (RDD), which is a fault-tolerant read-only multiset of data objects spread over a cluster of computers.
Microsoft Power BI	Microsoft's Power BI is a business analytics service. Its goal is to give dynamic visualisations and business intelligence capabilities through a user-friendly interface that enables end users to develop their own reports and dashboards.
R/RStudio	RStudio is an integrated development environment (IDE) for R, a statistical computing and programming language. It comes in two formats: RStudio Desktop is a traditional desktop application, whereas RStudio Server is a web-based application that operates on a remote server.
PyCharm	PyCharm is an integrated development environment for computer programming, with a focus on the Python programming language, developed by JetBrains, a Czech firm. It supports multiple plugins and provides the ability to program in various languages.
IoT Application	The IoT application generates IoT sensor data for five different types of sensors (Oxygen, Smoke, Luminosity, Parking and Garbage). The application generates sensor data every 10 seconds for each sensor type. This application is developed in Python.
Github	GitHub Inc. is a software development and version control company. It comprises Git's distributed version control and source code management functions.
Tableau	Tableau is another modern data visualisation and reporting tool which has helped various organisations to be more data driven.

Table 5.2 presents the list of tools required to create instances for each use case for evaluation along with a description and purpose of each tool. Once the tools are identified, the next step is to identify the instructions to deploy and create the instances required for the use case evaluation. For this purpose, Table 5.3 provides the detailed steps in creating the instances.

Table 5.3 Instructions to create the reference architecture instances

Step	Description	Details
1	Downloading and setting up Cloudera VM	Download and setup the Cloudera Virtual Machine. The VM provides a QuickStart platform for the big data analytics and comes built-in with big data tools (including Apache Spark, Flume, Hive, HDFS). The VM requires 8GB of RAM and hence a computing machine with at least 16GB is advisable to run the VM. The VM is available for download at https://studentutsedu-my.sharepoint.com/:u:/g/personal/12532293_student_uts_edu_au/EQ7VwJhc_G1OnEl_H2qYjEABSibEX1S8xRf6lOiny5iVxg?e=aSdGr4 To download the VM, the password is IoTSmartBuildings This VM has all the required tools and packages already setup. The tools include Python, PyCharm, Elasticsearch, Kopf plugin, Kibana.
2	Install VMWare Player	To run the Cloudera VM, download and instal VMWare Workstation Player.
4	R installation	Install R on the Windows machine. R is employed to develop a time series ARIMA model to forecast future values of the IoT sensors based on their historical values.
5	RStudio installation	Install RStudio. RStudio is the IDE for R language. We used RStudio to download batched data from the web and to perform data cleansing on the data.
6	Power BI installation	Install Power BI desktop. Power BI was used to visualise the incoming IoT sensor data. We also used R in the Power BI environment and developed the ARIMA model to make future predictions about the data.
7	Run Cloudera VM	Run Cloudera VM in VMWare player.
8	Open Cloudera terminal	When the Cloudera VM is launched properly, open the Cloudera terminal.
9	Open PyCharm	To open PyCharm in the Cloudera VM.

		Detailed instructions can be found on: https://github.com/c3212218/smartBuildings
10	Run Flume agents	To ingest IoT data generated by the IoT application into Elasticsearch and HDFS, we need the Flume agents up and running. Detailed instructions can be found on: https://github.com/c3212218/smartBuildings
11	Run Elasticsearch	Run Elasticsearch so that the IoT data can be indexed and stored in Elasticsearch. Detailed instructions can be found on: https://github.com/c3212218/smartBuildings
12	Run Kibana	Run Kibana so the incoming IoT can be visualised in near real-time. Detailed instructions can be found on: https://github.com/c3212218/smartBuildings
13	Run Spark code	Run the Spark code which makes Spark ready to analyse incoming data from IoT sensors. Spark code analyses whether the data falls under the acceptable range or not. Detailed instructions can be found on: https://github.com/c3212218/smartBuildings
14	Run IoT application	Run the Python IoT application by running the Python code in PyCharm. This application generates IoT data from 1000 sensors (with 200 of each of five different types of sensors) Detailed instructions can be found on: https://github.com/c3212218/smartBuildings

The detailed instruction for setting the end-to-end pipeline can be found on GitHub: https://github.com/c3212218/smartBuildings

Table 5.2 and 5.3 outline how to set up instances of the architecture for evaluation purposes. Table 5.4 outlines the tools required to create instances of the IBDMA metamodel and a description of each tool for the purpose of evaluation.

Table 5.4 IBDMA metamodel tools description

Tool	**Description**
TopBraid	TopBraid is a leading semantic web modelling tool which is used to implement linked services and metamodels.

	We defined classes and the concepts in TopBraid to develop the IBDMA model. An instance of the IBDMA metamodel for a practical use case is also represented in TopBraid.
Microsoft Visio	MS Visio is a tool to create flowcharts, network diagrams, organisation charts etc. We represented the concepts of the IBDMA metamodel into the metamodeling layers (M2, M1 and M0) using a diagram developed in Microsoft Visio. We also represented the metamodel including the concepts and the relationship between these concepts using Visio. The instantiation of the metamodel for a particular scenario was also presented in Visio.

Table 5.5 outlines the instructions to create instances of the IBDMA metamodel for each use case evaluation.

Table 5.5 Instructions to create the metamodel instances

Step	Description	Details
1	TopBraid Installation	Install and run the TopBraid application. Define the concepts and relationship between the concepts in TopBraid as presented in Section 4.3. Create a specific metamodel instance for a specific use case. More details are presented in Section 5.3.2.
2	MS Visio installation	Install and run Visio for the metamodel implementation and evaluation.

The next section presents the use cases and the evaluation results for each use case for the reference architecture and the metamodel.

5.3. IBDMA Evaluation

This section presents the IBDMA reference architecture and metamodel evaluation details. The evaluation of the reference architecture and metamodel using both the practical use cases and the industry experts is presented. The next two sub-sections present the evaluation of the reference architecture and metamodel using practical use cases. The third sub-section presents the evaluation by the industry experts. The use cases and the industry experts' feedback are evaluated against thirteen evaluation criteria, as presented in Figure 3.8.

5.3.1. Evaluation of reference architecture using practical use cases

The evaluation of the IBDMA Reference Architecture was conducted using practical use cases in a smart building context. The reference architecture was developed in five iterations and each iteration was evaluated against the thirteen evaluation criteria illustrated in Figure 3.8. Details of each of these iterations are presented in Appendix I and the final iteration was presented in Chapter 4 (Section 4.2). Hence, in this section, we present the evaluation of the final version of the reference architecture. The evaluation of the first four iterations is presented in Appendix J.

The process of practical use case evaluation is illustrated in Figure 5.1. The use case evaluation was conducted following specific steps presented in chapter 3 (Section 3.2.5.1):

1. Design

2. Preparation for data collection

3. Collecting data

4. Data analysis

5. Reporting

5.3.1.1. Design

Practical use cases are chosen so that they encompass a variety of practical scenarios that can occur in a smart building in a real-life environment. Each use case is designed to encompass a real-world scenario in a smart building environment. While defining the scope of use cases, special attention is given to ensure the use cases test and evaluate all the important elements of the framework to prove their completeness and effectiveness.

5.3.1.2. Preparation of data collection

Since this research focuses on the data generated by IoT sensors installed in a smart building, we used the data generated by the IoT sensors for the evaluation of the IBDMA reference architecture.

5.3.1.3. Collecting data

The data for the evaluation is collected using the reference architecture instance by developing a pipeline which ingests data from IoT sensors and analyses it for the effective management of a smart building. The details of the reference architecture and its implementation are presented in Chapter 4 (Section 4.2).

5.3.1.4. Data analysis

The data generated by the IoT sensors is analysed to control various aspects and facilities of the smart building depending on the specific instance of the use case.

5.3.1.5. Reporting

The findings of each use case are reported against the 13 evaluation criteria to check if each use case satisfies the criteria or not, as detailed in Figure 3.8. The detailed results of all the five iterations of the reference architecture against the thirteen evaluation criteria (Figure 3.8) are presented in subsequent sections.

- **IBDMA reference architecture use case and evaluation**

For the evaluation of the IBDMA reference architecture, we chose five practical use cases. The details of each use case are presented in Tables 5.6 – 5.10. The title for each use case, its context, the problem statement, the solution to the problem statement, test metrics and its detailed description are presented in the following tables.

Table 5.6 Test case 01 - Detection of oxygen levels

Test Case 01	**Detection of low oxygen level and autonomously activating HVAC system.**

Context	Oxygen levels fall below human comfort levels at a specified location in the building.
Problem	Low oxygen levels in the smart building may remain unnoticed by smart building management which could result in the discomfort of the residents and could potentially prove fatal.
Solution	The oxygen level falls and is detected by the IBDMA reference architecture. Management is notified, and the HVAC is autonomously activated within two minutes.
Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: the terminal displays the message saying the HVAC turned ON when the value of the sensor fell below the threshold.
Description	Oxygen levels in the smart building fall, the oxygen sensor senses that the levels of oxygen at a particular location have fallen below the threshold levels and sends the data to HDFS via Flume. The system detects the low oxygen levels, activates the associated HVAC system installed at that particular location and notifies the smart building management within two minutes. The system autonomously turns the HVAC system off when the oxygen levels are in an acceptable range.
Consequences	The residents enjoy a comfortable environment. The smart building management are notified within two minutes if the levels fall below the threshold levels.

Table 5.7 Test case 02 - Detection of fire

Test Case 02	**Detection of fire and alerting local community.**
Context	A fire erupts at a specified location in the building.
Problem	Fire erupts in the smart building and may remain unnoticed by the smart building management for a longer period, potentially resulting in the loss of infrastructure, investment and lives.
Solution	Fire erupts and is detected by the IBDMA reference architecture. Management and the Fire Brigade are notified within two minutes of the fire eruption and a fire alarm at that location is turned ON.

Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: The terminal displays a message saying Fire Alarm turned ON when the value of the sensor rose above the threshold.
Description	Fire erupts in a smart building; the IoT smoke detector sensor senses fire and sends the data to HDFS via Flume. The system detects fire, activates the associated fire alarm installed at that particular location alerting the local community and notifies the smart building management including the relevant fire brigade team within two minutes. The fire brigade team then acts on the notification and eradicates the fire at the location.
Consequences	The local community is alerted by activating the fire alarm so that they can move to safe areas. The fire brigade and building management are notified within two minutes so that they can respond to the fire to minimize loss of infrastructure, investment and lives.

Table 5.8 Test case 03 - Detection of luminosity

Test Case 03	**Detection of low luminosity level and autonomously activating smart lights.**
Context	Luminosity levels fall below human luminous comfort levels at a specified location in the smart building.
Problem	Low luminous levels in the smart building may remain unnoticed by the smart building management that may not only result in the discomfort of the residents but could also be a safety hazard for the residents.
Solution	The luminosity level falls and is detected by the IBDMA architecture. Management is notified and the smart lights are autonomously activated within two minutes.
Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: The terminal displays the message saying the Lights turned ON when the value of the luminosity sensor fell below the threshold.
Description	The luminosity sensor detects that the luminosity levels at a specified location have fallen below the threshold levels and sends the data to HDFS via Flume. The system detects the low luminous levels, activates the associated smart lights installed at that location and notifies the smart building management within two minutes. The system autonomously turns the smart lights off when the luminosity levels are in the acceptable range.
Consequences	The residents enjoy comfortable luminous levels. The smart building management are notified within two minutes if the levels fall below the threshold levels.

Table 5.9 Test case 04 - Detection of parking space usage

Test Case 04	**Detect if parking lot becomes full and alert the residents.**
Context	A parking lot of a smart building is full.
Problem	A parking lot of a smart building becomes full and may remain unnoticed by the residents for a long period, resulting in inconvenience for the residents.
Solution	A parking lot becomes full and is detected by the IBDMA architecture. The smart building residents are notified within two minutes of the parking lot becoming full and are notified to move to alternate parking areas.
Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: The terminal displays the message saying the parking space is occupied when the value of the sensor is '1'.
Description	A parking lot of a smart building is full and the IoT parking lot sensor senses this and sends the data to HDFS via Flume. The system alerts the local community and notifies the smart building management within two minutes.
Consequences	The residents are alerted so that they can move to alternative parking lots or areas. The building management is notified within two minutes so that they can respond to the growing needs of the residents by planning more parking lots in the area.

Table 5.10 Test case 05 - Detection of garbage

Test Case 05	**Detect if garbage bin becomes full and alert the residents.**
Context	A garbage bin in the smart building becomes full.
Problem	A garbage bin in the smart building becomes full and may remain unnoticed by the smart building garbage management team for a long period, resulting in inconvenience for the residents.
Solution	A garbage bin becomes full and is detected by the IBDMA reference architecture. The smart building management is notified within two minutes of the garbage bin becoming full and the residents are notified to either wait until the bin is emptied or throw their garbage in an alternative location.
Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: The terminal displays the message saying the Garbage Bin is full when the value of the sensor is '1'.

Description	A garbage bin in a smart building becomes full; the IoT garbage detection sensor senses this and sends the data to HDFS via Flume. The system alerts the local community and notifies the smart building management within two minutes. The smart building garbage management team sends the garbage collector to empty the garbage bin at the location.
Consequences	The residents are alerted so that they can throw their garbage in an alternative garbage bin. The smart building garbage management team is notified within two minutes so that they can respond to the growing needs of the residents by sending garbage collectors more often to the specified location.

The five above test cases were evaluated against the thirteen ECs presented in Figure 3.8 using IoT data. When a test case passes the evaluation criteria (EC1 – EC13), a tick or cross is marked against it to represent a pass or fail respectively. Some of these evaluation criteria are easier to evaluate and substantiate than others. For examples, evaluation criterion 'simplicity' (EC7) was evaluated based on the ease of use, while evaluation criterion 'clarity' (EC8) was evaluated based on the clarity of setup for re-producing the results. The outcomes of the evaluation are summarised in Table 5.11.

Table 5.11 Reference architecture evaluation results for evaluating each test case against the thirteen ECs

Test Case	**Evaluation Criteria**	**Pass✓/Fail✗**	**Research Question Addressed**
01			RQ1
	EC1 (Efficacy)	✓	RQ1
	EC2 (Validity)	✓	RQ1
	EC3 (Generality)	✓	RQ1
	EC4 (Consistency with People)	✓	RQ1
	EC5 (Consistency with Technology)	✓	RQ1
	EC6 (Completeness)	✓	RQ1
	EC7 (Simplicity)	✓	RQ1
	EC8 (Clarity)	✓	RQ1
	EC9 (Accuracy)	✓	RQ1
	EC10 (Performance)	✓	RQ1

	EC11 (Efficiency)	✓	RQ1
	EC12 (Robustness)	✓	RQ1
	EC13 (Scalability)	✓	RQ1
02			
	EC1 (Efficacy)	✓	RQ1
	EC2 (Validity)	✓	RQ1
	EC3 (Generality)	✓	RQ1
	EC4 (Consistency with People)	✓	RQ1
	EC5 (Consistency with Technology)	✓	RQ1
	EC6 (Completeness)	✓	RQ1
	EC7 (Simplicity)	✓	RQ1
	EC8 (Clarity)	✓	RQ1
	EC9 (Accuracy)	✓	RQ1
	EC10 (Performance)	✓	RQ1
	EC11 (Efficiency)	✓	RQ1
	EC12 (Robustness)	✓	RQ1
	EC13 (Scalability)	✓	RQ1
03			
	EC1 (Efficacy)	✓	RQ1
	EC2 (Validity)	✓	RQ1
	EC3 (Generality)	✓	RQ1
	EC4 (Consistency with People)	✓	RQ1
	EC5 (Consistency with Technology)	✓	RQ1
	EC6 (Completeness)	✓	RQ1
	EC7 (Simplicity)	✓	RQ1
	EC8 (Clarity)	✓	RQ1

	EC9 (Accuracy)	✓	RQ1
	EC10 (Performance)	✓	RQ1
	EC11 (Efficiency)	✓	RQ1
	EC12 (Robustness)	✓	RQ1
	EC13 (Scalability)	✓	RQ1
04			
	EC1 (Efficacy)	✓	RQ1
	EC2 (Validity)	✓	RQ1
	EC3 (Generality)	✓	RQ1
	EC4 (Consistency with People)	✓	RQ1
	EC5 (Consistency with Technology)	✓	RQ1
	EC6 (Completeness)	✓	RQ1
	EC7 (Simplicity)	✓	RQ1
	EC8 (Clarity)	✓	RQ1
	EC9 (Accuracy)	✓	RQ1
	EC10 (Performance)	✓	RQ1
	EC11 (Efficiency)	✓	RQ1
	EC12 (Robustness)	✓	RQ1
	EC13 (Scalability)	✓	RQ1
05			
	EC1 (Efficacy)	✓	RQ1
	EC2 (Validity)	✓	RQ1
	EC3 (Generality)	✓	RQ1
	EC4 (Consistency with People)	✓	RQ1
	EC5 (Consistency with Technology)	✓	RQ1
	EC6 (Completeness)	✓	RQ1

	EC7 (Simplicity)	✓	RQ1
	EC8 (Clarity)	✓	RQ1
	EC9 (Accuracy)	✓	RQ1
	EC10 (Performance)	✓	RQ1
	EC11 (Efficiency)	✓	RQ1
	EC12 (Robustness)	✓	RQ1
	EC13 (Scalability)	✓	RQ1

The above results clearly demonstrate that the IBDMA reference architecture satisfies all the evaluation criteria for all five use cases. The next section presents the evaluation results for the metamodel.

5.3.2. Evaluation of metamodel using practical use cases

The evaluation of the IBDMA metamodel was performed through three practical use cases. The details of each of these and their corresponding results are discussed in detail in this section. The evaluation process is illustrated in Figure 5.1. The use case evaluation was conducted following the specific steps presented in chapter 3 (Section 3.2.5.1):

1. Design

2. Preparation for data collection

3. Collecting data

4. Data analysis

5. Reporting

5.3.2.1. Design

The practical use cases are chosen so that they encompass various scenarios in a smart building in a real-world environment. While defining the scope of use cases, special attention is given to the fact that the use cases test and evaluate all the important elements of the framework to prove their completeness and effectiveness.

5.3.2.2. Preparation of data collection

Since this research focuses on the data produced by IoT sensors installed in the smart building, we used the data generated by the IoT sensors for the evaluation of the IBDMA reference architecture.

5.3.2.3. Collecting data

The data for the evaluation of the metamodel use cases is collected using the reference architecture instance by developing a pipeline which ingests data from IoT sensors and stores it into HDFS.

5.3.2.4. Data analysis

The data produced by the IoT sensors is analysed to control numerous aspects and facilities of the smart building depending on the specific instance of the use case.

5.3.2.5. Reporting

Based on the use case scenario, a metamodel instance is developed. For each scenario, we evaluate if the metamodel proposed in chapter 4 encompasses all the elements in the BDMA pipeline and is able to encompass all the relationships between various elements of the pipeline.

- **Metamodel evaluation and evaluation scenario 1**

We develop a metamodel instantiation for a specific scenario. The smart building we selected includes numerous types of IoT sensors deployed throughout the building. In scenario 1, however, we chose a single oxygen sensor deployed in the smart building that monitors the oxygen levels in a smart building room. We name this sensor 1 because it is located in Room 1 of the smart building which has been chosen. Cloudera VM is used to construct the BDMA architecture and develop an end-to-end pipeline, as shown in Figure 4.12. This pipeline consumes data from the IoT oxygen sensor and stores it into HDFS, from which the sensor's value is analysed using Spark code and the smart building HVAC system is regulated based on the sensor's value. Once sensor 1 gives a value that is below the human comfort threshold, the Spark code outputs "HVAC System 1 switched ON," demonstrating that the HVAC system that services room 1 where sensor 1 is installed has been turned on. Figure 5.2 illustrates this.

Using this scenario, we can now validate our metamodel. As previously stated, sensor 1 produces oxygen level data; as a device, this sensor generates data and interfaces with the data process. The sensor's "data is processed and relevant information is produced that originates' from the data. This beneficial information helps maintain the facility that contains physical entities. If the data process detects that the value obtained by the oxygen sensor is below the threshold in this situation, it activates the HVAC system to ensure that the oxygen level in room 1 is within the permissible range for the residents of the smart building facility.

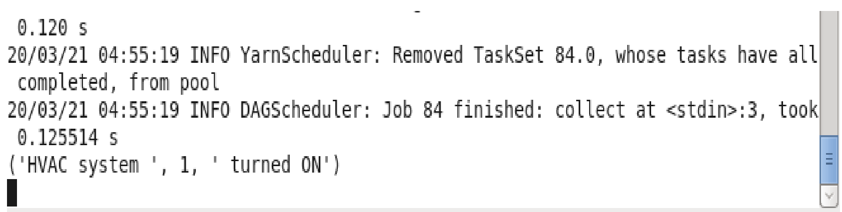


Figure 5.2 Output of evaluation scenario 1

Figure 5.3 represents the resulting metamodel for this particular situation. The IBDMA metamodel clearly contains and consolidates all the elements essential for verifying this particular scenario, as can be seen in this example scenario, and hence proves the applicability and validity of the metamodel. The metamodel use case when evaluated against the thirteen evaluation criteria satisfies all the criteria. The evaluation results are presented in Table 5.12.

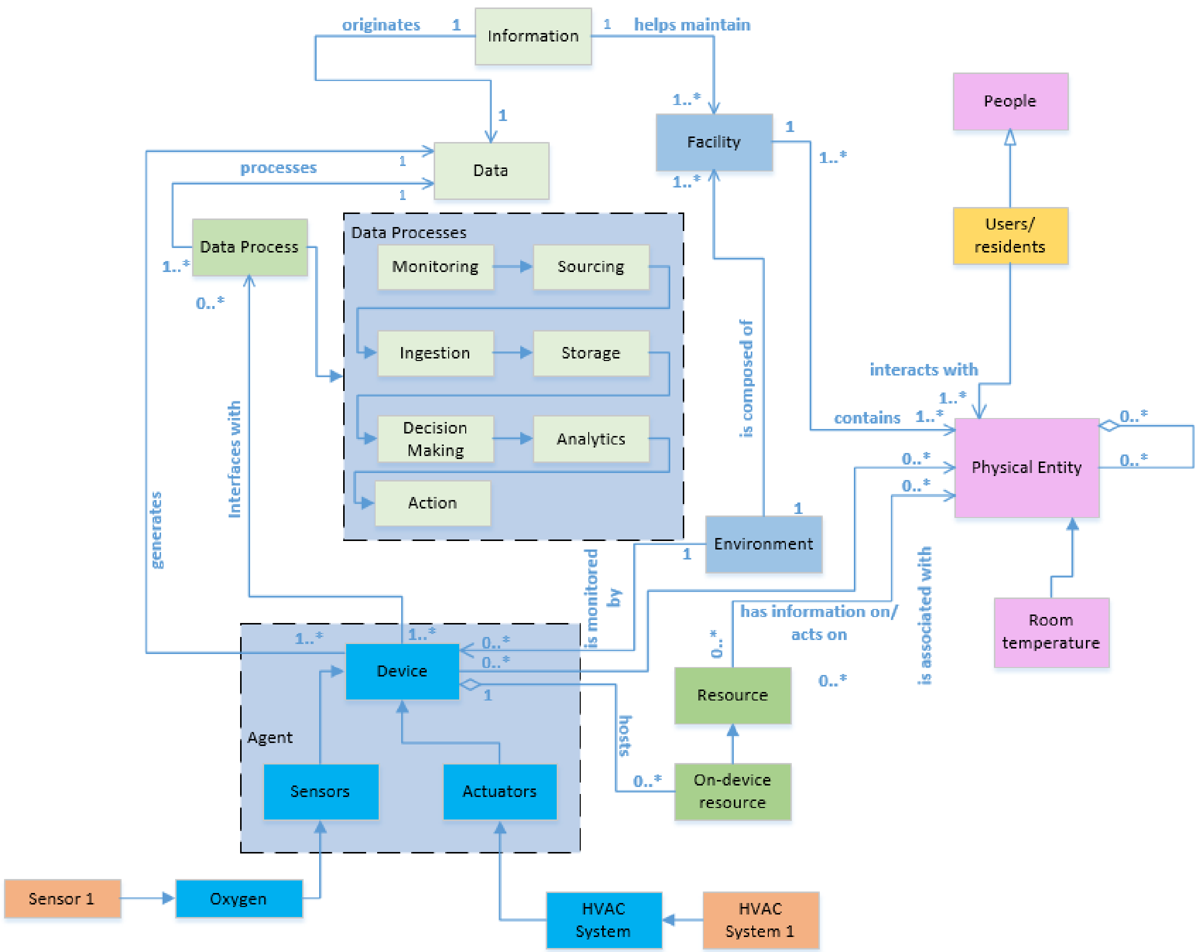


Figure 5.3 Resultant IBDMA metamodel for scenario 1

- **Metamodel evaluation and evaluation scenario 2**

In this section, we examine the second use case to evaluate how well the IBDMA metamodel performs. For this scenario, we choose a smart building from an Australian university as our smart building example. This building has 12 levels, and a variety of sensors are deployed throughout the building. Waspmote sensors are one of many sensor types deployed in the building. Oxygen sensors, carbon dioxide sensors, luminosity sensors, temperature sensors, humidity sensors, and other Waspmote sensors are installed in the building. We chose a luminosity sensor situated on floor 6 of the University's smart building for this scenario. This sensor outputs a binary value, with 1 indicating an acceptable brightness level and 0 indicating a low luminosity level in the sensor's installation room. We call this sensor 601 to keep things simple. Once sensor 601 returns a '1', the lights in this smart building room turn off. When the sensor produces a '0' result, the light in the room turns on.

This is where we create the metamodel for this use case. As previously stated, sensor 601 produces data about luminance levels; as a "Device," this sensor 'generates "data" and 'interfaces with' the "data process." The sensor's "data" is processed, and relevant "information" is produced that 'originates' from the "data." This beneficial "information" 'assists' in the "maintenance" of the "facility" that 'contains' "physical entities”. If the "data process" determines that the luminosity sensor's value is too low in this situation, the "light" is turned on to ensure that the luminance level in room 1 is within the allowed range for the "residents" of the smart building "facility". Figure 5.4 depicts the resulting metamodel for this particular example. As a result, it is obvious from this instance that the IBDMA metamodel covers and encompasses all the elements needed to validate this use case and hence proves the applicability and validity of the metamodel. The metamodel use case when evaluated against the thirteen evaluation criteria satisfies all the criteria. The evaluation results are presented in Table 5.12.

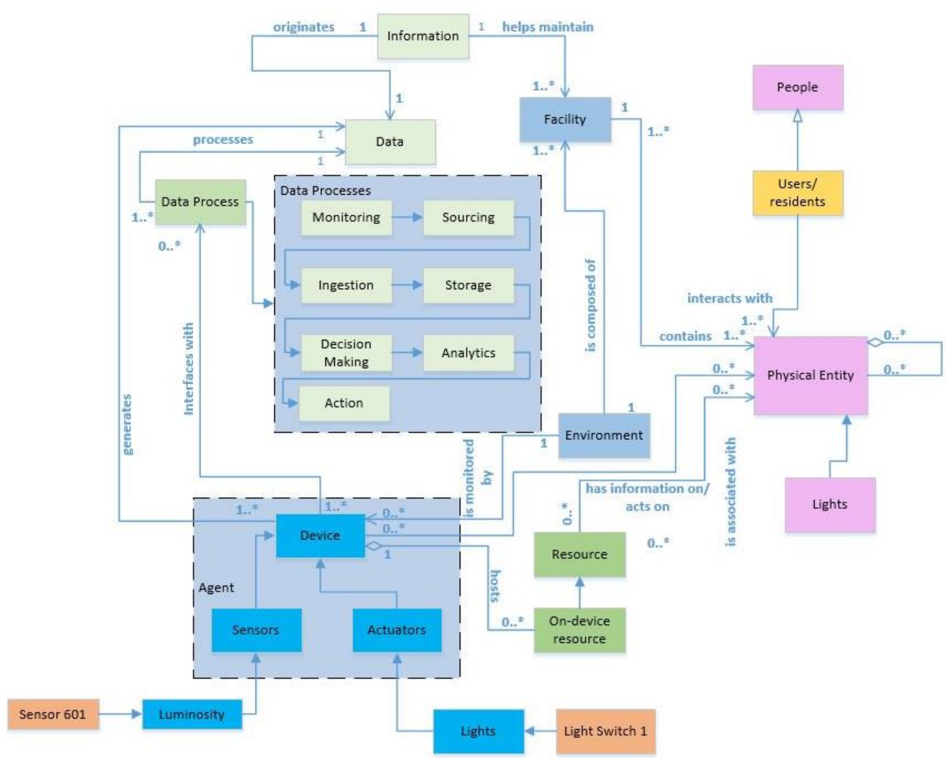


Figure 5.4 Resultant IBDMA metamodel for scenario 2

- **Metamodel evaluation and evaluation scenario 3**

We imported the metamodel into TopBraid EDG (Enterprise Data Governance) (TopQuadrant 2020) to operationalize it. Topbraid is a modular collection of various graphs that represent related information on various aspects of data management and governance. It facilitates the rapid creation of dynamic ontology-driven applications by supporting the whole semantic application lifecycle from development to deployment. To evaluate use case 3 with a smoke detection sensor deployed in the smart building, we use TopBraid metamodel evaluation.

In TopBraid, importing the metamodel entails building classes for the metamodel's concepts and then creating instances of the classes. Figure 5.5 depicts the IBDMA metamodel concepts defined into TopBraid as classes.

Figure 5.5 IBDMA metamodel concepts in TopBraid

The IBDMA metamodel's relationships are then defined in TopBraid. The properties (relationships) between the classes are displayed in the pane on the right side of the TopBraid pane (concepts) as shown in Figure 5.6.

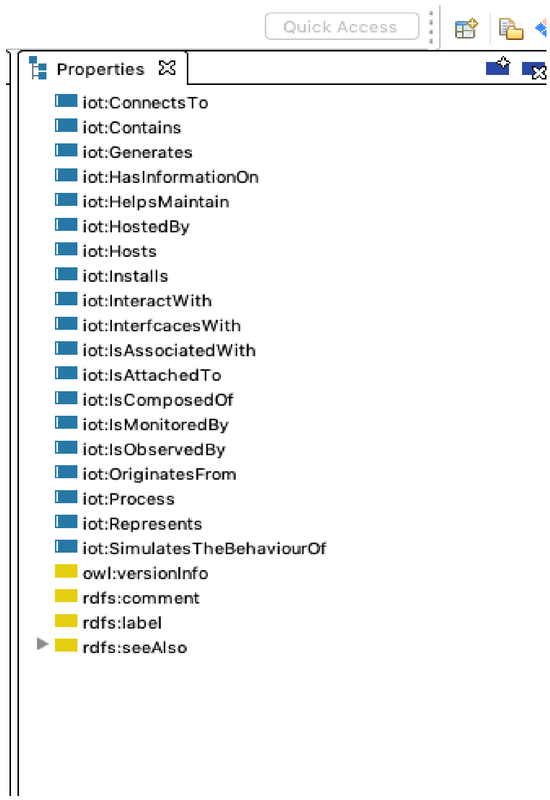


Figure 5.6 IBDMA metamodel relationships

As illustrated in Figure 5.7, the TopBraid bottom pane provides details of the instances of a particular TopBraid class (idea) when selected in TopBraid's class window.

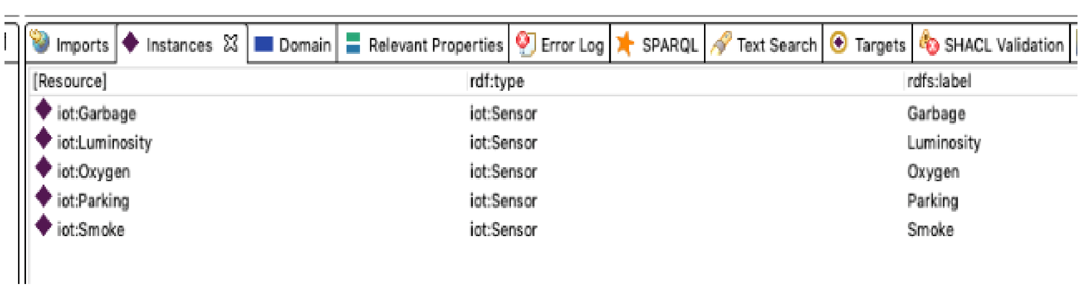


Figure 5.7 Instances of concepts in TopBraid

On importing into TopBraid, the metamodel concepts and the relationships between them are shown in Figure 5.8.

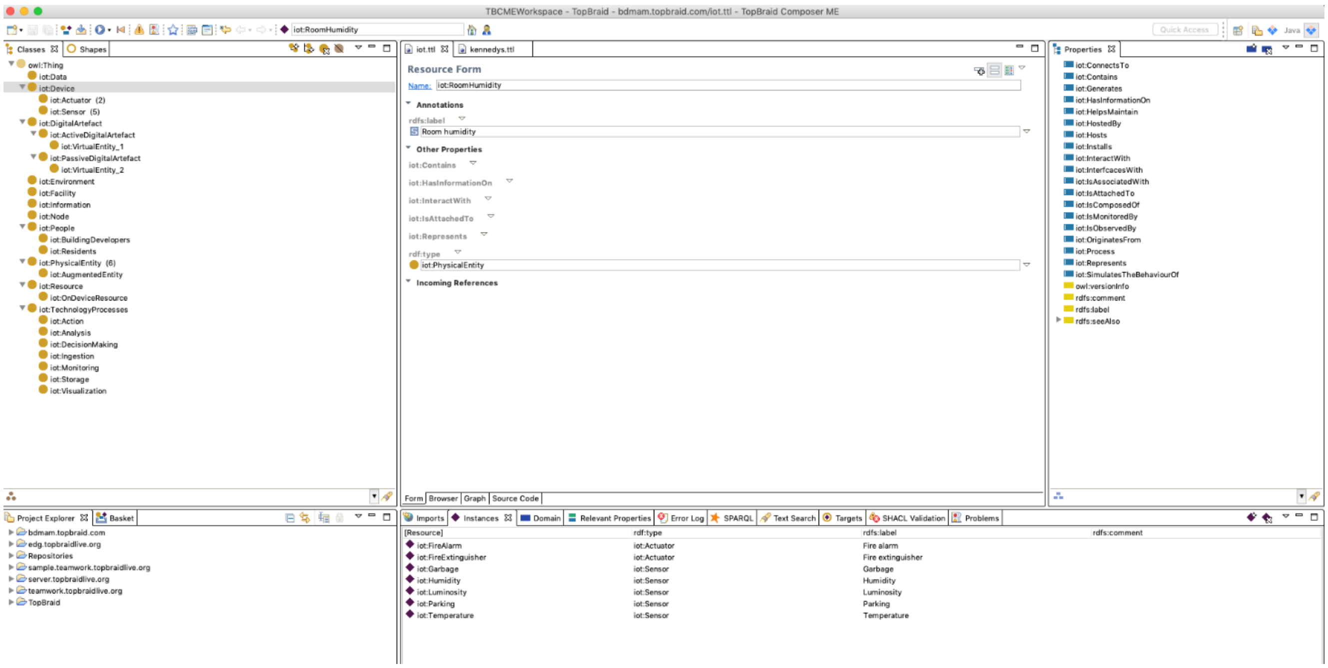


Figure 5.8 IBDMA metamodel imported in TopBraid

As seen in the previous section, we define an instance of the metamodel for smoke detection sensor 201 for the use case 3 evaluation. Figure 5.9 shows the metamodel evaluation scenario's resultant instantiation of the IBDMA metamodel.

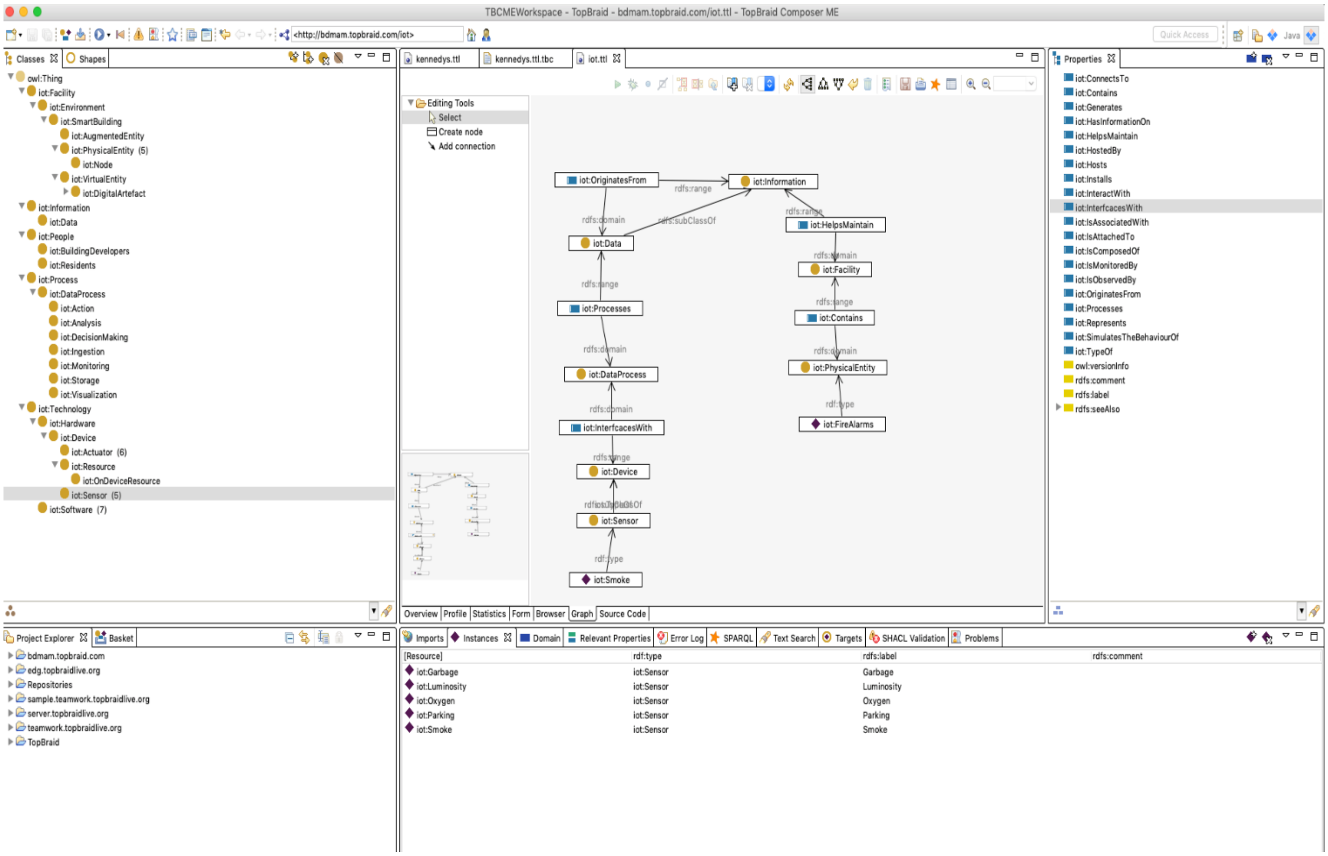


Figure 5.9 Metamodel evaluation in TopBraid – Evaluation and validation scenario 3

For this scenario, Figure 5.10 illustrates a detailed representation of the IBDMA metamodel. The metamodel created in TopBraid contains all the concepts as well as their relationships. This consistent operationalization of the IBDMA metamodel makes it simpler to utilise it, demonstrating that it is valid for the third scenario. The metamodel use case when evaluated against the thirteen evaluation criteria satisfies all the criteria. The evaluation results are presented in Table 5.12.

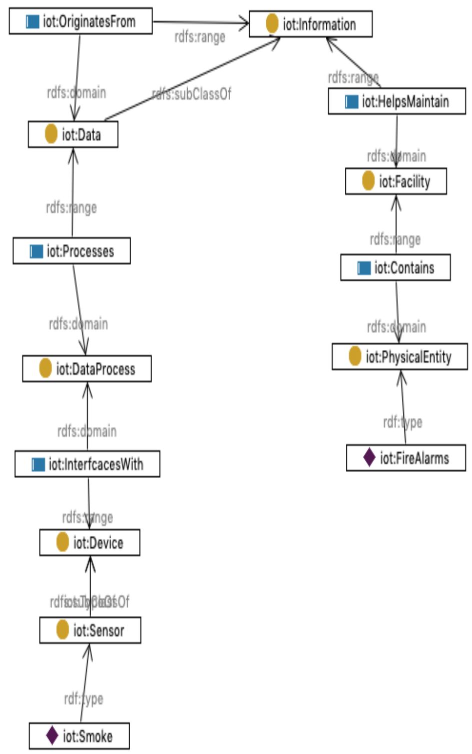


Figure 5.10 Metamodel evaluation scenario 3 version

Table 5.12: Metamodel evaluation results for evaluating each test case against the thirteen ECs

Use Case	**Evaluation Criteria**	**Pass✓/Fail✗**	**Research Question Addressed**
01			
	EC1 (Efficacy)	✓	RQ2
	EC2 (Validity)	✓	RQ2
	EC3 (Generality)	✓	RQ2
	EC4 (Consistency with People)	✓	RQ2
	EC5 (Consistency with Technology)	✓	RQ2
	EC6 (Completeness)	✓	RQ2
	EC7 (Simplicity)	✓	RQ2
	EC8 (Clarity)	✓	RQ2
	EC9 (Accuracy)	✓	RQ2

	EC10 (Performance)	✓	RQ2
	EC11 (Efficiency)	✓	RQ2
	EC12 (Robustness)	✓	RQ2
	EC13 (Scalability)	✓	RQ2
02			
	EC1 (Efficacy)	✓	RQ2
	EC2 (Validity)	✓	RQ2
	EC3 (Generality)	✓	RQ2
	EC4 (Consistency with People)	✓	RQ2
	EC5 (Consistency with Technology)	✓	RQ2
	EC6 (Completeness)	✓	RQ2
	EC7 (Simplicity)	✓	RQ2
	EC8 (Clarity)	✓	RQ2
	EC9 (Accuracy)	✓	RQ2
	EC10 (Performance)	✓	RQ2
	EC11 (Efficiency)	✓	RQ2
	EC12 (Robustness)	✓	RQ2
	EC13 (Scalability)	✓	RQ2
03			
	EC1 (Efficacy)	✓	RQ2
	EC2 (Validity)	✓	RQ2
	EC3 (Generality)	✓	RQ2
	EC4 (Consistency with People)	✓	RQ2
	EC5 (Consistency with Technology)	✓	RQ2
	EC6 (Completeness)	✓	RQ2
	EC7 (Simplicity)	✓	RQ2

	EC8 (Clarity)	✓	RQ2
	EC9 (Accuracy)	✓	RQ2
	EC10 (Performance)	✓	RQ2
	EC11 (Efficiency)	✓	RQ2
	EC12 (Robustness)	✓	RQ2
	EC13 (Scalability)	✓	RQ2

This section presented the use case evaluation for the IBDMA metamodel. The next section presents the IBDMA framework evaluation by industry experts.

5.3.3. Evaluation based on industry expert feedback

This section details the feedback provided by industry experts on the IBDMA framework. The framework was presented to three industry experts who had extensive experience working in IoT or big data domains. The experts were given a presentation and demo about the research conducted and the IBDMA framework. At the end of the presentation and demo, each participant was given a questionnaire to provide their feedback. The details of each participant are provided in Table 5.13:

Table 5.13 Participants details

Participant #	Experience	Country of residence
Participant 1	20 years of experience in the IT industry, with a primary focus on big data, data engineering and solution architecture for IoT enabled smart buildings.	Australia
Participant 2	Over 15 years of experience in data modelling, big data and data architecture. More recently, the participant has implemented data solutions focussing on smart buildings and other IoT applications.	Australia
Participant 3	Over 12 years of experience in the telecom industry with a primary focus on developing business strategy for IoT and 5G in particular for smart buildings (shopping malls, research campuses).	Australia

Before inviting the participants to participate, ethics approval was requested from the UTS ethics committee. The ethics approval is provided in Appendix B. After the ethics approval was granted, an invitation to participate letter was sent to the participants. The invitation letter is provided in Appendix D. The first three experts who accepted the invitation were selected for this evaluation. The selected participants were given the participant information sheet and consent form (Appendix C). On the day of the evaluation, the participants were given a presentation on the IBDMA framework. After the presentation, a demo was given to the participants. Any questions raised by the participants were answered. At the conclusion of the session, a questionnaire comprising five questions was presented to all the participants to provide written feedback regarding the IBDMA framework. The sample questionnaire is presented in Appendix G.

The process of this expert evaluation is illustrated in Figure 5.1. The industry expert evaluation was conducted following the steps outlined in chapter 3 (DSR Evaluation step):

1. Design

2. Preparation for data collection

3. Collecting data

4. Data analysis

5. Reporting

5.3.3.1. Design

The participants were contacted through LinkedIn by searching for people based in Australia who have been working in the big data or IoT domains, who had extensive experience in either of these two domains and who were at the manager level or above. The participants were requested to participate in the research by providing their feedback. The participants who agreed were given the Participant Information Sheet and Consent Form (Appendix C).

5.3.3.2. Preparation for data collection

The participant information sheet gave the participants preliminary details including a brief about the research work being conducted and what their participation would involve. The participants were advised that their participation was voluntary and that they could withdraw at any stage if they so desired.

5.3.3.3. Collecting data

On the day of the evaluation, the participants were given a presentation on the IBDMA framework including the reference architecture and metamodel. At the conclusion of the presentation, a demo was presented to the participants to demonstrate the applicability of the IBDMA reference architecture and metamodel in a smart building context. All questions raised by the participants were answered during the presentation and demo. At the end of the session, a questionnaire with five questions was presented to all the participants for them to provide written feedback regarding the IBDMA framework. The sample questionnaire is presented in Appendix G. The sample questionnaire had five questions to which the participants were asked to provide their written responses. The five questions were:

1. Please provide comments about the IBDMA Reference Architecture. Does it fulfil its intended purpose?
2. Please provide comments about the IBDMA Metamodel. Does it fulfil its intended purpose?
3. Please provide comments on the tools used to implement the IBDMA framework.
4. Please provide feedback on the setup and implementation process of the IBDMA framework.
5. Please provide your overall feedback on the research (including any areas of improvement).

The responses from each participant are provided as follows:

- **Response from participant 1**

1- Please provide comments about the IBDMA Reference Architecture. Does it fulfil its intended purpose?

The reference architecture happens to address practical challenges in the area of big data analytics and management and covers major challenges of the data pipeline from data ingestion to analytics and decision making in an effective and efficient manner. The IBDMA architecture seems to have the ability to be scaled up for buildings of any size such as shopping malls, industrial buildings etc., with slight modifications and can be deployed in multiple smart city application scenarios. It was really interesting to know how the final version of the reference architecture was developed and how it evolved through various iterations following the design science research methodology.

Overall, the reference architecture meets its goal to provide effective management and analysis of data generated from IoT devices.

2- Please provide comments about the IBDMA Metamodel. Does it fulfil its intended purpose?

The IBDMA metamodel provides a comprehensive metamodel which encompasses all major elements and the relationship between these elements. The metamodel fulfils its intended purpose and takes into account essential elements that could be used in a smart building environment. In terms of its usefulness, I believe that the metamodel can be utilised in several practical use cases in industry, including smart buildings and other smart environments. The six-step process of developing and evaluating the metamodel explained by the researcher is quite extensive and ensures broader coverage to address the underlying research question. I was particularly interested in how various concepts and relationships of the metamodel compared to the other relevant architectures and models e.g. TOGAF. I also liked the idea of using a professional tool TopBraid for the implementation and evaluation of the metamodel.

3- Please provide comments on the tools used to implement the IBDMA framework.

I believe that the tools used for the implementation of the IBDMA framework are the latest modern-day tools and are relevant to what is used in industry. I understand that for the PhD project, the availability of enterprise-level tools is difficult. But the researchers have used an open-source version of the enterprise-level tools from leading vendors in the market. For example, the use of Elasticsearch and Kibana, the use of Spark, Flume and Hadoop shows that *the researcher is well aware of the tools available in the market as well as the current leading market tools.*

4- Please provide feedback on the setup and implementation process of the IBDMA framework.

I found the setup and implementation of the reference architecture and metamodel to be satisfactory. The instructions were provided on GitHub regarding the setup and implementation of the framework. The Cloudera virtual machine required for the reference architecture implementation was provided by the researcher and comes pre-installed with all the required tools for the implementation.

5- Please provide your overall feedback on the research (including any areas of improvement).

The IBDMA framework is an integrated framework providing both a reference architecture and a metamodel which is very useful in a real-world scenario. Overall, the research work done is relevant to both industry and academics. The framework design seems to be very well thought out and could easily be used in the IoT and big data spaces with minor modifications for each particular use case. The framework covers the BDMA aspects in a smart building scenario very effectively and efficiently. For future research, the introduction of privacy and security aspects will also add further value to the research.

- **Response from participant 2**

1- Please provide comments about the IBDMA Reference Architecture. Does it fulfil its intended purpose?

IBDMA reference architecture focuses on and addresses big data ingestion, management and analysis challenges in smart buildings and hence fulfils its intended purpose. It is scalable and can be widely adopted both in industry and academia. The researchers understood the challenges at hand and addressed the challenges effectively. The reference architecture is simplistic in nature and encompasses the data ingestion, data storage and data analysis challenges very well.

2- Please provide comments about the IBDMA Metamodel. Does it fulfil its intended purpose?

Yes, the metamodel accomplishes the goal and includes critical features that might be utilised in a smart building. It was interesting to see the implementation of the metamodel done using a *professional tool. In terms of utility, I believe the metamodel may be used in a variety of real-world scenarios in industry. The researcher's six-step approach for building and assessing the metamodel is extremely detailed, ensuring greater coverage to answer the underlying study topic. The researchers have ensured they compared the metamodel with the relevant literature and the models available in the literature.*

3- Please provide comments on the tools used to implement the IBDMA framework.

The implementation of the IBDMA framework was undertaken using the appropriate selection of tools. I found these tools to be cutting-edge and up to date according to the current market. This demonstrates that the researchers have not only extensively reviewed the literature to identify the research challenges, they have also done a good job implementing the solutions to address these challenges. The use of tools like Python, Cloudera, Spark, Tableau, and TopBraid demonstrates the breadth of technical skills the researcher has in addition to research activity.

4- Please provide feedback on the setup and implementation process of the IBDMA framework.

After the presentation given by the researcher, the researcher clearly explained the implementation steps for the reference architecture and the metamodel. I was particularly interested in how they were implemented and after the presentation and demo session with the researcher, both the setup and implementation process seemed appropriate. Written steps and instructions were provided in a clear manner and are available on GitHub. The virtual machine provided by the researcher minimises the setup time required for the implementation.

5- Please provide your overall feedback on the research (including any areas of improvement).

Overall, I found the IBDMA framework very relevant in various IoT use cases. The reference architecture and the metamodel were designed, developed and implemented in a well thought out manner by considering all aspects of the associated challenges in the big data and analytics domains. I believe people working in industry and well as in research will find it quite suitable and useful in various applications.

- **Response from participant 3**

1- Please provide comments about the IBDMA Reference Architecture. Does it fulfil its intended purpose?

The IBDMA reference architecture presents a reference architecture for the BDMA of IoT-enabled smart buildings. The data from various IoT sensors that are installed in smart buildings can be effectively ingested, managed and then can be analysed to autonomously trigger various controls in the smart building. The reference architecture addresses all the challenges for which it was designed, and hence fulfils its intended purpose. The reference architecture is easy to scale and can be used to control many other facilities in an autonomous manner.

2- Please provide comments about the IBDMA Metamodel. Does it fulfil its intended purpose?

Yes, the IBDMA metamodel serves its intended purpose by providing a consolidated metamodel which takes into account all the essential concepts/elements utilized in a smart building big data ecosystem. The metamodel presents a clear relationship between these elements and provides an understanding about how various elements in the smart building interact with each other to provide a safe environment for the residents of the building. The researcher provided a detailed overview of the metamodel development process which I found to be quite interesting and extensive.

3- Please provide comments on the tools used to implement the IBDMA framework.

The researcher has chosen the appropriate tools for the implementation of the IBDMA framework. Tools like Spark, Hadoop, Cloudera, ELK stack and Power BI are undoubtedly the best and leading tools in the market for their relevant usage. This demonstrates that the researcher has extensive knowledge in using these tools.

4- Please provide feedback on the setup and implementation process of the IBDMA framework.

I found the setup and implementation process of the IBDMA framework was reasonably easy and smooth. Following the steps and instructions to create instances of the framework was easy. However, in the future, the setup can be simplified.

5- Please provide your overall feedback on the research (including any areas of improvement).

The IBDMA framework is an integrated framework that provides both a reference architecture and a metamodel, both of which are extremely valuable in a real-world scenario. The researchers have done a good job in identifying the challenges in this domain and to come up *with the IBDMA framework to address the challenges. The setup and implementation details are provided in a clear and concise manner.*

However, as a future research area or as an area of research expansion, the privacy and security aspects of the IoT ecosystem can be incorporated into the framework.

5.3.3.4. Data Analysis

The experts' feedback collected as part of the IBDMA evaluation as discussed in the previous section is summarised and analysed in Table 5.14. The relevant research question that was addressed as part of this feedback evaluation is also listed in Table 5.14. The data was analysed using the cross-examination method between participants' feedback and the thirteen evaluation criteria in Figure 3.8. This analysis aims to connect or relate the hypotheses (evaluation criteria) to the experts' feedback to evaluate the IBDMA framework. The output of the analysis is organised into two columns: participant feedback which is the experts' feedback on the evaluation of IBDMA framework, and IBDMA evaluation criteria categories, which is the relationship between the feedback and the evaluation criteria. This relationship provides a clear view of how the industry experts' feedback relates to the evaluation criteria and whether the IBDMA framework satisfies all thirteen evaluation criteria as per the experts.

Table 5.14 Evaluation criteria satisfied based on participants' feedback

Participants' feedback	**IBDMA evaluation criteria satisfied**	**Research Question Addressed**
Participant 1: "The reference architecture happens to address practical challenges in the area of big data analytics and management and covers major challenges of the data pipeline from data ingestion to analytics and decision making in an effective and efficient manner." Participant 2: "IBDMA reference architecture focuses on and addresses big data ingestion, management and analysis challenges in smart buildings and hence fulfils its intended purpose." Participant 2: "Yes, the metamodel accomplishes the goal and includes critical features that might be utilised in a smart building."	Efficacy (EC1), Validity (EC2)	RQ1 and RQ2

Participant 3: "The IBDMA framework is an integrated framework that provides both a reference architecture and a metamodel, both of which are extremely valuable in a real-world scenario."		
Participant 1: "The IBDMA architecture seems to have the ability to be scaled up for buildings of any size such as shopping malls, industrial buildings etc., with slight modifications and can be deployed in multiple smart city application scenarios." Participant 2: "IBDMA reference architecture focuses on and addresses big data ingestion, management and analysis challenges in smart buildings and hence fulfils its intended purpose. It is scalable and can be widely adopted both in industry and academia." Participant 3: "The reference architecture is easy to scale and can be used to control many other facilities in an autonomous manner."	Scalability (EC13)	RQ1 and RQ2
Participant 1: "Overall, the reference architecture meets its goal to provide effective management and analysis of data generated from IoT devices." Participant 2: "IBDMA metamodel provides a comprehensive metamodel which encompasses all major elements and the relationship between these elements." Participant 3: "Yes, the IBDMA metamodel serves its intended purpose by providing a consolidated metamodel which takes into account all the essential concepts/elements utilized in a smart building big data ecosystem."	Completeness (EC6)	RQ1 and RQ2
Participant 1: "The metamodel fulfils its intended purpose and takes into account essential elements that could be used in a smart building environment. In terms of its usefulness, I believe that the metamodel can be utilised in several practical use cases in industry including smart buildings and other smart environments."	Generality (EC3) Robustness (EC12)	RQ1 and RQ2

Participant 2: "In terms of utility, I believe the metamodel may be used in a variety of real-world scenarios in industry. The researcher's six-step approach for building and assessing the metamodel is extremely detailed, ensuring greater coverage to answer the underlying study topic." Participant 3: "In terms of utility, I believe the metamodel may be used in a variety of real-world scenarios in industry." Participant 3: "The framework design seems to be very well thought out and can be easily used in the IoT and big data spaces with minor modifications for each particular use case."		
Participant 1: "The six-step process of developing and evaluating the metamodel explained by the researcher is quite extensive and ensured broader coverage to address the underlying research question. I was particularly interested in how various concepts and relationships of the metamodel compared to the other relevant architectures and models e.g. TOGAF." Participant 2: "After the presentation given by the researcher, the researcher clearly explained the implementation steps for the reference architecture and the metamodel. I was particularly interested in how they were implemented and after the presentation and demo session with the researcher, both the setup and implementation process seemed appropriate. Written steps and instructions were provided in a clear manner and are available on GitHub." Participant 3: "Yes, the IBDMA metamodel serves its intended purpose by providing a consolidated metamodel which takes into account all the essential concepts/elements utilized in a smart building big data ecosystem."	Simplicity (EC7) Clarity (EC8) Accuracy (EC9)	RQ1 and RQ2
Participant 1: "I believe that the tools used for the implementation of the IBDMA framework are the latest modern-day tools and are relevant to what is used in industry. I understand that for the PhD project, the availability of enterprise-level tools is difficult. But	Consistency with technology (EC5)	RQ1 and RQ2

the researchers have used an open-source version of the enterprise-level tools from leading vendors in the market. For example, the use of Elasticsearch and Kibana, and the use of Spark, Flume and Hadoop shows that the researcher is well aware of the tools available on the market as well as the current market leading tools in the market." Participant 2: "The implementation of the IBDMA framework was undertaken using the appropriate selection of tools. I found these tools to be cutting-edge and up to date according to the current market." Participant 3: "The researcher has chosen the appropriate tools for the implementation of the IBDMA framework. Tools like Spark, Hadoop, Cloudera, ELK stack and Power BI are undoubtedly the best and leading tools in the market for their relevant usage".		
Participant 1: "The IBDMA framework is an integrated framework providing both a reference architecture and a metamodel which is very useful in a real-world scenario. Overall, the research work is relevant both in industry and academia." Participant 2: "Overall, I found the IBDMA framework very relevant in various IoT use cases. The reference architecture and the metamodel were designed, developed and implemented in a well thought manner by considering all aspects of the associated challenges in the big data and analytics domains. I believe people working in industry and well as in research will find it quite suitable and useful in various applications." Participant 3: "The IBDMA framework is an integrated framework that provides both a reference architecture and a metamodel, both of which are extremely valuable in a real-world scenario."	Consistency with People (EC4)	RQ1 and RQ2
Participant 1: "The framework covers BDMA aspects in a smart building scenario very effectively and efficiently."	Performance (EC10) Efficiency (EC11)	RQ1 and RQ2

Participant 2: "Overall, the reference architecture meets its goal to provide effective management and analysis of data generated from IoT devices."		

5.3.3.5. Reporting

The feedback from the industry experts is compiled and reported in this section. The report aims to draw a conclusion from the industry experts' point of view and feedback about the IBDMA to demonstrate its applicability and usability in real-world scenarios. Table 5.15 provides details about the systematic evaluation steps including the expected outcome, the actual outcome and the observations from the industry experts' feedback.

Table 5.15 Industry experts' evaluation report

Expert Evaluation reporting	Description
Evaluation dates	Evaluation was conducted between 1 July 2021 – 31 July 2021.
Participants' details	All three participants reside in Australia and have extensive experience in IoT or big data domains.
Evaluation objective	The objective is to evaluate the IBDMA in view of industry experts so its real-life applicability can be gauged.
Evaluation package	To evaluate the IBDMA framework, a presentation slide pack and demo were designed to demonstrate the solution to the participants. This demo included an IoT software application that was used to generate IoT data.
Evaluation components	The participant evaluated the following IBDMA components: IBDMA reference architecture IBDMA metamodel
Evaluation method	The evaluation comprises obtaining qualitative feedback from industry experts on the IBDMA framework.
Evaluation duration	60 mins (presentation, demo, discussion, questions)
Data type	Qualitative feedback provided by participant
Key activities	The participants verify: The IBDMA framework satisfies all thirteen ECs presented in Figure 3.8 and is applicable in real-world applications. The IBDMA framework enables the autonomous monitoring, management and analytics of IoT data generated in the context of IoT-enabled smart buildings.

Expected outcome	The expected outcome is to the evaluate that the IBDMA framework satisfies the BDMA challenges for IoT-enabled smart buildings and addresses the research gaps highlighted in Chapter 3.
Actual outcome	The actual outcome is determined in the data analysis section (see Table 5.14). The participants provided valuable qualitative feedback and evaluated the IBDMA framework. The feedback satisfied all 13 evaluation criteria presented in Figure 3.8. The participants mentioned that the IBDMA framework adds new knowledge and can be used for the BDMA of IoT data in the context of smart buildings.
Observation and interpretation	The participants evaluated the IBDMA framework against the presentation and the demo presented to them. Based on the feedback from the participants, it is safe to say that the IBDMA framework is not fixed for any particular situation or scenario and can be scaled up and generalized for other scenarios. The participants indicted that the choice of tools used to implement both the reference architecture and the metamodel was excellent, and that these tools serve the purpose well. The observations drawn from the participants' feedback indicate that the IBDMA framework seems to provide a satisfactory and adequate solution to address the BDMA challenges in IoT-enabled smart buildings and can be employed in real-world scenarios.

5.4. Summary

This chapter evaluated the IBDMA framework using an empirical evaluation. The evaluation of the reference architecture and metamodel was performed separately. The IBDMA framework was evaluated using practical use cases and by industry experts. The practical use cases and the industry experts' feedback was considered against thirteen distinct evaluation criteria (Figure 3.8). As mentioned in Chapter 4, the reference architecture was finalised in five iterations, the first four iterations of the IBDMA reference architecture did not satisfy all the thirteen evaluation criteria and hence the final (fifth iteration) was developed which satisfied all thirteen evaluation criteria. The metamodel was similarly evaluated using practical uses cases for smart buildings. The qualitative feedback provided by the industry experts confirmed that it satisfies all the evaluation criteria. Both the empirical evaluation and the industry experts' evaluation resulted in the same outcome, which proved that the results were aligned for both these evaluation methods. Both the evaluation methods satisfied all 13 evaluation criteria that we discussed in Chapter 2. It is evident that the IBDMA framework performs well in real-world scenarios. The evaluation results also highlighted some limitations of the IBDMA framework. In the next chapter, these are discussed in the context of delineating key contributions of this thesis.

Chapter 6 : Discussion and Conclusion

This chapter outlines the research journey that started in February 2016 and concluded with the development of the IBDMA framework. The IBDMA framework is the main output of this research and aims at addressing the big data management and analysis challenges in smart buildings. Secondly, the contributions and publications produced as a result of this research are listed in this chapter. Thirdly, the limitations of this research are discussed based on the feedback from the empirical evaluation. Finally, the chapter concludes with an overall summary of the research project.

6.1. Research Journey and Output

This section outlines the research journey and the output of this research. It is organised into two sections as follows:

- The research journey which started in February 2016
- The main output of this research which is the IBDMA framework.

6.1.1. Research Journey

The research journey is illustrated in Figure 6.1. It began in February 2016 as a Master by Research degree and was later upgraded to a PhD in Spring 2019. The initial step was to perform a systematic literature review on the chosen research topic. The systematic literature review helped to identify the research gap. The actual design and development work on the IBDMA framework started in Autumn 2018. Initially, the focus was on the IBDMA reference architecture and in Autumn 2020, the metamodel design and development work was initiated. During this journey, three research publications were published with the fourth one currently in the submitted state. The CA3 (Candidature Assessment 3) and submission of the thesis is expected by the end of Autumn 2022.

Figure 6.1 Research journey

6.1.2. Research Output

The main research output of the thesis is the IBDMA framework which aims to enable the autonomous monitoring and control of different facilities in smart buildings using IoT data. The framework development was discussed in Chapter 4 in detail and the evaluation results were presented in Chapter 5. The IBDMA framework was developed using the DSR method outlined in Chapter 3.

The framework comprises a reference architecture and a metamodel as shown in Figure 6.2. The details of the framework are provided in Section 4.1.

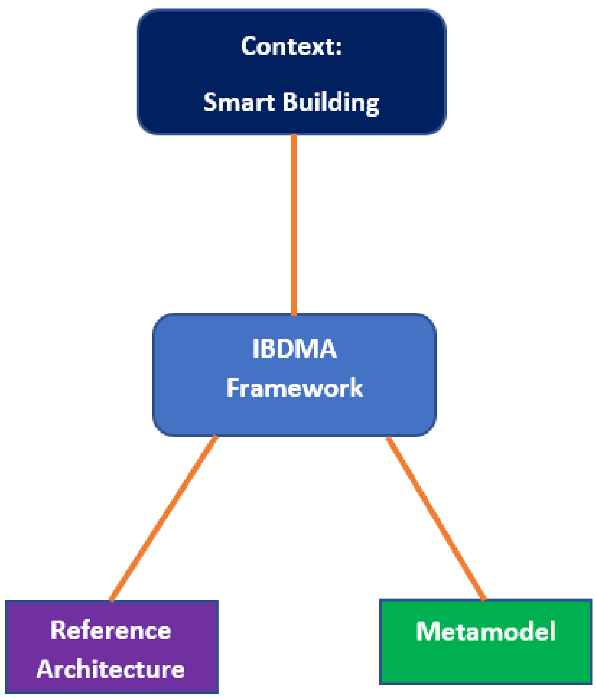


Figure 6.2 IBDMA framework elements

6.1.2.1. Framework elements

The IBDMA framework has five key conceptual level elements: people, process, technology, information and facility, as outlined in Chapter 4. These elements form the foundations of the development of the reference architecture and the metamodel.

Table 6.1 Description of framework elements

IBDMA element	Description
People	The *people* element of IBDMA represents people in the smart building environment. This includes the residents as well as the developers of the smart building. Section 4.1.1 discusses this in detail.
Process	The *process* element of IBDMA represents the data processes employed to address the BDMA challenges in the context of smart buildings. The details are provided in Section 4.1.2.
Technology	The *technology* element of the IBDMA represents the tool stack used to implement the data processes (as mentioned in the *process* element). The details are provided in Section 4.1.3.
Information	The *information* element of IBDMA represents the useful information and knowledge obtained from the management and processing of IoT data in the context of smart buildings. The details are provided in Section 4.1.4.
Facility	The *facility* element represents numerous facilities in the smart building that aim to address the safety, security, and comfort of the smart building residents. The details are provided in Section 4.1.5.

6.1.2.2. Framework design

The IBDMA framework was designed using the DSR approach (Chapter 3). The framework aims at addressing the challenges associated with BDMA in smart buildings. The framework consists of a reference architecture and a metamodel.

- **Reference architecture design**

The development of the reference architecture was an iterative process following the DSR approach, which was explained in Section 4.2. Each of the iterations was evaluated against the thirteen evaluation criteria adopted from the literature, as presented in Figure 3.8. The details of each iteration are presented in Table 6.2:

Table 6.2 Reference architecture design iterations

Iteration	Description
Iteration 1	Building data was extracted from the web portal using R and the data was analysed. Details of iteration 1 are presented in Section 4.2.1.

Iteration 2	A sensor and actuator prototype solution were developed using a senor, microcontroller, and a linear actuator. The data was analysed, and the actuator was activated/deactivated based on the data generated by the sensor. Details of iteration 2 are presented in Section 4.2.2.
Iteration 3	The reference architecture was modified to obtain data from fifteen IoT sensors. Data pipelining tools were introduced into the architecture. BDMA tools were introduced into the architecture. Details of iteration 3 are presented in Section 4.2.3.
Iteration 4	They five conceptual elements of the IBDMA (people, process, technology, information and facility) were introduced and linked to the reference architecture. The reference architecture on the physical layer was expanded to include 200 sensors. The number of facilities that can be controlled in an automated manner was increased. Details of iteration 4 are presented in Section 4.2.4.
Iteration 5	The reference architecture was extended to provide near real visualisation of the IoT data. The number of sensors on the physical layer was increased to 1000 sensors. The ability to predict upcoming data from the sensors was also introduced. Moreover, the reference architecture was expanded to address both the batch and streaming data. Details of iteration 5 are presented in Section 4.2.5.

- **Metamodel design**

The design of the metamodel consisted of a six-step process, as explained in Section 4.4. The concepts and the relationship between them were identified for the development of the metamodel. These concepts were arranged in three metamodeling layers M2, M1 and M0 as presented in Figure 4.23. The details of each metamodel layer and the concepts in each layer are presented in Table 6.3:

Table 6.3 Metamodel layers description

Metamodeling layer	Description
M2	This layer being the high-level layer consisted of IBDMA conceptual elements i.e., people, process, technology, information and facility. This is consistent with the five elements of the reference architecture.
M1	The concepts defined in the M1 layer have a definite relationship between the concepts of the M2 layer. The concepts at the M1 layer are categorizations of instances at the M2 layer. Details of this layer and the concepts included in this layer are in Chapter 4.
M0	This layer contains the actual instances of the concepts defined in the M1 layer. Details of this layer and the concepts included in this layer are in Chapter 4.

6.1.2.3. Framework Evaluation:

The framework evaluation results were presented in Chapter 5. Both components of the framework i.e., the reference architecture and the metamodel were evaluated using practical use cases and by obtaining qualitative feedback from industry experts with extensive experience in big data or IoT domains. The evaluation results were evaluated against thirteen evaluation criteria extracted from the literature. The evaluation results from the use cases as well as from the industry experts satisfied the evaluation criteria and proved the applicability and practicality of the developed framework to address the BDMA challenges in the context of IoT-enabled smart buildings.

6.2. Key Contributions and Publications

Key publications contributed to the development of the IBDMA framework. This section discusses the key contributions of the research which includes both theoretical and practical contributions. The theoretical contributions are listed in Table 6.4.

Table 6.4 Research contribution and publications

Contribution	Reference	Source
IBDMA Framework	The main contribution of this research is the development and evaluation of the IBDMA framework.	Chapter 4 and 5.
Conference	Towards an IoT big data analytics framework: smart buildings systems	IEEE HPCC/SmartCity/DSS (2016)
Conference	IoT-enabled smart buildings: A systematic review	IEEE IntelliSys (2017)
Journal	Big Data Management and Analytics Metamodel for IoT-Enabled Smart Buildings	IEEE Access (2020)
Journal (Accepted for publication)	A reference architecture for IoT-enabled smart buildings	SN Computer Science (2022)

The practical contributions of the research include:

- Reference Architecture (to effectively manage and analyse the IoT data generated in smart buildings).
- Metamodel (to holistically identify all elements and their relationship that are employed in IoT-enabled smart buildings).

Based on the feedback from the industry experts, both the reference architecture and the metamodel can be employed in industry to address big data management and analytics challenges for IoT enabled smart buildings.

6.3. Research Limitations and Future Work

The IBDMA framework was tested and evaluated against practical use cases and was also evaluated by industry experts. The research work has been published in prestigious conferences and journals including peer-reviewed journals. However, despite the key contributions provided in the previous section, the current version of the framework has some limitations as in any research project that will lead the path to the future research work. This section discusses these limitations and provides an opportunity for future research directions.

1. Security and privacy of sensors and data

The framework addresses the IoT data management and analytics issues for smart buildings. However, the research does not encompass the security and privacy aspects of IoT sensors as well as the IoT data generated by the sensors. The research not only covers the data management and analytics challenges, it also addresses the autonomous monitoring and control of facilities within the smart building. These areas are very broad in nature and privacy and security is a separate research topic in itself, hence this research does not consider the security and privacy of IoT sensors, and the data produced by these sensors. It would be interesting to extend the IBDMA framework to include the security and privacy of the sensors and the data generated by these sensors as future research work.

2. Data processing scalability (did not use the cloud)

The IBDMA framework addresses the challenges of data management and analytics associated with smart buildings. If the framework is used to process data from a large number of sensors, such as in the case of a smart city, the data processing will need to be supported by a more scalable architecture. Hence, the next version of the IBDMA framework can potentially include the use of a cloud platform with load balancing and autoscaling options as future research work to see how performance benefits regarding data processing and analysis can be achieved.

3. Edge computing

This research encompasses IoT sensors which do not require high bandwidth or heavy payloads such as video cameras, hence the data processing was conducted at a centralized location. The edge computing aspect was therefore not considered in this research. IBDMA framework can be further extended to consider edge computing as a future area of work.

The aforementioned limitations and future research directions will further enhance the capability of the IBDMA framework.

6.4.Summary

This research was conducted between 2016 and 2022. In Chapter 1, an introduction and background to the research was presented which showed that the concepts of IoT, big data and smart buildings are related to each other. The research questions presented in Section 1.1.4 were developed after a comprehensive review of these concepts. The research problem was then identified, and the aims, objectives, significance, and scope of the research were discussed. The users and applications of this research were presented in Chapter 1.

In Chapter 2, the SLR was conducted to study and review IoT, big data and smart building concepts. The SLR identified the research work done by other researchers to help identify the research gap in the literature which helped in answering the first RQ (Section 1.1.4). The SLR identified that there was no existing coherent framework which consists of both a reference architecture and a metamodel to address the BDMA challenges for smart buildings while incorporating the autonomous monitoring and control of smart buildings. The output of the SLR was published in a conference paper for IEEE IntelliSys (2017), as mentioned in Section 6.2.

This research was conducted using the DSR process based on the guidelines derived from (Gregor & Hevner 2013) and (Peffers et al. 2007). Chapter 3 discusses the research method employed for the development of the framework to address the research gap. The details of the DSR process and the steps involved in the DSR are presented in Figure 3.2. The output of the DSR is the IBDMA framework which is the main contribution of this research.

This research was carried out to develop a new framework called the IBDMA framework for IoT using a constructive and iterative DSR (see Chapter 3). The IBDMA framework aims to resolve the IoT BDMA challenges for smart buildings. As discussed in Chapter 4, the framework has two components: reference architecture and metamodel. Chapter 4 includes the development and implementation details of the reference architecture and the metamodel. The IBDMA framework will aid researchers and industry experts to address and resolve the BDMA challenges in the context of smart buildings.

The IBDMA framework was evaluated using practical use cases and also by industry experts. The detailed evaluation and the outcomes of the evaluation are presented in Chapter 5. For the use case evaluation, practical use cases in the smart building environment were considered for evaluation. Each use case was evaluated against thirteen evaluation criteria extracted from the literature. For the industry expert evaluation, the industry experts provided qualitative feedback. The industry experts were given a presentation and a demo about the research and the IBDMA framework. After the presentation and demo, the industry experts were asked to answer five questions related to the framework. The qualitative responses provided by the industry experts were evaluated against the same evaluation criteria that were used in the use case evaluation. Both the use case evaluation results and the outcome of the industry experts proved that the IBDMA framework was fit for purpose and was beneficial in addressing the management and analytics challenges for smart building IoT data.

Finally, in Chapter 6, the research work is summarised by presenting the research journey that started in autumn 2016 and how the journey continued till Spring 2021. The research output is discussed briefly, and the key contributions and publications are also listed. The framework limitations based on the use case evaluation and the industry experts' feedback are discussed.

6.5. Conclusion

This thesis presented the IBDMA framework to provide both researchers and industry experts working in the big data and IoT domains with a method to address the BDMA challenges. The framework consists of reference architecture and metamodel. The evaluation of both the reference architecture and the metamodel was conducted using practical use-cases as well as by the industry experts. The evaluation results prove that the framework is fit for its intended purpose and provides a comprehensive reference architecture and metamodel that can be applied in any organizational context with slight modifications based on the business requirements however, the basic concepts and architecture will remain the same. The framework can be further extended and scaled to address other smart environments, including smart homes and smart cities, and in areas as outlined in Section 6.3 (Research Limitations and Future Work). The IBDMA framework will be further extended based on future learning, research and experience.

Bibliography

5 Most Important Difference Between Apache Kafka vs Flume 2020, viewed August 10 2018, <https://www.educba.com/apache-kafka-vs-flume/>.

Aberdeen, T. 2013, 'Yin, RK (2009). Case study research: Design and methods . Thousand Oaks, CA: Sage', *The Canadian Journal of Action Research*, vol. 14, no. 1, pp. 69-71.

Adeli, H. & Jiang, X. 2009, 'Intelligent infrastructure: neural networks, wavelets, and chaos theory for intelligent transportation systems and smart structures', CRC Press.

Adeli, H. & Vishnubhotla, P.R. 2020, 'Parallel processing and parallel machines', *Parallel processing in computational mechanics*, CRC Press, pp. 1-20.

Agarwal, Y., Balaji, B., Gupta, R., Lyles, J., Wei, M. & Weng, T. 2010, 'Occupancy-driven energy management for smart building automation', *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*, pp. 1-6.

Ahmed, F. 2017, 'An IoT-big data based machine learning technique for forecasting water requirement in irrigation field', *International conference on research and practical issues of enterprise information systems*, Springer, pp. 67-77.

Akkaya, K., Guvenc, I., Aygun, R., Pala, N. & Kadri, A. 2015, 'IoT-based occupancy monitoring techniques for energy-efficient smart buildings', *Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE*, pp. 58-63.

Al-Sai, Z.A., Abdullah, R. & Husin, M.H. 2019, 'Big Data Impacts and Challenges: A Review', *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pp. 150-5.

Alahakoon, D. & Yu, X. 2013, 'Advanced analytics for harnessing the power of smart meter big data', *Intelligent Energy Systems (IWIES), 2013 IEEE International Workshop on*, pp. 40-5.

Alzoubi, Y.I., Gill, A.Q. & Al-Ani, A. 2016, 'Empirical studies of geographically distributed agile development communication challenges: A systematic review', *Information & Management*, vol. 53, no. 1, pp. 22-37.

Amanullah, M.A., Habeeb, R.A.A., Nasaruddin, F.H., Gani, A., Ahmed, E., Nainar, A.S.M., Akim, N.M. & Imran, M. 2020, 'Deep learning and big data technologies for IoT security', *Computer Communications*, vol. 151, pp. 495-517.

Apache Flume 2018, viewed August 10 2018, <https://flume.apache.org/>.

Apache Lucene 2018, viewed July 07 2018, <http://lucene.apache.org/>.

Apache Pig, viewed August 10 2018, <https://pig.apache.org/>.

Apache Spark, viewed August 10 2018, <http://spark.apache.org/>.

Arnone, D., Croce, V., Paternó, G., Rossi, A., Emma, S., Miceli, R. & Tommaso, A.O.D. 2016, 'Energy management of multi-carrier smart buildings for integrating local renewable energy systems', *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 845-50.

Arnott, D. & Pervan, G.J.D.S.S. 2008, 'Eight key issues for the decision support systems discipline', vol. 44, no. 3, pp. 657-72.

Atkinson, C. & Kuhne, T.J.I.s. 2003, 'Model-driven development: a metamodeling foundation', vol. 20, no. 5, pp. 36-41.

Autoregressive integrated moving average 2018, viewed March 05 2018, <https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average>.

Bansal, M. & Gandhi, B. 2019, 'IoT & Big Data in Smart Healthcare (ECG Monitoring)', *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, IEEE, pp. 390-6.

Basha, A.J., Malathi, M., Balaganesh, S. & Maheshwari, R. 2018, 'Patient Tracking Using IoT and Big Data', *Proceedings of First International Conference on Smart System, Innovations and Computing*, Springer, pp. 613-21.

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.

Bashir, R. 2017, *smartBuildings*, viewed August 10 2018, <https://github.com/c3212218/smartBuildings>.

Bassi, A., Bauer, M., Fiedler, M. & Kranenburg, R.v. 2013, *Enabling things to talk*, Springer-Verlag GmbH.

Bellini, P., Bugli, F., Nesi, P., Pantaleo, G., Paolucci, M. & Zaza, I. 2019, 'Data flow management and visual analytic for big data smart city/IOT', *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, IEEE, pp. 1529-36.

Benbasat, I., Goldstein, D.K. & Mead, M.J.M.q. 1987, 'The case research strategy in studies of information systems', pp. 369-86.

Bera, A., Kundu, A., De Sarkar, N.R. & Mou, D. 2017, 'Experimental analysis on big data in IoT-based architecture', *Proceedings of the international conference on data engineering and communication technology*, Springer, pp. 1-9.

Beydoun, G., Low, G., Henderson-Sellers, B., Mouratidis, H., Gomez-Sanz, J.J., Pavon, J. & Gonzalez-Perez, C.J.I.T.o.S.E. 2009, 'FAML: a generic metamodel for MAS development', vol. 35, no. 6, pp. 841-63.

Bin, N. 2018, 'Research on Methods and Techniques for IoT Big Data Cluster Analysis', *2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*, IEEE, pp. 184-8.

Borrmann, A., König, M., Koch, C. & Beetz, J. 2015, *Building Information Modeling: Technologische Grundlagen und industrielle Praxis*, Springer-Verlag.

Boulakbech, M., Messai, N., Sam, Y., Devogele, T. & Hammoudeh, M. 2017, 'IoT Mashups: From IoT Big Data to IoT Big Service', paper presented to the Proceedings of the International Conference on Future Networks and Distributed Systems, Cambridge, United Kingdom, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3102304.3102324>.

Boumlik, A. & Bahaj, M. 2017, 'Big data and iot: A prime opportunity for banking industry', *International Conference on Advanced Information Technology, Services and Systems*, Springer, pp. 396-407.

Braun, P., Cuzzocrea, A., Leung, C.K., Pazdor, A., Souza, J. & Tanbeer, S.K. 2019, 'Pattern mining from big IoT data with fog computing: models, issues, and research perspectives', *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pp. 584-91.

Buckman, A.H., Mayfield, M. & Beck, S.B.M. 2014, 'What is a Smart Building?', *Smart and Sustainable Built Environment*, vol. 3, no. 2, pp. 92-109.

Cai, H., Xu, B., Jiang, L. & Vasilakos, A.V. 2016, 'IoT-based big data storage systems in cloud computing: perspectives and challenges', *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 75-87.

Caplinskas, A. & Vasilecas, O. 2004, 'Information systems research methodologies and models', *The 5th international conference on Computer systems and technologies*, Citeseer, pp. 1-6.
Caragliu, A., Del Bo, C. & Nijkamp, P. 2011, 'Smart cities in Europe', *Journal of urban technology*, vol. 18, no. 2, pp. 65-82.
Carrillo, E., Benitez, V., Mendoza, C. & Pacheco, J. 2015, 'IoT framework for smart buildings with cloud computing', *2015 IEEE First International Smart Cities Conference (ISC2)*, IEEE, pp. 1-6.
Chatuphale, P. & Armstrong, L. 2017, 'Indian mobile agricultural services using big data and internet of things (IoT)', *International Conference on Intelligent Systems Design and Applications*, Springer, pp. 1028-37.
Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A.V. & Rong, X. 2015, 'Data Mining for the Internet of Things: Literature Review and Challenges', *International Journal of Distributed Sensor Networks*, vol. 2015, p. 14.
Chen, S.-y., Song, S.-f., Li, L.-x. & Shen, J. 2009, 'Survey on smart grid technology [J]', *Power System Technology*, vol. 8, pp. 1-7.
Chen, Y. 2020, 'IoT, cloud, big data and AI in interdisciplinary domains', *Simulation Modelling Practice and Theory*, vol. 102, p. 102070.
Chhowa, T.T., Rahman, M.A., Paul, A.K. & Ahmmed, R. 2019, 'A Narrative Analysis on Deep Learning in IoT based Medical Big Data Analysis with Future Perspectives', *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, pp. 1-6.
Chou, J.-S. & Ngo, N.-T., 'Smart grid data analytics framework for increasing energy savings in residential buildings', *Automation in Construction*.
Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J.R., Mellouli, S., Nahon, K., Pardo, T.A. & Scholl, H.J. 2012, 'Understanding smart cities: An integrative framework', *System Science (HICSS), 2012 45th Hawaii International Conference on*, IEEE, pp. 2289-97.
Cisco 2011, *The Internet of Things How the Next Evolution of the Internet Is Changing Everything*.
Cloudera 2018, viewed March 05 2018, <www.cloudera.com>.
Cozzi, L., Gould, T., Bouckart, S., Crow, D., Kim, T., Mcglade, C., Olejarnik, P., Wanner, B. & Wetzel, D. 2020, 'World Energy Outlook 2020', IEA Paris.
Daissaoui, A., Boulmakoul, A., Karim, L. & Lbath, A. 2020, 'IoT and Big Data Analytics for Smart Buildings: A Survey', *Procedia Computer Science*, vol. 170, pp. 161-8.
Das, S.K. & Yamana, H. 2017, 'Securing big data and IoT networks in smart cyber-physical environments', *Proceedings of the 2017 International Conference on Smart Digital Environment*, pp. 189-94.
Dasgupta, A., Gill, A. & Hussain, F.K. 2019, 'A Conceptual Framework for Data Governance in IoT-enabled Digital IS Ecosystems', *DATA*, pp. 209-16.
Dattana, V., Gupta, K. & Kush, A. 2019, 'A Probability based Model for Big Data Security in Smart City', *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*, IEEE, pp. 1-6.
Demirkan, H. 2013, 'A smart healthcare systems framework', *It Professional*, vol. 15, no. 5, pp. 38-45.
Deng, Q. & Ji, S.J.P.A.j.o.t.a.f.i.s. 2018, 'A review of design science research in information systems: concept, process, outcome, and evaluation', vol. 10, no. 1, p. 2.
Desai, P.V. 2018, 'A survey on big data applications and challenges', *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 737-40.
Dineshreddy, V. & Gangadharan, G.R. 2016, 'Towards an "Internet of Things" framework for financial services sector', *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, pp. 177-81.
Dittakavi, S., Bhamidipati, G. & Neelam, V.S.K. 2018, 'Big Data Analytics via IoT with Cloud Service', *Big Data Analytics*, Springer, pp. 319-28.

Dlamini, N.N. & Johnston, K. 2016, 'The use, benefits and challenges of using the Internet of Things (IoT) in retail businesses: A literature review', *2016 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, pp. 430-6.
Dresch, A., Lacerda, D.P. & Antunes, J.A.V. 2015, 'Design science research', *Design science research*, Springer, pp. 67-102.
Dryjanski, M., Buczkowski, M., Ould-Cheikh-Mouhamedou, Y. & Kliks, A. 2020, 'Adoption of smart cities with a practical smart building implementation', *IEEE Internet of Things Magazine*, vol. 3, no. 1, pp. 58-63.
Duffy, A. & O'Donnel, F.J. 1998, 'A Design Research Approach', *Proceedings of the AID'98 Workshop on Research Methods*, pp. 20-7.
Dupont, C., Bures, T., Sheikhalishahi, M., Pham, C. & Rahim, A. 2017, 'Low-cost IoT, Big Data, and Cloud Platform for Developing Countries', *International Conference on the Economics of Grids, Clouds, Systems, and Services*, Springer, pp. 285-99.
Dupont, C., Sheikhalishahi, M., Biswas, A.R. & Bures, T. 2017, 'IoT, big data, and cloud platform for rural African needs', *2017 IST-Africa Week Conference (IST-Africa)*, IEEE, pp. 1-7.
Dutta, J. & Roy, S. 2017, 'IoT-fog-cloud based architecture for smart city: Prototype of a smart building', *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*, IEEE, pp. 237-42.
Dwivedi, A., Pant, R., Pandey, S. & Kumar, K. 2018, 'Internet of Things'(IoT's) Impact on Decision Oriented Applications of Big Data Sentiment Analysis', *2018 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, IEEE, pp. 1-10.
Dybå, T. & Dingsøyr, T. 2008, 'Empirical studies of agile software development: A systematic review', *Information and Software Technology*, vol. 50, no. 9–10, pp. 833-59.
El-Shafie, M. & Fakeih, L. 2018, 'The smart environment of commercial buildings', *2018 15th Learning and Technology Conference (L&T)*, pp. 161-6.
Elasticsearch 2018, viewed March 05 2018, <https://www.elastic.co/products/elasticsearch>.
Ericsson 2011, *More than 50 Billion Connected Devices*, viewed 29/05/2016.
Fan, R., Li, Y., Cao, Y., Xie, W., Tan, Y. & Cai, Y. 2016, 'An optimization management strategy for energy efficiency of air conditioning loads in smart building', *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, pp. 1-5.
Giacobbe, M., Puliafito, C. & Scarpa, M. 2016, 'The big bucket: An iot cloud solution for smart waste management in smart cities', *European Conference on Service-Oriented and Cloud Computing*, Springer, pp. 43-58.
Gill, A.Q. 2015a, *Adaptive cloud enterprise architecture*, World Scientific.
Gill, A.Q. 2015b, 'Agile enterprise architecture modelling: Evaluating the applicability and integration of six modelling standards', *Information and Software Technology*, vol. 67, pp. 196-206.
Gill, A.Q. & Chew, E. 2019, 'Configuration information system architecture: Insights from applied action design research', *Information & Management*, vol. 56, no. 4, pp. 507-25.
Gill, A.Q., Chew, E.K., Kricker, D. & Bird, G. 2016, 'Adaptive enterprise resilience management: Adaptive action design research in financial services case study', *2016 IEEE 18th Conference on Business Informatics (CBI)*, vol. 1, IEEE, pp. 113-22.
Goes, P.B. 2014, 'Editor's comments: big data and IS research'.
González-Vidal, A., Moreno-Cano, V., Terroso-Sáenz, F. & Skarmeta, A.F. 2016, 'Towards Energy Efficiency Smart Buildings Models Based on Intelligent Data Analytics', *Procedia Computer Science*, vol. 83, pp. 994-9.
Gregor, S. & Hevner, A.R. 2013, 'Positioning and presenting design science research for maximum impact', *MIS quarterly*, pp. 337-55.

Groenli, T.-M., Flesch, B., Mukkamala, R., Vatrapu, R., Klavestad, S. & Bergner, H. 2018, 'Internet of Things big data analytics: The case of noise level measurements at the Roskilde music festival', *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 5153-8.

Guba, E.G. & Lincoln, Y.S. 1994, 'Competing paradigms in qualitative research', *Handbook of qualitative research*, vol. 2, no. 163-194, p. 105.

Gummesson, E. 2000, *Qualitative methods in management research*, Sage.

Haj-Bolouri, A., Bernhardsson, L. & Rossi, M. 2016, 'PADRE: A method for participatory action design research', *International Conference on Design Science Research in Information System and Technology*, Springer, pp. 19-36.

Hajiheydari, N., Talafidaryani, M. & Khabiri, S. 2019, 'IoT Big Data Value Map: How to Generate Value from IoT Data', paper presented to the Proceedings of the 2019 the 5th International Conference on e-Society, e-Learning and e-Technologies, Vienna, Austria, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3312714.3312728>.

Hernández-Ramos, J.L., Moreno, M.V., Bernabé, J.B., Carrillo, D.G. & Skarmeta, A.F. 2015, 'SAFIR: Secure access framework for IoT-enabled services on smart buildings', *Journal of Computer and System Sciences*, vol. 81, no. 8, pp. 1452-63.

Hevner, A. & Chatterjee, S. 2010, 'Design science research in information systems', *Design research in information systems*, Springer, pp. 9-22.

Hevner, A.R. 2007, 'A three cycle view of design science research', *Scandinavian journal of information systems*, vol. 19, no. 2, p. 4.

Hevner, A.R., March, S.T., Park, J. & Ram, S.J.M.q. 2004, 'Design science in information systems research', pp. 75-105.

Hilpert, H., Beckers, C., Kolbe, L.M. & Schumann, M. 2013, 'Green IS for GHG emission reporting on product-level? An action design research project in the meat industry', *International Conference on Design Science Research in Information Systems*, Springer, pp. 324-39.

Horban, V. 2016, 'A multifaceted approach to smart energy city concept through using big data analytics', *2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP)*, pp. 392-6.

Huhtamäki, J. 2016, 'Visualizing co-authorship networks for actionable insights: action design research experiment', *Proceedings of the 20th International Academic Mindtrek Conference*, pp. 208-15.

Iivari, J. & Venable, J.R. 2009, 'Action research and design science research-Seemingly similar but decisively dissimilar'.

Iivari, J.J.E.J.o.I.S. 2015, 'Distinguishing and contrasting two strategies for design science research', vol. 24, no. 1, pp. 107-15.

Jamil, A., Abdullah, M., Javed, M.A. & Hassan, M.S. 2018, 'Comprehensive Review of Challenges & Technologies for Big Data Analytics', *2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET)*, pp. 229-33.

Jedlitschka, A. & Pfahl, D. 2005, 'Reporting guidelines for controlled experiments in software engineering', *2005 International Symposium on Empirical Software Engineering, 2005.*, IEEE, p. 10 pp.

Jell, T., Baumgartner, C., Bröring, A. & Mitic, J. 2018, 'BIG IoT: interconnecting IoT platforms from different domains—first success story', *Information Technology-New Generations*, Springer, pp. 721-4.

Jia, M., Srinivasan, R.S. & Raheem, A.A. 2017, 'From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency', *Renewable and Sustainable Energy Reviews*, vol. 68, Part 1, pp. 525-40.

Jin, J., Gubbi, J., Marusic, S. & Palaniswami, M. 2014, 'An Information Framework for Creating a Smart City Through Internet of Things', *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 112-21.

Jung, J.-j., Kim, K. & Park, J. 2019, 'Framework of Big data Analysis about IoT-Home-device for supporting a decision making an effective strategy about new product design', *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, IEEE, pp. 582-4.

Kakiz, M.T., Öztürk, E. & Çavdar, T. 2017, 'A novel SDN-based IoT architecture for big data', *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, IEEE, pp. 1-5.

Kalamaras, I., Kaklanis, N., Votis, K. & Tzovaras, D. 2018, 'Towards Big Data Analytics in Large-Scale Federations of Semantically Heterogeneous IoT Platforms', *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer, pp. 13-23.

Kassela, E., Provatas, N., Tsiourvas, A., Konstantinou, I. & Koziris, N. 2019, 'BigOptiBase: Big Data Analytics for Base Station Energy Consumption Optimization', *2019 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 6098-100.

Kaur, H. & Kushwaha, A.S. 2018, 'A review on integration of big data and IoT', *2018 4th International Conference on Computing Sciences (ICCS)*, IEEE, pp. 200-3.

Kaur, M.J. & Maheshwari, P. 2016, 'Building smart cities applications using IoT and cloud-based architectures', *2016 International Conference on Industrial Informatics and Computer Systems (CIICS)*, pp. 1-5.

Kavousian, A., Rajagopal, R. & Fischer, M. 2015, 'Ranking appliance energy efficiency in households: Utilizing smart meter data and energy efficiency frontiers to estimate and identify the determinants of appliance energy efficiency in residential buildings', *Energy and Buildings*, vol. 99, pp. 220-30.

Keijzer-Broers, W., Florez-Atehortua, L. & De Reuver, M. 2016, 'Prototyping a health and wellbeing platform: An action design research approach', *2016 49th Hawaii International Conference on System Sciences (HICSS)*, IEEE, pp. 3462-71.

Khare, S. & Totaro, M. 2019, 'Big Data in IoT', *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, pp. 1-7.

Kho, D.D., Lee, S. & Zhong, R.Y. 2018, 'Big data analytics for processing time analysis in an IoT-enabled manufacturing shop floor', *Procedia Manufacturing*, vol. 26, pp. 1411-20.

Kibana 2018, viewed March 05 2018, <https://www.elastic.co/products/kibana>.

Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M. & Linkman, S. 2010, 'Systematic literature reviews in software engineering – A tertiary study', *Information and Software Technology*, vol. 52, no. 8, pp. 792-805.

Ku, T.-Y., Park, W.-K. & Choi, H. 2017, 'Energy information collecting agent for IoT big data system', *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, IEEE, pp. 1134-6.

Lachhab, F., Bakhouya, M., Ouladsine, R. & Essaaidi, M. 2018, 'Towards an Intelligent Approach for Ventilation Systems Control using IoT and Big Data Technologies', *Procedia Computer Science*, vol. 130, pp. 926-31.

Lanka, S. & Bajracharya, S. 2018, 'Incorporating Big Data with IoT in Data Center Analytics', *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, IEEE, pp. 240-5.

Lempinen, H., Rossi, M. & Tuunainen, V.K. 2012, 'Design principles for inter-organizational systems development–case Hansel', *International Conference on Design Science Research in Information Systems*, Springer, pp. 52-65.

Li, F., Xie, R., Wang, Z., Guo, L., Ye, J., Ma, P. & Song, W. 2019, 'Online Distributed IoT Security Monitoring with Multidimensional Streaming Big Data', *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4387-94.

Lin, C.-L., Lin, H.-L., Lin, S.-C. & Liu, Y.-T. 2019, 'Long Term Healthcare System for Elders by Using Internet of Things with Big Data', paper presented to the Proceedings of the 3rd International Conference on Computer Science and Application Engineering, Sanya, China, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3331453.3362042>.

Lin, C.Y., Li, C.C., Huang, W.H., Liao, W.C. & Chen, W.M. 2014, 'A Sensor Data Processing and Access Platform Based on Hadoop for Smart Environments', *2014 17th International Conference on Network-Based Information Systems*, pp. 455-60.

Linder, L., Vionnet, D., Bacher, J.-P. & Hennebert, J. 2017, 'Big building data-a big data platform for smart buildings', *Energy Procedia*, vol. 122, pp. 589-94.

Litchfield, A. & Althouse, J. 2014, 'A systematic review of cloud computing, big data and databases on the cloud'.

Liu, G., Zhu, W., Saunders, C., Gao, F. & Yu, Y. 2015, 'Real-time Complex Event Processing and Analytics for Smart Grid', *Procedia Computer Science*, vol. 61, pp. 113-9.

Logstash 2018, viewed August 10 2018, <https://www.elastic.co/products/logstash>.

Luo, X.J., Oyedele, L.O., Ajayi, A.O., Monyei, C.G., Akinade, O.O. & Akanbi, L.A. 2019, 'Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands', *Advanced Engineering Informatics*, vol. 41, p. 100926.

Lwin, K.K., Sekimoto, Y., Takeuchi, W. & Zettsu, K. 2019, 'City Geospatial Dashboard: IoT and Big Data Analytics for Geospatial Solutions Provider in Disaster Management', *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, IEEE, pp. 1-4.

Magruk, A. 2015, 'The Most Important Aspects of Uncertainty in the Internet of Things Field – Context of Smart Buildings', *Procedia Engineering*, vol. 122, pp. 220-7.

Malik, V. & Singh, S. 2019, 'Cloud, Big Data & IoT: Risk Management', *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, IEEE, pp. 258-62.

March, S.T. & Smith, G.F.J.D.s.s. 1995, 'Design and natural science research on information technology', vol. 15, no. 4, pp. 251-66.

Mazzara, M., Afanasyev, I., Sarangi, S.R., Distefano, S., Kumar, V. & Ahmad, M. 2019, 'A reference architecture for smart and software-defined buildings', *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, pp. 167-72.

Mendoza, N. 2016, *New £24 million "Internet of Things" research hub announced*, 2021, <https://www.imperial.ac.uk/news/170033/new-24-million-internet-things-research/>.

Microsoft 2020, *Microsoft Power BI*, 2020, viewed August 10 2018, <https://powerbi.microsoft.com/>.

Mohammadi, M., Al-Fuqaha, A., Sorour, S. & Guizani, M. 2018, 'Deep learning for IoT big data and streaming analytics: A survey', *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2923-60.

Morales, G.D.F., Bifet, A., Khan, L., Gama, J. & Fan, W. 2016, 'IoT Big Data Stream Mining', paper presented to the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/2939672.2945385>.

Moreira, F., Ferreira, M.J. & Cardoso, A. 2017, 'Higher education disruption through IoT and Big Data: A conceptual approach', *International Conference on Learning and Collaboration Technologies*, Springer, pp. 389-405.

Moreno, M.V., Dufour, L., Skarmeta, A.F., Jara, A.J., Genoud, D., Ladevie, B. & Bezian, J.-J. 2016, 'Big data: the key to energy efficiency in smart buildings', *Soft Computing*, vol. 20, no. 5, pp. 1749-62.

Moumen, A. 2020, 'Adoption of Big Data, Cloud Computing & IoT in Morocco Perception of Public Administrations Collaborators', *Proceedings of the 4th Edition of International Conference on Geo-IT and Water Resources 2020, Geo-IT and Water Resources 2020*, pp. 1-4.

Mullassery, D.J. 2016, 'Sensors And Analytics for Smart Buildings'.
Nath, N.R., Narayanaswami, P. & Mohan, G.G. 2016, 'Intelligent query placement strategy for progressive-real time analytics in big data', *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, pp. 1-4.
Nicolalde, F.C., Silva, F., Herrera, B. & Pereira, A. 2018, 'Big data analytics in IOT: challenges, open research issues and tools', *World Conference on Information Systems and Technologies*, Springer, pp. 775-88.
Nie, X., Fan, T., Wang, B., Li, Z., Shankar, A. & Manickam, A. 2020, 'Big Data analytics and IoT in Operation safety management in Under Water Management', *Computer Communications*, vol. 154, pp. 188-96.
Node-RED 2021, *Node-RED*, Low-code programming for event-driven applications, viewedviewed August 10 2018, <https://nodered.org/>.
Ntehelang, G., Isong, B., Lugayizi, F. & Dladlu, N. 2019, 'IoT-based big data analytics issues in healthcare', *Proceedings of the 3rd International Conference on Telecommunications and Communication Engineering*, pp. 16-21.
Onal, A.C., Sezer, O.B., Ozbayoglu, M. & Dogdu, E. 2017, 'Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning', *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 2037-46.
Onizuka, M. 2017, 'Trend and applications of Big Data and IoT techniques', *Proceedings of the Eighth International Symposium on Information and Communication Technology*, pp. 5-.
Orlikowski, W.J. & Baroudi, J.J.J.I.s.r. 1991, 'Studying information technology in organizations: Research approaches and assumptions', vol. 2, no. 1, pp. 1-28.
Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A. & Sinz, E.J.J.E.j.o.i.s. 2011, 'Memorandum on design-oriented information systems research', vol. 20, no. 1, pp. 7-10.
Othman, S.H. & Beydoun, G. 2010, 'Metamodelling approach to support disaster management knowledge sharing'.
Pandharipande, A., Caicedo, D. & Wang, X. 2014, 'Sensor-Driven Wireless Lighting Control: System Solutions and Services for Intelligent Buildings', *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4207-15.
Parvin, S., Venkatraman, S., de Souza-Daw, T., Fahd, K., Jackson, J., Kaspi, S., Cooley, N., Saleem, K. & Gawanmeh, A. 2019, 'Smart food security system using iot and big data analytics', *16th International Conference on Information Technology-New Generations (ITNG 2019)*, Springer, pp. 253-8.
Peffers, K., Tuunanen, T., Rothenberger, M.A. & Chatterjee, S. 2007, 'A design science research methodology for information systems research', *Journal of management information systems*, vol. 24, no. 3, pp. 45-77.
Peña, M., Biscarri, F., Guerrero, J.I., Monedero, I. & León, C. 2016, 'Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach', *Expert Systems with Applications*, vol. 56, pp. 242-55.
Plageras, A.P., Psannis, K.E., Stergiou, C., Wang, H. & Gupta, B.B. 2018, 'Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings', *Future Generation Computer Systems*, vol. 82, pp. 349-57.
Power BI 2018, viewed March 05 2018, <https://powerbi.microsoft.com/en-us/>.
Prat, N., Comyn-Wattiau, I. & Akoka, J. 2014, 'Artifact Evaluation in Information Systems Design-Science Research-a Holistic View', *PACIS*, p. 23.

Puthal, D., Ranjan, R., Nepal, S. & Chen, J. 2017, 'IoT and big data: An architecture with data flow and security issues', *Cloud infrastructures, services, and IoT systems for smart cities*, Springer, pp. 243-52.

Pycharm 2018, viewed March 05 2018, <https://www.jetbrains.com/pycharm/>.

Rajeswari, S., Suthendran, K. & Rajakumar, K. 2017, 'A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics', *2017 international conference on intelligent computing and control (I2C2)*, IEEE, pp. 1-5.

Rani, M.S. & Geetavani, B. 2017, 'Design and analysis for improving reliability and accuracy of big-data based peripheral control through IoT', *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, IEEE, pp. 749-53.

Rathore, M.M., Ahmad, A., Paul, A. & Rho, S. 2016, 'Urban planning and building smart cities based on the Internet of Things using Big Data analytics', *Computer Networks*, vol. 101, pp. 63-80.

Rathore, M.M., Paul, A., Hong, W.-H., Seo, H., Awan, I. & Saeed, S. 2018, 'Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data', *Sustainable Cities and Society*, vol. 40, pp. 600-10.

Rogerson, C., Scott, E.J.I.T. & Education, S. 2014, 'Motivating an action design research approach to implementing online training in an organisational context'.

Romero, C.D.G., Barriga, J.K.D. & Molano, J.I.R. 2016, 'Big data meaning in the architecture of IoT for smart cities', *International Conference on Data Mining and Big Data*, Springer, pp. 457-65.

Romme, A.G.L.J.O.s. 2003, 'Making a difference: Organization as design', vol. 14, no. 5, pp. 558-73.

Roy, S., Ray, R., Roy, A., Sinha, S., Mukherjee, G., Pyne, S., Mitra, S., Basu, S. & Hazra, S. 2017, 'IoT, big data science & analytics, cloud computing and mobile app based hybrid system for smart agriculture', *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, IEEE, pp. 303-4.

Runeson, P. & Höst, M. 2009, 'Guidelines for conducting and reporting case study research in software engineering', *Empirical software engineering*, vol. 14, no. 2, p. 131.

Rzadca, K., Findeisen, P., Swiderski, J., Zych, P., Broniek, P., Kusmierek, J., Nowak, P., Strack, B., Witusowski, P. & Hand, S. 2020, 'Autopilot: workload autoscaling at Google', *Proceedings of the Fifteenth European Conference on Computer Systems*, pp. 1-16.

Saha, H.N., Auddy, S., Pal, S., Kumar, S., Pandey, S., Singh, R., Singh, A.K., Banerjee, S., Ghosh, D. & Saha, S. 2017, 'Disaster management using Internet of Things', *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, pp. 81-5.

Santoso, F.K. & Vun, N.C.H. 2015, 'Securing IoT for smart home system', *2015 International Symposium on Consumer Electronics (ISCE)*, pp. 1-2.

Sein, M.K., Henfridsson, O., Purao, S., Rossi, M. & Lindgren, R.J.M.q. 2011, 'Action design research', pp. 37-56.

Seo, S. & Chung, J.-M. 2019, 'Outlier Detection Technique for IoT Sensor-Driven Big Data Systems', *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, IEEE, pp. 1-2.

Shah, J. & Mishra, B. 2016, 'Customized IoT-Enabled Wireless Sensing and Monitoring Platform for Smart Buildings', *Procedia Technology*, vol. 23, pp. 256-63.

Shyam, R., Ganesh H.B, B., Kumar S, S., Poornachandran, P. & Soman, K.P. 2015, 'Apache Spark a Big Data Analytics Platform for Smart Grid', *Procedia Technology*, vol. 21, pp. 171-8.

Singh, P.K., Verma, R.K. & Prasad, P.K. 2019, 'IoT-Based Smartbots for Smart City Using MCC and Big Data', *Smart Intelligent Computing and Applications*, Springer, pp. 525-34.

Singh, S. & Yassine, A. 2018, 'IoT big data analytics with fog computing for household energy management in smart grids', *International Conference on Smart Grid and Internet of Things*, Springer, pp. 13-22.

Sivaraman, E. & Manickachezian, R. 2014, 'High Performance and Fault Tolerant Distributed File System for Big Data Storage and Processing Using Hadoop', *Intelligent Computing Applications (ICICA), 2014 International Conference on*, pp. 32-6.

Smith, N.C.J.J.o.I.T. 1990, 'The case study: a useful research method for information management', vol. 5, no. 3, pp. 123-33.

Sood, S.K., Sandhu, R., Singla, K. & Chang, V. 2018, 'IoT, big data and HPC based smart flood management framework', *Sustainable Computing: Informatics and Systems*, vol. 20, pp. 102-17.

Srivastava, L. 2004, 'Japan's ubiquitous mobile information society', *info*, vol. 6, no. 4, pp. 234-51.

Stannard, J., Writer-Davies, R., Spielman, D. & Nurse, J. 2020, *Consumer Attitudes Towards IoT Security*.

Tableau 2018, viewed March 05 2018, <https://www.tableau.com/>.

Tableau 2020, *Tableau*, viewed 2020 December 10 2020, <https://www.tableau.com/>.

Taherkordi, A., Eliassen, F. & Horn, G. 2017, 'From IoT big data to IoT big services', *Proceedings of the Symposium on Applied Computing*, pp. 485-91.

Theodoridis, E., Mylonas, G. & Chatzigiannakis, I. 2013, 'Developing an IoT Smart City framework', *Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on*, pp. 1-6.

TheThingsNetwork 2021, *TheThingsNetwork*, Building a global open LoRaWAN Network, viewed 19/06/2021 2021, <https://www.thethingsnetwork.org/>.

ThingsBoard 2021, *ThingsBoard*, ThingsBoard Open-source IoT Platform, viewed 19/06/2021 2021, <https://thingsboard.io/>.

Toomer, E. & Bowen, K.J.J.o.t.O.R.S. 1993, 'Qualitative methods in management research', vol. 44, no. 7, pp. 735-6.

TopQuadrant 2020, *TopBraid Enterprise Data Governance*, TopQuadrant, viewed 13 July 2020 2020, <https://www.topquadrant.com/products/topbraid-enterprise-data-governance/>.

Ullah, F., Habib, M.A., Farhan, M., Khalid, S., Durrani, M.Y. & Jabbar, S. 2017, 'Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare', *Sustainable Cities and Society*, vol. 34, pp. 90-6.

United-Nations 2009, *Summary for Decision-Makers*, p. 62.

Uusitalo, M.A. 2006, 'Global Vision for the Future Wireless World from the WWRF', *IEEE Vehicular Technology Magazine*, vol. 1, no. 2, pp. 4-8.

Van Aken, J.E.J.B.j.o.m. 2005, 'Management research as a design science: Articulating the research products of mode 2 knowledge production in management', vol. 16, no. 1, pp. 19-36.

Varma, C. 2018, 'Performance Analysis and Challenges of the Map Reduce Framework in Big Data Analytics', *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pp. 1-5.

Vasicek, D., Jalowiczor, J., Sevcik, L. & Voznak, M. 2018, 'IoT Smart Home Concept', *2018 26th Telecommunications Forum (TELFOR)*, pp. 1-4.

Venkatraman, S., Overmars, A., Fahd, K., Parvin, S. & Kaspi, S. 2020, 'Security Challenges for Big Data and IoT', *Proceedings of the 2020 2nd International Conference on Big Data Engineering and Technology*, pp. 1-6.

Walker, M. 2014, *Descriptive Predictive Prescriptive Analytics*, viewed 19/06/2021 2021, <https://www.datascienceassn.org/content/descriptive-predictive-prescriptive-analytics>.

Wazid, M., Das, A.K., Hussain, R., Succi, G. & Rodrigues, J.J.P.C. 2019, 'Authentication in cloud-driven IoT-based big data environment: Survey and outlook', *Journal of Systems Architecture*, vol. 97, pp. 185-96.

Weng, T. & Agarwal, Y. 2012, 'From Buildings to Smart Buildings—Sensing and Actuation to Improve Energy Efficiency', *IEEE Design & Test of Computers*, vol. 29, no. 4, pp. 36-44.

Winter, R.J.E.J.o.I.S. 2008, 'Design science research in Europe', vol. 17, no. 5, pp. 470-5.

Xia, F., Yang, L.T., Wang, L. & Vinel, A. 2012, 'Internet of Things', *International Journal of Communication Systems*, vol. 25, no. 9, pp. 1101-2.

Yacchirema, D., de Puga, J.S., Palau, C. & Esteve, M. 2018, 'Fall detection system for elderly people using IoT and Big Data', *Procedia Computer Science*, vol. 130, pp. 603-10.

Yadav, P. & Vishwakarma, S. 2018, 'Application of Internet of Things and big data towards a smart city', *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, IEEE, pp. 1-5.

Yang, H.-F., Chen, C.-H.K. & Chen, K.-L.B. 2019, 'Using Big Data Analytics and Visualization to Create IoT-enabled Science Park Smart Governance Platform', *International Conference on Human-Computer Interaction*, Springer, pp. 459-72.

Yassine, A., Singh, S., Hossain, M.S. & Muhammad, G. 2019, 'IoT big data analytics for smart homes with fog and cloud computing', *Future Generation Computer Systems*, vol. 91, pp. 563-73.

Yin, R.K. 2012, 'Case study methods'.

Zainal, Z.J.J.k. 2007, 'Case study as a research method', vol. 5, no. 1.

Zeng, D., Guo, S. & Cheng, Z. 2011, 'The Web of Things: A Survey', *Journal of Communications*, vol. 6, no. 6, pp. 424-38.

Zhang, K., Ni, J., Yang, K., Liang, X., Ren, J. & Shen, X.S. 2017, 'Security and Privacy in Smart City Applications: Challenges and Solutions', *IEEE Communications Magazine*, vol. 55, no. 1, pp. 122-9.

Zhou, K., Fu, C. & Yang, S. 2016, 'Big data driven smart energy management: From big data to big insights', *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 215-25.

Zhou, K. & Yang, S. 2016, 'Understanding household energy consumption behavior: The contribution of energy big data analytics', *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 810-9.

Appendix A – selected studies for review

[S1] (Zhou, Fu & Yang 2016) - Zhou, K., Fu, C. & Yang, S. 2016, 'Big data driven smart energy management: From big data to big insights', Renewable and Sustainable Energy Reviews, vol. 56, pp. 215-25.

[S2] (Alahakoon & Yu 2013) - Alahakoon, D. & Yu, X. 2013, 'Advanced analytics for harnessing the power of smart meter big data', Intelligent Energy Systems (IWIES), 2013 IEEE International Workshop on, pp. 40-5.

[S3] (Shyam et al. 2015) - Shyam, R., Ganesh H.B, B., Kumar S, S., Poornachandran, P. & Soman, K.P. 2015, 'Apache Spark a Big Data Analytics Platform for Smart Grid', Procedia Technology, vol. 21, pp. 171-8.

[S4] (Magruk 2015) - Magruk, A. 2015, 'The Most Important Aspects of Uncertainty in the Internet of Things Field – Context of Smart Buildings', Procedia Engineering, vol. 122, pp. 220-7.

[S5] (Hernández-Ramos et al. 2015) - Hernández-Ramos, J.L., Moreno, M.V., Bernabé, J.B., Carrillo, D.G. & Skarmeta, A.F. 2015, 'SAFIR: Secure access framework for IoT-enabled services on smart buildings', Journal of Computer and System Sciences, vol. 81, no. 8, pp. 1452-63.

[S6] (Akkaya et al. 2015) - Akkaya, K., Guvenc, I., Aygun, R., Pala, N. & Kadri, A. 2015, 'IoT-based occupancy monitoring techniques for energy-efficient smart buildings', Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE, pp. 58-63.

[S7] (Rathore et al. 2016) - Rathore, M.M., Ahmad, A., Paul, A. & Rho, S. 2016, 'Urban planning and building smart cities based on the Internet of Things using Big Data analytics', Computer Networks, vol. 101, pp. 63-80.

[S8] (Theodoridis, Mylonas & Chatzigiannakis 2013) - Theodoridis, E., Mylonas, G. & Chatzigiannakis, I. 2013, 'Developing an IoT Smart City framework', Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on, pp. 1-6.

[S9] (Lin et al. 2014) - Lin, C.Y., Li, C.C., Huang, W.H., Liao, W.C. & Chen, W.M. 2014, 'A Sensor Data Processing and Access Platform Based on Hadoop for Smart Environments', 2014 17th International Conference on Network-Based Information Systems, pp. 455-60.

[S10] (Jin et al. 2014) - Jin, J., Gubbi, J., Marusic, S. & Palaniswami, M. 2014, 'An Information Framework for Creating a Smart City Through Internet of Things', IEEE Internet of Things Journal, vol. 1, no. 2, pp. 112-21.

[S11] (Kaur & Maheshwari 2016) - Kaur, M.J. & Maheshwari, P. 2016, 'Building smart cities applications using IoT and cloud-based architectures', 2016 International Conference on Industrial Informatics and Computer Systems (CIICS), pp. 1-5.

[S12] (Liu et al. 2015) - Liu, G., Zhu, W., Saunders, C., Gao, F. & Yu, Y. 2015, 'Real-time Complex Event Processing and Analytics for Smart Grid', Procedia Computer Science, vol. 61, pp. 113-9.

[S13] (Weng & Agarwal 2012) - Weng, T. & Agarwal, Y. 2012, 'From Buildings to Smart Buildings—Sensing and Actuation to Improve Energy Efficiency', IEEE Design & Test of Computers, vol. 29, no. 4, pp. 36-44.

[S14] (Moreno et al. 2016) - Moreno, M.V., Dufour, L., Skarmeta, A.F., Jara, A.J., Genoud, D., Ladevie, B. & Bezian, J.-J. 2016, 'Big data: the key to energy efficiency in smart buildings', Soft Computing, vol. 20, no. 5, pp. 1749-62.

[S15] (González-Vidal et al. 2016) - González-Vidal, A., Moreno-Cano, V., Terroso-Sáenz, F. & Skarmeta, A.F. 2016, 'Towards Energy Efficiency Smart Buildings Models Based on Intelligent Data Analytics', Procedia Computer Science, vol. 83, pp. 994-9.

[S16] (Chou & Ngo) - Chou, J.-S. & Ngo, N.-T., 'Smart grid data analytics framework for increasing energy savings in residential buildings', Automation in Construction.

[S17] (Zhou & Yang 2016) - Zhou, K. & Yang, S. 2016, 'Understanding household energy consumption behavior: The contribution of energy big data analytics', Renewable and Sustainable Energy Reviews, vol. 56, pp. 810-9.

[S18] (Kavousian, Rajagopal & Fischer 2015) - Kavousian, A., Rajagopal, R. & Fischer, M. 2015, 'Ranking appliance energy efficiency in households: Utilizing smart meter data and energy efficiency frontiers to estimate and identify the determinants of appliance energy efficiency in residential buildings', Energy and Buildings, vol. 99, pp. 220-30.

[S19] (Peña et al. 2016) - Peña, M., Biscarri, F., Guerrero, J.I., Monedero, I. & León, C. 2016, 'Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach', Expert Systems with Applications, vol. 56, pp. 242-55.

[S20] (Santoso & Vun 2015) - Santoso, F.K. & Vun, N.C.H. 2015, 'Securing IoT for smart home system', 2015 International Symposium on Consumer Electronics (ISCE), pp. 1-2.

[S21] (Pandharipande, Caicedo & Wang 2014) - Pandharipande, A., Caicedo, D. & Wang, X. 2014, 'Sensor-Driven Wireless Lighting Control: System Solutions and Services for Intelligent Buildings', IEEE Sensors Journal, vol. 14, no. 12, pp. 4207-15.

[S22] (Shah & Mishra 2016) - Shah, J. & Mishra, B. 2016, 'Customized IoT-Enabled Wireless Sensing and Monitoring Platform for Smart Buildings', Procedia Technology, vol. 23, pp. 256-63.

[S23] (Chhowa et al. 2019) - Chhowa, T.T., Rahman, M.A., Paul, A.K. & Ahmmed, R. 2019, 'A Narrative Analysis on Deep Learning in IoT based Medical Big Data Analysis with Future Perspectives', 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, pp. 1-6.

[S24] (Kakiz, Öztürk & Çavdar 2017) - Kakiz, M.T., Öztürk, E. & Çavdar, T. 2017, 'A novel SDN-based IoT architecture for big data', 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), IEEE, pp. 1-5

[S25] (Dattana, Gupta & Kush 2019) - Dattana, V., Gupta, K. & Kush, A. 2019, 'A Probability based Model for Big Data Security in Smart City', 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), IEEE, pp. 1-6.

[S26] (Kaur & Kushwaha 2018) - Kaur, H. & Kushwaha, A.S. 2018, 'A review on integration of big data and IoT', 2018 4th International Conference on Computing Sciences (ICCS), IEEE, pp. 200-3.

[S27] (Rajeswari, Suthendran & Rajakumar 2017) - Rajeswari, S., Suthendran, K. & Rajakumar, K. 2017, 'A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics', 2017 international conference on intelligent computing and control (I2C2), IEEE, pp. 1-5.

[S28] (Yadav & Vishwakarma 2018) - Yadav, P. & Vishwakarma, S. 2018, 'Application of Internet of Things and big data towards a smart city', 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU), IEEE, pp. 1-5.

[S29] (Khare & Totaro 2019) - Khare, S. & Totaro, M. 2019, 'Big Data in IoT', 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp. 1-7.

[S30] (Kassela et al. 2019) - Kassela, E., Provatas, N., Tsiourvas, A., Konstantinou, I. & Koziris, N. 2019, 'BigOptiBase: Big Data Analytics for Base Station Energy Consumption Optimization', 2019 IEEE International Conference on Big Data (Big Data), IEEE, pp. 6098-100.

[S31] (Lwin et al. 2019) - Lwin, K.K., Sekimoto, Y., Takeuchi, W. & Zettsu, K. 2019, 'City Geospatial Dashboard: IoT and Big Data Analytics for Geospatial Solutions Provider in Disaster Management', 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), IEEE, pp. 1-4.

[S32] (Malik & Singh 2019) - Malik, V. & Singh, S. 2019, 'Cloud, Big Data & IoT: Risk Management', 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, pp. 258-62.

[S33] (Bellini et al. 2019) - Bellini, P., Bugli, F., Nesi, P., Pantaleo, G., Paolucci, M. & Zaza, I. 2019, 'Data flow management and visual analytic for big data smart city/IOT', 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), IEEE, pp. 1529-36.

[S34] (Mohammadi et al. 2018) - Mohammadi, M., Al-Fuqaha, A., Sorour, S. & Guizani, M. 2018, 'Deep learning for IoT big data and streaming analytics: A survey', IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2923-60.

[S35] (Rani & Geetavani 2017) - Rani, M.S. & Geetavani, B. 2017, 'Design and analysis for improving reliability and accuracy of big-data based peripheral control through IoT', 2017 International Conference on Trends in Electronics and Informatics (ICEI), IEEE, pp. 749-53.

[S36] (Ku, Park & Choi 2017) - Ku, T.-Y., Park, W.-K. & Choi, H. 2017, 'Energy information collecting agent for IoT big data system', 2017 International Conference on Information and Communication Technology Convergence (ICTC), IEEE, pp. 1134-6.

[S37] (Jung, Kim & Park 2019) - Jung, J.-j., Kim, K. & Park, J. 2019, 'Framework of Big data Analysis about IoT-Home-device for supporting a decision making an effective strategy about new product design', 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), IEEE, pp. 582-4.

[S38] (Lanka & Bajracharya 2018) - Lanka, S. & Bajracharya, S. 2018, 'Incorporating Big Data with IoT in Data Center Analytics', 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE, pp. 240-5

[S39] (Groenli et al. 2018) - Groenli, T.-M., Flesch, B., Mukkamala, R., Vatrapu, R., Klavestad, S. & Bergner, H. 2018, 'Internet of Things big data analytics: The case of noise level measurements at the Roskilde music festival', 2018 IEEE International Conference on Big Data (Big Data), IEEE, pp. 5153-8.

[S40] (Dwivedi et al. 2018) - Dwivedi, A., Pant, R., Pandey, S. & Kumar, K. 2018, 'Internet of Things'(IoT's) Impact on Decision Oriented Applications of Big Data Sentiment Analysis', 2018 3rd International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), IEEE, pp. 1-10.

[S41] (Bansal & Gandhi 2019) - Bansal, M. & Gandhi, B. 2019, 'IoT & Big Data in Smart Healthcare (ECG Monitoring)', 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, pp. 390-6.

[S42] (Roy et al. 2017) - Roy, S., Ray, R., Roy, A., Sinha, S., Mukherjee, G., Pyne, S., Mitra, S., Basu, S. & Hazra, S. 2017, 'IoT, big data science & analytics, cloud computing and mobile app based hybrid system for smart agriculture', 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), IEEE, pp. 303-4.

[S43] (Dupont, Sheikhalishahi, et al. 2017) - Dupont, C., Sheikhalishahi, M., Biswas, A.R. & Bures, T. 2017, 'IoT, big data, and cloud platform for rural African needs', 2017 IST-Africa Week Conference (IST-Africa), IEEE, pp. 1-7.

[S44] (Cai et al. 2016) - Cai, H., Xu, B., Jiang, L. & Vasilakos, A.V. 2016, 'IoT-based big data storage systems in cloud computing: perspectives and challenges', IEEE Internet of Things Journal, vol. 4, no. 1, pp. 75-87.

[S45] (Li et al. 2019) - Li, F., Xie, R., Wang, Z., Guo, L., Ye, J., Ma, P. & Song, W. 2019, 'Online Distributed IoT Security Monitoring with Multidimensional Streaming Big Data', IEEE Internet of Things Journal, vol. 7, no. 5, pp. 4387-94.

[S46] (Seo & Chung 2019) - Seo, S. & Chung, J.-M. 2019, 'Outlier Detection Technique for IoT Sensor-Driven Big Data Systems', 2019 International Conference on Electronics, Information, and Communication (ICEIC), IEEE, pp. 1-2.

[S47] (Braun et al. 2019) - Braun, P., Cuzzocrea, A., Leung, C.K., Pazdor, A., Souza, J. & Tanbeer, S.K. 2019, 'Pattern mining from big IoT data with fog computing: models, issues, and research perspectives', 2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), pp. 584-91.

[S48] (Bin 2018) - Bin, N. 2018, 'Research on Methods and Techniques for IoT Big Data Cluster Analysis', 2018 International Conference on Information Systems and Computer Aided Education (ICISCAE), IEEE, pp. 184-8.

[S49] (Onal et al. 2017) - Onal, A.C., Sezer, O.B., Ozbayoglu, M. & Dogdu, E. 2017, 'Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning', 2017 IEEE International Conference on Big Data (Big Data), IEEE, pp. 2037-46.

[S50] (Ahmed 2017) - Ahmed, F. 2017, 'An IoT-big data based machine learning technique for forecasting water requirement in irrigation field', International conference on research and practical issues of enterprise information systems, Springer, pp. 67-77.

[S51] (Nicolalde et al. 2018) - Nicolalde, F.C., Silva, F., Herrera, B. & Pereira, A. 2018, 'Big data analytics in IOT: challenges, open research issues and tools', World Conference on Information Systems and Technologies, Springer, pp. 775-88.

[S52] (Dittakavi, Bhamidipati & Neelam 2018) - Dittakavi, S., Bhamidipati, G. & Neelam, V.S.K. 2018, 'Big Data Analytics via IoT with Cloud Service', Big Data Analytics, Springer, pp. 319-28.

[S53] (Boumlik & Bahaj 2017) - Boumlik, A. & Bahaj, M. 2017, 'Big data and iot: A prime opportunity for banking industry', International Conference on Advanced Information Technology, Services and Systems, Springer, pp. 396-407.

[S54] (Romero, Barriga & Molano 2016) - Romero, C.D.G., Barriga, J.K.D. & Molano, J.I.R. 2016, 'Big data meaning in the architecture of IoT for smart cities', International Conference on Data Mining and Big Data, Springer, pp. 457-65.

[S55] (Jell et al. 2018) - Jell, T., Baumgartner, C., Bröring, A. & Mitic, J. 2018, 'BIG IoT: interconnecting IoT platforms from different domains—first success story', Information Technology-New Generations, Springer, pp. 721-4.

[S56] (Bera et al. 2017) - Bera, A., Kundu, A., De Sarkar, N.R. & Mou, D. 2017, 'Experimental analysis on big data in iot-based architecture', Proceedings of the international conference on data engineering and communication technology, Springer, pp. 1-9.

[S57] (Moreira, Ferreira & Cardoso 2017) - Moreira, F., Ferreira, M.J. & Cardoso, A. 2017, 'Higher education disruption through IoT and Big Data: A conceptual approach', International Conference on Learning and Collaboration Technologies, Springer, pp. 389-405.

[S58] (Chatuphale & Armstrong 2017) - Chatuphale, P. & Armstrong, L. 2017, 'Indian mobile agricultural services using big data and internet of things (IoT)', International Conference on Intelligent Systems Design and Applications, Springer, pp. 1028-37.

[S59] (Puthal et al. 2017) - Puthal, D., Ranjan, R., Nepal, S. & Chen, J. 2017, 'IoT and big data: An architecture with data flow and security issues', Cloud infrastructures, services, and IoT systems for smart cities, Springer, pp. 243-52.

[S60] (Singh & Yassine 2018) - Singh, S. & Yassine, A. 2018, 'IoT big data analytics with fog computing for household energy management in smart grids', International Conference on Smart Grid and Internet of Things, Springer, pp. 13-22.

[61] (Singh, Verma & Prasad 2019) - Singh, P.K., Verma, R.K. & Prasad, P.K. 2019, 'IoT-Based Smartbots for Smart City Using MCC and Big Data', Smart Intelligent Computing and Applications, Springer, pp. 525-34.

[S62] (Dupont, Bures, et al. 2017) - Dupont, C., Bures, T., Sheikhalishahi, M., Pham, C. & Rahim, A. 2017, 'Low-cost IoT, Big Data, and Cloud Platform for Developing Countries', International Conference on the Economics of Grids, Clouds, Systems, and Services, Springer, pp. 285-99.

[S63] (Basha et al. 2018) - Basha, A.J., Malathi, M., Balaganesh, S. & Maheshwari, R. 2018, 'Patient Tracking Using IoT and Big Data', Proceedings of First International Conference on Smart System, Innovations and Computing, Springer, pp. 613-21.

[S64] (Parvin et al. 2019) - Parvin, S., Venkatraman, S., de Souza-Daw, T., Fahd, K., Jackson, J., Kaspi, S., Cooley, N., Saleem, K. & Gawanmeh, A. 2019, 'Smart food security system using iot and big data analytics', 16th International Conference on Information Technology-New Generations (ITNG 2019), Springer, pp. 253-8.

[S65] (Giacobbe, Puliafito & Scarpa 2016) - Giacobbe, M., Puliafito, C. & Scarpa, M. 2016, 'The big bucket: An iot cloud solution for smart waste management in smart cities', European Conference on Service-Oriented and Cloud Computing, Springer, pp. 43-58.

[S66] (Ahmed 2017) - Ahmed, F. 2017, 'An IoT-big data based machine learning technique for forecasting water requirement in irrigation field', International conference on research and practical issues of enterprise information systems, Springer, pp. 67-77.

[S67] (Kalamaras et al. 2018) - Kalamaras, I., Kaklanis, N., Votis, K. & Tzovaras, D. 2018, 'Towards Big Data Analytics in Large-Scale Federations of Semantically Heterogeneous IoT Platforms', IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, pp. 13-23.

[S68] (Yang, Chen & Chen 2019) - Yang, H.-F., Chen, C.-H.K. & Chen, K.-L.B. 2019, 'Using Big Data Analytics and Visualization to Create IoT-enabled Science Park Smart Governance Platform', International Conference on Human-Computer Interaction, Springer, pp. 459-72.

[S69] (Moumen 2020) - Moumen, A. 2020, 'Adoption of Big Data, Cloud Computing & IoT in Morocco Perception of Public Administrations Collaborators', Proceedings of the 4th Edition of International Conference on Geo-IT and Water Resources 2020, Geo-IT and Water Resources 2020, pp. 1-4.

[S70] (Taherkordi, Eliassen & Horn 2017) - Taherkordi, A., Eliassen, F. & Horn, G. 2017, 'From IoT big data to IoT big services', Proceedings of the Symposium on Applied Computing, pp. 485-91.

[S71] (Morales et al. 2016) - Morales, G.D.F., Bifet, A., Khan, L., Gama, J. & Fan, W. 2016, 'IoT Big Data Stream Mining', paper presented to the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/2939672.2945385>.

[S72] (Hajiheydari, Talafidaryani & Khabiri 2019) - Hajiheydari, N., Talafidaryani, M. & Khabiri, S. 2019, 'IoT Big Data Value Map: How to Generate Value from IoT Data', paper presented to the Proceedings of the 2019 the 5th International Conference on e-Society, e-Learning and e-

Technologies, Vienna, Austria, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3312714.3312728>.

[S73] (Boulakbech et al. 2017) - Boulakbech, M., Messai, N., Sam, Y., devogele, T. & Hammoudeh, M. 2017, 'IoT Mashups: From IoT Big Data to IoT Big Service', paper presented to the Proceedings of the International Conference on Future Networks and Distributed Systems, Cambridge, United Kingdom, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3102304.3102324>.

[S74] (Ntehelang et al. 2019) - Ntehelang, G., Isong, B., Lugayizi, F. & Dladlu, N. 2019, 'IoT-based big data analytics issues in healthcare', Proceedings of the 3rd International Conference on Telecommunications and Communication Engineering, pp. 16-21.

[S75] (Lin et al. 2019) - Lin, C.-L., Lin, H.-L., Lin, S.-C. & Liu, Y.-T. 2019, 'Long Term Healthcare System for Elders by Using Internet of Things with Big Data', paper presented to the Proceedings of the 3rd International Conference on Computer Science and Application Engineering, Sanya, China, <https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3331453.3362042>.

[S76] (Das & Yamana 2017) - Das, S.K. & Yamana, H. 2017, 'Securing big data and IoT networks in smart cyber-physical environments', Proceedings of the 2017 International Conference on Smart Digital Environment, pp. 189-94.

[S77] (Venkatraman et al. 2020) - Venkatraman, S., Overmars, A., Fahd, K., Parvin, S. & Kaspi, S. 2020, 'Security Challenges for Big Data and IoT', Proceedings of the 2020 2nd International Conference on Big Data Engineering and Technology, pp. 1-6.

[S78] (Onizuka 2017) - Onizuka, M. 2017, 'Trend and applications of Big Data and IoT techniques', Proceedings of the Eighth International Symposium on Information and Communication Technology, pp. 5.

[S79] (Wazid et al. 2019) - Wazid, M., Das, A.K., Hussain, R., Succi, G. & Rodrigues, J.J.P.C. 2019, 'Authentication in cloud-driven IoT-based big data environment: Survey and outlook', Journal of Systems Architecture, vol. 97, pp. 185-96.

[S80] (Nie et al. 2020) - Nie, X., Fan, T., Wang, B., Li, Z., Shankar, A. & Manickam, A. 2020, 'Big Data analytics and IoT in Operation safety management in Under Water Management', Computer Communications, vol. 154, pp. 188-96.

[S81] (Kho, Lee & Zhong 2018) - Kho, D.D., Lee, S. & Zhong, R.Y. 2018, 'Big data analytics for processing time analysis in an IoT-enabled manufacturing shop floor', Procedia Manufacturing, vol. 26, pp. 1411-20.

[S82] (Amanullah et al. 2020) - Amanullah, M.A., Habeeb, R.A.A., Nasaruddin, F.H., Gani, A., Ahmed, E., Nainar, A.S.M., Akim, N.M. & Imran, M. 2020, 'Deep learning and big data technologies for IoT security', Computer Communications, vol. 151, pp. 495-517.

[S83] (Luo et al. 2019) - Luo, X.J., Oyedele, L.O., Ajayi, A.O., Monyei, C.G., Akinade, O.O. & Akanbi, L.A. 2019, 'Development of an IoT-based big data platform for day-ahead prediction of building heating and cooling demands', Advanced Engineering Informatics, vol. 41, p. 100926.

[S84] (Plageras et al. 2018) - Plageras, A.P., Psannis, K.E., Stergiou, C., Wang, H. & Gupta, B.B. 2018, 'Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings', Future Generation Computer Systems, vol. 82, pp. 349-57.

[S85] (Rathore et al. 2018) - Rathore, M.M., Paul, A., Hong, W.-H., Seo, H., Awan, I. & Saeed, S. 2018, 'Exploiting IoT and big data analytics: Defining Smart Digital City using real-time urban data', Sustainable Cities and Society, vol. 40, pp. 600-10.

[S86] (Yacchirema et al. 2018) - Yacchirema, D., de Puga, J.S., Palau, C. & Esteve, M. 2018, 'Fall detection system for elderly people using IoT and Big Data', Procedia Computer Science, vol. 130, pp. 603-10.

[S87] (Daissaoui et al. 2020) - Daissaoui, A., Boulmakoul, A., Karim, L. & Lbath, A. 2020, 'IoT and Big Data Analytics for Smart Buildings: A Survey', Procedia Computer Science, vol. 170, pp. 161-8.

[S88] (Yassine et al. 2019) - Yassine, A., Singh, S., Hossain, M.S. & Muhammad, G. 2019, 'IoT big data analytics for smart homes with fog and cloud computing', Future Generation Computer Systems, vol. 91, pp. 563-73.

[S89] (Sood et al. 2018) - Sood, S.K., Sandhu, R., Singla, K. & Chang, V. 2018, 'IoT, big data and HPC based smart flood management framework', Sustainable Computing: Informatics and Systems, vol. 20, pp. 102-17.

[S90] (Chen 2020) - Chen, Y. 2020, 'IoT, cloud, big data and AI in interdisciplinary domains', Simulation Modelling Practice and Theory, vol. 102, p. 102070

[S91] (Ullah et al. 2017) - Ullah, F., Habib, M.A., Farhan, M., Khalid, S., Durrani, M.Y. & Jabbar, S. 2017, 'Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare', Sustainable Cities and Society, vol. 34, pp. 90-6.

[S92] (Lachhab et al. 2018) - Lachhab, F., Bakhouya, M., Ouladsine, R. & Essaaidi, M. 2018, 'Towards an Intelligent Approach for Ventilation Systems Control using IoT and Big Data Technologies', Procedia Computer Science, vol. 130, pp. 926-31.

[S93] (Plageras et al. 2018) - Plageras, A.P., Psannis, K.E., Stergiou, C., Wang, H. & Gupta, B.B. 2018, 'Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings', Future Generation Computer Systems, vol. 82, pp. 349-57.

[S94] (Linder et al. 2017) - Linder, L., Vionnet, D., Bacher, J.-P. & Hennebert, J. 2017, 'Big building data-a big data platform for smart buildings', Energy Procedia, vol. 122, pp. 589-94.

[S95] (Mazzara et al. 2019) - Mazzara, M., Afanasyev, I., Sarangi, S.R., Distefano, S., Kumar, V. & Ahmad, M. 2019, 'A reference architecture for smart and software-defined buildings', 2019 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, pp. 167-72.

[S96] (Dutta & Roy 2017) - Dutta, J. & Roy, S. 2017, 'IoT-fog-cloud based architecture for smart city: Prototype of a smart building', 2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence, IEEE, pp. 237-42.

[S97] (Dryjanski et al. 2020) - Dryjanski, M., Buczkowski, M., Ould-Cheikh-Mouhamedou, Y. & Kliks, A. 2020, 'Adoption of smart cities with a practical smart building implementation', IEEE Internet of Things Magazine, vol. 3, no. 1, pp. 58-63

[S98] (Carrillo et al. 2015) - Carrillo, E., Benitez, V., Mendoza, C. & Pacheco, J. 2015, 'IoT framework for smart buildings with cloud computing', 2015 IEEE First International Smart Cities Conference (ISC2), IEEE, pp. 1-6.

Appendix B: Ethics Approval (email approval from Ethics committee) ETH21-6056

"Dear Applicant,

Re: ETH21-6056 - "Ethics Approval for Big Data Management and Analytics Framework for IoT-enabled Smart Buildings"

Your local research office has reviewed your application and agreed that it now meets the requirements of the National Statement on Ethical Conduct in Human

Research (2007) and has been approved on that basis. You are therefore authorised to commence activities as outlined in your application, subject to any conditions detailed in this document.

You are reminded that this letter constitutes ethics approval only. This research project must also be undertaken in accordance with all UTS policies and guidelines including the Research Management Policy.

Your approval number is UTS HREC REF NO. ETH21-6056

Approval will be for a period of five (5) years from the date of this correspondence subject to the submission of annual progress reports.

The following standard conditions apply to your approval:

- Your approval number must be included in all participant material and advertisements. Any advertisements on Staff Connect without an approval number will be removed.
- The Principal Investigator will immediately report anything that might warrant review of ethical approval of the project to the Ethics Secretariat.
- The Principal Investigator will notify the Committee of any event that requires a modification to the protocol or other project documents and submit any required amendments prior to implementation. Instructions on how to submit an amendment application can be found here.

- The Principal Investigator will promptly report adverse events to the Ethics Secretariat. An adverse event is any event (anticipated or otherwise) that has a negative impact on participants, researchers or the reputation of the University. Adverse events can also include privacy breaches, loss of data and damage to property.
- The Principal Investigator will report to the UTS HREC or UTS MREC annually and notify the Committee when the project is completed at all sites. The Principal Investigator will notify the Committee of any plan to extend the duration of the project past the approval period listed above.
- The Principal Investigator will obtain any additional approvals or authorisations as required (e.g. from other ethics committees, collaborating institutions, supporting organisations).
- The Principal Investigator will notify the Committee of his or her inability to continue as Principal Investigator including the name of and contact information for a replacement.

This research must be undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research.

You should consider this your official letter of approval.

If you have any queries about this approval, or require any amendments to your approval in future, please do not hesitate to contact your local research office or the Ethics Secretariat."

Appendix C: Participant Information Sheet and Consent Form

PARTICIPANT INFORMATION SHEET

[ETH21-6056] – Big Data Management and Analytics Framework for IoT-enabled Smart Buildings

WHO IS CONDUCTING THIS RESEARCH?

My name is *Rizwan Bashir* and I am a student at UTS. My supervisor is *Professor Ghassan Beydoun*.

WHAT IS THE RESEARCH ABOUT?

The aim of the research is to develop a big data management and analytics framework for IoT-enabled smart buildings. The framework will be known as IBDMA (Integrated Big Data Management and Analytics) framework and will consist of two elements. One is the reference architecture and the other one is the metamodel. The IBDMA framework will allow the autonomous monitoring and control of various facilities in the smart building.

WHY HAVE I BEEN INVITED?

You have been invited to participate as an industry expert because of your distinguished expertise in the areas of big data and IoT.
Your contact details were obtained from LinkedIn by searching for people based in Australia with expertise in IoT or big data.
Before you decide to participate in this research study, please check the selection criteria.

Selection criteria:

Only participants who have extensive industry experience in IoT or big data will be requested and invited to provide their feedback. The participants must be based in Australia and must have a designation of manager or above in the relevant domain.

FUNDING

Funding for this project has been received from Commonwealth Government as student funding help for higher education and research students.

WHAT DOES MY PARTICIPATION INVOLVE?

If you decide to participate, I will invite you to participate in a one-and-a-half-hour session with the student for this activity. The activity will involve

- The research student to give a presentation on the research conducted and the developed framework.
- The research student will give a demo on the developed framework.
- At the end of the demo, you will be given a questionnaire seeking your feedback on the developed framework. The questionnaire will record qualitative feedback on the developed framework.

Further information:

- No travelling or payment is required
- The presentation and demo will be conducted online via a Teams/Zoom meeting.
- The questionnaire will be provided via email, and the participant will be asked to fill in the questionnaire and send the completed questionnaire back to the student.
- The data will not include any information that may identify you in any way. No personal data will be collected; the data collected will be completely anonymous.
- The data will be stored in UTS systems as per the UTS research data management policy on the UTS-recommended cloud storage CloudStor. Only the student and student's supervisor will have secure login access to CloudStor.
- The collected anonymised data will be used for the publication of conference papers, journal papers and a research thesis.

ARE THERE ANY RISKS/INCONVENIENCE?

There is no risk (low category) because participation only involves an online presentation and an online demo of the IBDMA framework. It only involves technical and software content. Therefore, it is highly unlikely that any risk will occur.

DO I HAVE TO TAKE PART IN THIS RESEARCH PROJECT?

Participation in this study is voluntary. It is completely up to you whether or not you decide to take part.
If you decide not to participate or to withdraw from the study, it will not affect your relationship with the researchers or the University of Technology Sydney. Participation or non-participation will have no bearing on student course progression or assessment.

WHAT IF I WITHDRAW FROM THIS RESEARCH PROJECT?

If you wish to withdraw from the study once it has started, you can do so at any time without having to give a reason by contacting Rizwan Bashir (@student.uts.edu.au). If you withdraw from the study, your data and any feedback you provided will be discarded and destroyed.

WHAT WILL HAPPEN TO INFORMATION ABOUT ME?

By signing the consent form, you consent to the research team collecting and using personal information about you for the research project. All this information will be treated confidentially. The responses will be stored in UTS systems as per UTS research data management policy. Only my supervisor and I have access to data via a UTS secure login. The collected anonymous data from your response to the online survey form will not identify you in any way and will only be used for the purpose of this research project (thesis) and publications (conference and journal papers).

WHAT IF I HAVE CONCERNS OR A COMPLAINT?

If you have concerns about the research that you think the research student or his supervisor can help you with, please feel free to contact me on @student.uts.edu.au or ghassan.beydoun@uts.edu.au

You will be given a copy of this form to keep.

NOTE:

This study has been approved in line with the University of Technology Sydney Human Research Ethics Committee [UTS HREC] guidelines. If you have any concerns or complaints about any aspect of the conduct of this research, please contact the Ethics Secretariat on ph.: +61 2 9514 2478 or email: Research.Ethics@uts.edu.au] and quote the UTS HREC reference number. Any matter raised will be treated confidentially, investigated and you will be informed of the outcome.

CONSENT FORM

[ETH21-6056] – Big Data Management and Analytics Framework for IoT-enabled Smart Buildings

I ____________________ agree to participate in the research project being conducted by Rizwan Bashir, [redacted]@student.uts.edu.au (phone: [redacted]). I understand that funding for this research has been provided by the Commonwealth Government as student funding help for higher education and research students.

I have read the Participant Information Sheet or someone has read it to me in a language that I understand.

I understand the purposes, procedures and risks of the research as described in the Participant Information Sheet.

I have had an opportunity to ask questions and I am satisfied with the answers I have received.

I freely agree to participate in this research project as described and understand that I am free to withdraw at any time without affecting my relationship with the researchers or the University of Technology Sydney.

I understand that I will be given a signed copy of this document to keep.

I am aware that I can contact Rizwan Bashir on [redacted]@student.uts.edu.au if I have any concerns about the research.

____ ____/____/____

Name and Signature [participant] Date

____ _____/_____/_____

Name and Signature [researcher or delegate] Date

Appendix D: Invitation letter

“Dear *(Participant Name)*,

Hope you are well.

We have been connected on LinkedIn for some time and I have been closely watching your LinkedIn posts and the LinkedIn articles you have written on IoT and the cloud edge.

I am working in the industry in the big data domain. In addition, I am also completing a PhD from the University of Technology Sydney (UTS). As part of my PhD research, I have developed and evaluated a big data management and analytics framework for IoT-enabled smart buildings.

I would like to invite you to participate in a one-hour session to provide feedback on my developed framework. During this session, I will give a presentation about the framework I have developed and will also give you a short demo.

At the end of the presentation and demo, you will be given a questionnaire with five questions regarding the framework. You will be required to provide written responses to the questionnaire.

Your participation is voluntary and you can back out from participation at any stage during this process. Your personal data will remain anonymous; however, your feedback may be used in research publications and a thesis.

I am attaching the participant information sheet and the consent form with this email.

If you are happy to participate, please read and sign the form and send back to me on [redacted]@student.uts.edu.au.

Once you agree and have provided me with your written consent, I will schedule a time with you to go through the framework and to obtain your feedback.

Thanks and regards,

Rizwan Bashir

UTS PhD research student

Appendix E: Empirical Study Data

This section contains the source of the data used in the empirical evaluation in this thesis. The data have been stored on CloudStor, the UTS-recommended cloud storage service. Only the thesis author (Rizwan Bashir), the joint supervisors (Dr Ghassan Beydoun and Dr Asif Gill) have access to the data files on CloudStor. The empirical original data files are organised as follows.

- Ethics approval letter: https://cloudstor.aarnet.edu.au/plus/s/CHibYP439Ou6Qpw
- Industry experts' feedback: https://cloudstor.aarnet.edu.au/plus/f/5722317790
- IBDMA presentation: https://cloudstor.aarnet.edu.au/plus/f/6799096010

Note: The author's LinkedIn page has been used to communicate with industry experts working in the big data or IoT domains to request the participants to provide their feedback based on the invitation letter (see Appendix C).

Note: All links in Appendix D were active at the time of the thesis publication.

Appendix F: Research Papers

This appendix lists the publications that were produced as outcomes of this research (2016–2021):

- Bashir, M.R., Gill, A.Q. & Beydoun, G. 2021, ‘A reference architecture for IoT-enabled Smart Buildings’, *Complex and Intelligent Systems*, (submitted March 2021).
- 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. (https://ieeexplore.ieee.org/document/9195814)
- Bashir, M.R. & Gill, A.Q. 2017, 'IoT-enabled smart buildings: A systematic review', *2017 Intelligent Systems Conference (IntelliSys)*, IEEE, pp. 151-9. (https://ieeexplore.ieee.org/document/8324283)
- 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. (https://ieeexplore.ieee.org/document/7828529)

Appendix G: Test cases

Test Case 01 - Detection of oxygen levels

Test Case 01	**Detection of low oxygen level and autonomously activating HVAC system.**
Context	Oxygen levels fall below human comfort levels at a specific location in the smart building.
Problem	Low oxygen levels in the smart building may remain unnoticed by smart building management which could result in the discomfort of the residents and could potentially prove fatal.
Solution	The oxygen level falls and is detected by the IBDMA framework. Management is notified, and the HVAC is autonomously activated within two minutes.
Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: the terminal displays the message saying the HVAC turned ON when the value of the sensor fell below the threshold.
Description	Oxygen levels in the smart building fall, the oxygen sensor senses that the levels of oxygen at a particular location have fallen below the threshold levels and sends the data to HDFS via Flume. The system detects the low oxygen levels, activates the associated HVAC system installed at that particular location and notifies the smart building management within two minutes. The system autonomously turns the HVAC system off when the oxygen levels are in an acceptable range.
Consequences	The residents enjoy a comfortable environment. The smart building management are notified within two minutes if the levels fall below the threshold levels.

Test Case 02 - Detection of fire

Test Case 02	**Detection of fire and alerting local community.**
Context	A fire erupts at a specific location in the smart building.
Problem	Fire erupts in the smart building and may remain unnoticed by the smart building management for a longer period potentially resulting in the loss of infrastructure, investment and lives.
Solution	Fire erupts and is detected by the IBDMA framework. Management and the Fire Brigade are notified within two minutes of the fire eruption and a fire alarm at that location is turned ON.
Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: The terminal displays a message saying Fire Alarm turned ON when the value of the sensor rose above the threshold.

Description	Fire erupts in a smart building; the IoT smoke detector sensor senses fire and sends the data to HDFS via Flume. The system detects fire, activates the associated fire alarm installed at that particular location alerting the local community and notifies the smart building management including the relevant fire brigade team within two minutes. The fire brigade team then acts on the notification and eradicates the fire at the location.
Consequences	The local community is alerted by activating the fire alarm so that they can move to safe areas. The fire brigade and building management are notified within two minutes so that they can respond to the fire to minimize loss of infrastructure, investment and lives.

Test Case 03 - Detection of luminosity

Test Case 03	**Detection of low luminosity level and autonomously activating smart lights.**
Context	Luminosity levels fall below human luminous comfort levels at a particular location in the smart building.
Problem	Low luminous levels in the smart building may remain unnoticed by the smart building management that may not only result in the discomfort of the residents but could also be a safety hazard for the residents.
Solution	The luminosity level falls and is detected by the IBDMA framework. Management is notified, and the smart lights are autonomously activated within two minutes.
Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: The terminal displays the message saying the Lights turned ON when the value of the luminosity sensor fell below the threshold.
Description	Luminosity levels in the building fall, the luminosity sensor senses that the luminosity levels at a location have fallen below the threshold levels and sends the data to HDFS via Flume. The system detects the low luminous levels, activates the associated smart lights installed at that location and notifies the smart building management within two minutes. The system autonomously turns the smart lights off when the luminosity levels are in the acceptable range.
Consequences	The residents enjoy comfortable luminous levels. The smart building management are notified within two minutes if the levels fall below the threshold levels.

Test Case 04 - Detection of parking space usage

Test Case 04	**Detect if parking lot becomes full and alert the residents.**
Context	A parking lot of a smart building is full.
Problem	A parking lot of a smart building becomes full and may remain unnoticed by the residents for a long period resulting in inconvenience for the residents.
Solution	A parking lot becomes full and is detected by the IBDMA framework. The smart building residents are notified within two minutes of the parking lot becoming full and are notified to move to alternate parking areas.

Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: The terminal displays the message saying the parking space is occupied when the value of the sensor is '1'.
Description	A parking lot of a smart building is full and the IoT parking lot sensor senses this and sends the data to HDFS via Flume. The system alerts the local community and notifies the smart building management within two minutes.
Consequences	The residents are alerted so that they can move to alternative parking lots or areas. The building management is notified within two minutes so that they can respond to the growing needs of the residents by planning more parking lots in the areas.

Test Case 05 - Detection of garbage

Test Case 05	**Detect if garbage bin becomes full and alert the residents.**
Context	A garbage bin in the smart building becomes full.
Problem	A garbage bin in the smart building becomes full and may remain unnoticed by the smart building garbage management team for a long period, resulting in inconvenience for the residents.
Solution	A garbage bin becomes full and is detected by the IBDMA framework. The smart building management is notified within two minutes of the garbage bin becoming full and the residents are notified to either wait until the bin is emptied or throw their garbage in an alternative location.
Test Metrics	Test duration: 60 mins, number of records in data: 60, detection measure: The terminal displays the message saying the Garbage Bin is full when the value of the sensor is '1'.
Description	A garbage bin in a smart building becomes full; the IoT garbage detection sensor senses this and sends the data to HDFS via Flume. The system alerts the local community and notifies the smart building management within two minutes. The smart building garbage management team sends the garbage collector to empty the garbage bin at the location.
Consequences	The residents are alerted so that they can throw their garbage in an alternative garbage bin. The smart building garbage management team is notified within two minutes so that they can respond to the growing needs of the residents by sending garbage collectors more often to the specified location.

The five test cases discussed above were evaluated against the thirteen ECs presented in Fig. 2 and the evaluation results are shown in the following table.

Evaluation results for evaluating each test case against the thirteen ECs

Test Case	Evaluation Criteria	Pass✓/Fail✗
01		
	EC1	✓
	EC2	✓

	EC3	✓
	EC4	✓
	EC5	✓
	EC6	✓
	EC7	✓
	EC8	✓
	EC9	✓
	EC10	✓
	EC11	✓
	EC12	✓
	EC13	✓
02		
	EC1	✓
	EC2	✓
	EC3	✓
	EC4	✓
	EC5	✓
	EC6	✓
	EC7	✓
	EC8	✓
	EC9	✓
	EC10	✓
	EC11	✓
	EC12	✓
	EC13	✓
03		
	EC1	✓
	EC2	✓
	EC3	✓
	EC4	✓
	EC5	✓
	EC6	✓
	EC7	✓
	EC8	✓
	EC9	✓
	EC10	✓
	EC11	✓
	EC12	✓
	EC13	✓

04		
	EC1	✓
	EC2	✓
	EC3	✓
	EC4	✓
	EC5	✓
	EC6	✓
	EC7	✓
	EC8	✓
	EC9	✓
	EC10	✓
	EC11	✓
	EC12	✓
	EC13	✓
05		
	EC1	✓
	EC2	✓
	EC3	✓
	EC4	✓
	EC5	✓
	EC6	✓
	EC7	✓
	EC8	✓
	EC9	✓
	EC10	✓
	EC11	✓
	EC12	✓
	EC13	✓

Appendix H: Questionnaire for Industry Experts to provide feedback

[ETH21-6056] – Big Data Management and Analytics Framework for IoT-enabled Smart Buildings

The participants will be given an online presentation and an online demo of the IBDMA framework. After the presentation and the demo, the participants will be given the following questionnaire to record their responses.

The participant will be allowed to ask any questions about the framework any time during this process. The participants will be asked the following questions.

1- Please provide comments about the IBDMA Reference Architecture. Does it fulfil its intended purpose?

____

____

____

2- Please provide comments about the IBDMA Metamodel. Does it fulfil its intended purpose?

____

____

____

3- Please provide comments on the tools used to implement the IBDMA framework.

4- Please provide feedback on the setup and implementation process of the IBDMA framework.

5- Please provide your overall feedback on the research (including any areas of improvement).

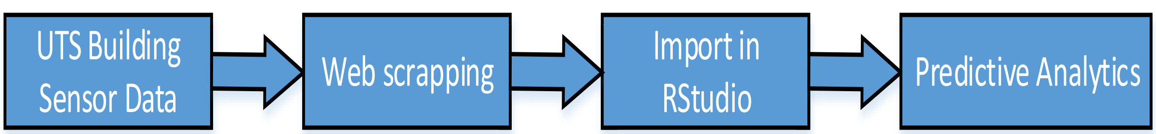
Appendix I: IBDMA Reference Architecture Iteration Process

The five iterations for the development of the reference architecture are discussed in detail below:

- Iteration 1 (Physical)
- Iteration 2 (Physical, real-time)
- Iteration 3 (Virtual, Smart Building)
- Iteration 4 (Virtual, Smart Building, Improved)
- Iteration 5 (Virtual, Smart Building, Improved and Finalized)

Iteration 1:

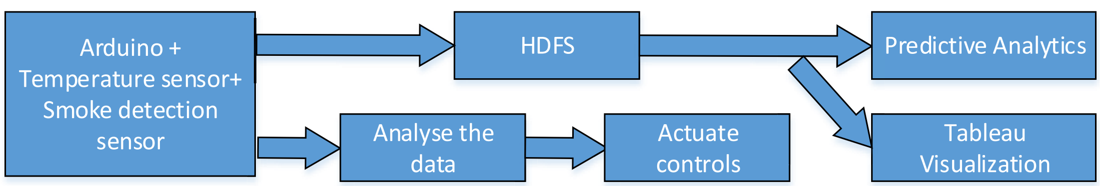
In the first iteration, we extracted UTS sensor data from UTS building 11 and imported these into RStudio by utilizing web-scrapping techniques. The data is available publicly on UTS's web portal. This data was graphed and plotted for various sensors and predictions were made about the sensor data using ARIMA (AutoRegressive Integrated Moving Average) (*Autoregressive integrated moving average* 2018) model. The implementation of graphs and the ARIMA model was done in RStudio using R. This exercise was done to become familiar with the sensor types and the data available. It also helped us familiarize ourselves with the tools and techniques that we could use for future research at that time. Figure 4.8 demonstrates the components and the steps that were taken in Iteration 1.



Iteration 1 – Initial architecture design (analysis of UTS Building 11 sensor data)

Iteration 2:

In the second iteration, we prototyped a physical system consisting of an Arduino microprocessor board, physical sensors and a linear actuator. The data was sent to HDFS (Hadoop Distributed File System) for storage from where it was imported to RStudio for predictive analytics. The problem with Iteration 1 was that the data available was only batched data. Our focus was on both batched as well as real-time data, so we decided to prototype a system with a couple of physical sensors connected to a microprocessor. The sensors considered for this iteration were temperature and smoke detection sensors. The sensors produced real-time data after regular intervals. A linear actuator was also connected to the system to simulate the behaviour of a fire extinguisher scenario. The sensors generated the data in real-time, the data was stored in HDFS, from HDFS we could perform predictive analytics as well as visualize it in Tableau (*Tableau* 2018; Ahmed 2017). The data generated from the sensors were also analyzed in real time as it was generated. If the values generated by the temperature sensor and the smoke detection sensor went above the threshold (simulating a fire scenario), the linear actuator was activated for 5 seconds simulating that the fire extinguisher is activated to rectify the fire. The linear actuator will go off if the sensors read a value below the threshold. The steps and processes followed in Iteration 2 are shown in Figure 4.9.



Iteration 2 – Real-time data analysis and actuation using physical devices

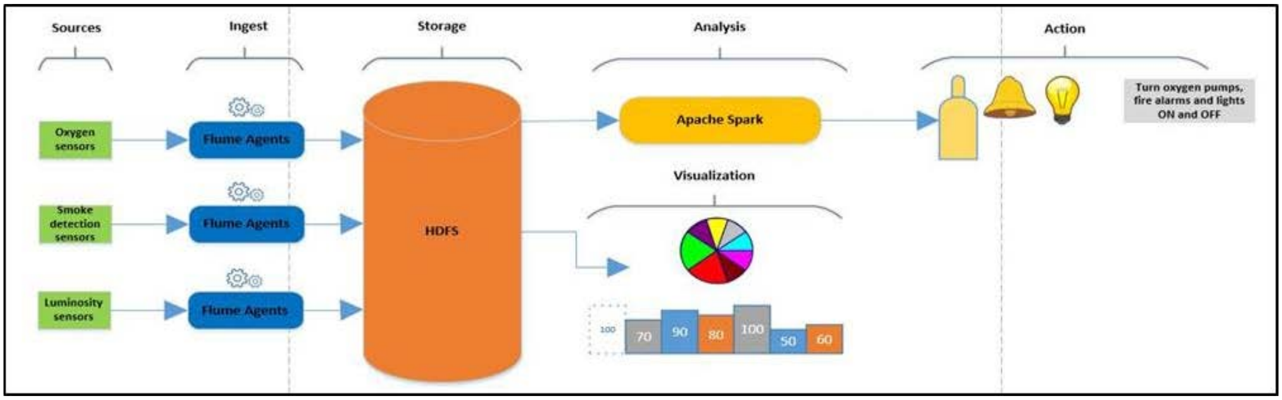
Iteration 3:

In the third iteration, we focused on scaling up the architecture developed in the second iteration. For this iteration, we considered a smart building application scenario by introducing big data pipelining, storage and analysis tools. It was not possible to have access to a large number of physical IoT sensors and actuators in a lab environment, thus we decided to virtualize the IoT sensors by simulating the sensor data. Similarly, we simulated (virtualized) the actions taken based on the data obtained from the virtual sensors. This work has been published in (Bashir & Gill 2016). The architecture developed for this iteration is shown in Figure 4.10.

For this iteration, we virtualized the data generation from fifteen virtual sensors using a Python application. These fifteen sensors include five (IoT) oxygen sensors, five smoke detection sensors and five luminosity sensors deployed in a smart building. These fifteen sensors are assumed to be deployed at five different locations (e.g. different rooms or floors) of the smart building in such a way that each location has a set of these three different sensor types i.e. oxygen, smoke and luminosity.

IoT sensor (source) data is ingested into HDFS (sink) through Apache Flume over a TCP (Transmission Control Protocol) port. For the implementation, we made use of the Cloudera (*Cloudera* 2018; Weng & Agarwal 2012) Big Data platform (Virtual Machine for the Apache Hadoop environment) for extraction, ingesting, data pipelining, storing and analysing the data. To ingest data into HDFS, Flume was the tool of choice because of Flume's robust integration with HDFS compared to Kafka (*5 Most Important Differences Between Apache Kafka vs Flume* 2020). MQTT is a widely used protocol for IoT data, however, MQTT is primarily used as a machine-to-machine protocol for transferring data between two physical systems. Since our goal is to move data to HDFS, we use Flume for data ingestion. There are a number of other tools available including Apache Beam, Apache Flink, Apache Storm, Apache NiFi and Apache Ignite that can be used for streaming data analysis and event processing. However, for this research and proof of concept prototype, we used Flume to ingest data and Apache Spark for its analysis. For the analysis of data to enable decision making, we developed an Apache Spark algorithm using PySpark [20]. The algorithm reads and analyses the data from three different types of IoT sensors stored in HDFS in near real-time to enable effective decision making. For instance, if the oxygen sensors generate data indicating a low oxygen concentration in a given location of the smart building, the Spark algorithm will in turn enable the HVAC system to turn ON to ensure comfortable oxygen concentration levels are attained in that location. The system represents this by outputting "HVAC X turned ON" on the Cloudera terminal, where X denotes the room or floor in a smart building. For the oxygen concentration threshold levels to turn the HVAC system ON or OFF, we defined the oxygen concentration threshold value of 14. However, if the oxygen concentration levels are above the threshold levels, the deployed infrastructure will represent this as "Oxygen level at X ok" on the Cloudera terminal, demonstrating that the oxygen concentration level in a specific location is above the comfortable threshold levels and that no further action is required to enable or disable the HVAC system. If the smoke detection is turned ON by the system due to high smoke, the system would turn the fire alarm ON and will represent this by outputting "Fire alarm X turned ON" where X represents the room or level where smoke is detected.

Likewise, if during the data analytics process, a particular luminosity sensor detects lower than the minimum luminosity levels, the system will turn the lights in that location ON. This is represented by the system displaying "Lights at X turned ON" at the Cloudera terminal where X denotes the location of the smart building.

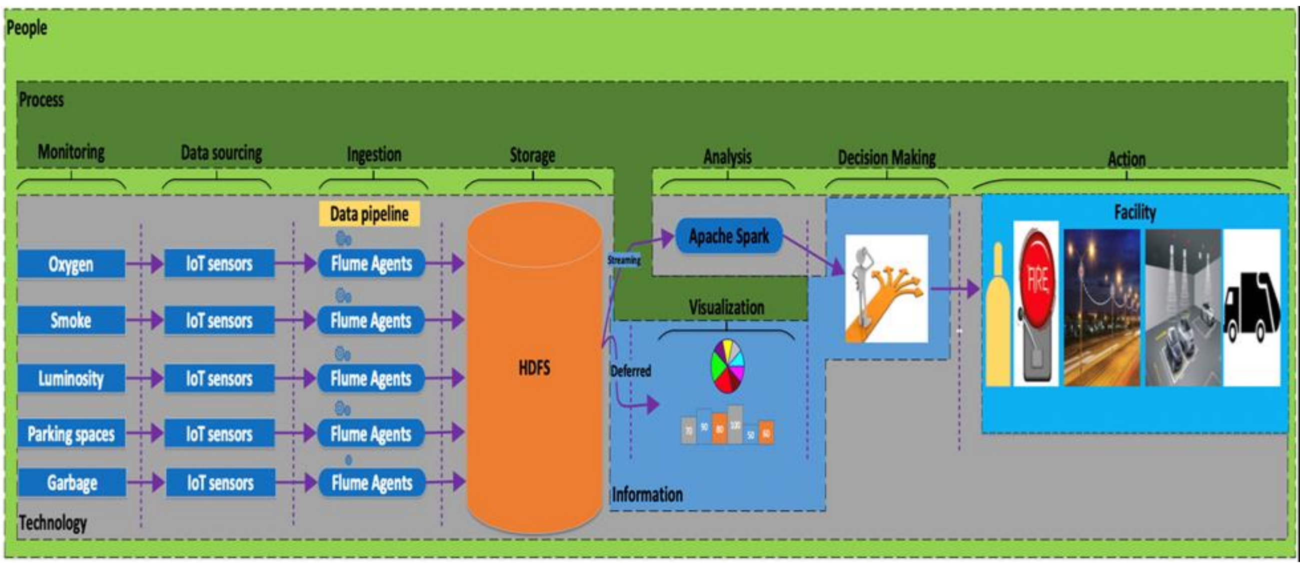


Iteration 3 – Real-time data management, analysis and actuation for smart building

Iteration 4:

In iteration 4, we improved and extended the architecture to conceptualize the elements in terms of people, process, technology, information and facility to link the conceptual elements (Figure 4.2) with the physical layer components. The architecture for iteration 4 is shown in Figure 4.11. The architecture developed in iteration 3 was scaled up and tested for a smart building application scenario by considering 1000 virtual IoT sensors. For this iteration, we considered 200 of each of the five types of IoT sensors, namely oxygen sensors, smoke detection sensors, light sensors, parking space sensors and garbage detection sensors. Ten Flume agents were configured with IoT sensor data as the source and HDFS as the sink. The data were then visualized in Tableau. Apache Spark was used to analyse the data in near real-time as it gets stored in HDFS. Based on the algorithm developed in PySpark, various messages were printed on the terminal screen simulating the feedback actuation behaviour.

For oxygen sensors, if the value sent by a sensor is below a threshold, the PySpark algorithm prints out a message on the terminal stating the HVAC system associated with that particular oxygen sensor has been turned ON. For smoke detection sensors, if the value generated by a sensor exceeds a threshold (i.e., occurrence of a fire), the PySpark algorithm detects that and outputs a message on the terminal stating that the fire alarm connected at the location of that particular smoke detection sensor is turned ON. In the case of luminosity sensors, if a particular luminosity sensor generates an output value below a threshold indicating it is dark, the PySpark algorithm outputs a message on the terminal stating that the lights associated with that particular luminosity sensor are turned ON. For the parking space sensors, if the value generated by a particular parking space sensor is a 1, the PySpark algorithm displays a message on the terminal stating that a car has been parked at that particular parking spot. For the garbage detection sensors, if the value generated by a particular sensor is above the threshold, the PySpark algorithm displays a message on the terminal stating that the garbage bin associated with the particular sensor which generated an above threshold value is full.



Iteration 4 – Updated real-time data analysis and actuation architecture

Iteration 5:

The earlier iterations had limitations in terms of real-time data visualization and performing predictive analytics. In the fifth and final iteration, we worked on improving the architecture to enable real-time visualizations by introducing Elasticsearch (*Elasticsearch* 2018; Adeli & Vishnubhotla 2020) and Kibana (*Kibana* 2018; Agarwal et al. 2010). We also introduced MS (Microsoft) Power BI (*Power BI* 2018; Akkaya et al. 2015) for the visualization of data stored in HDFS. The main reason for introducing MS Power BI was because Power BI integrates well with R scripts. This integration provides the ability to do data analysis and predictive analytics within Power BI in an interactive way. Moreover, we chose a hybrid model considering both batched as well as streaming data sources. We chose open data as a batched data source for our architecture. This open data is scrapped using R (can also be done using Python) and ingested into HDFS.

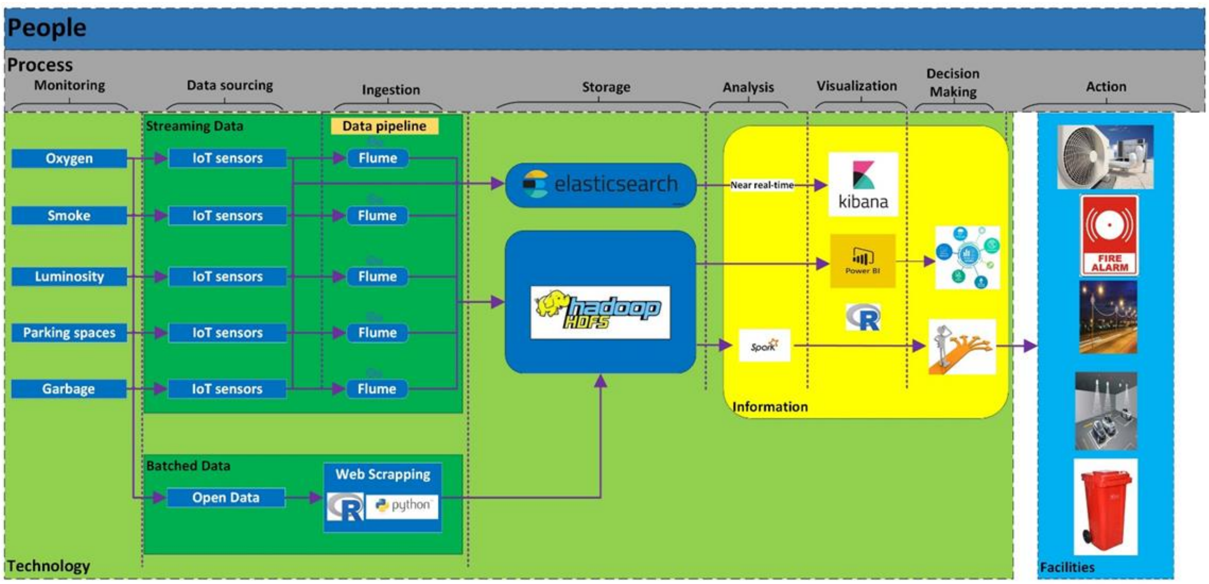
For streaming data sources, the virtual IoT sensors send the data to two sinks: 1) HDFS and 2) Elasticsearch. As the data from the IoT sensors land into HDFS, it is analysed in near real-time by the Apache Spark algorithm to enable decision making for the effective management and control of the five facilities earlier described within the smart building. This data once stored in HDFS is visualized in batches using Power BI. For predictive analytics, we used R scripts to develop an ARIMA model within Power BI. For the second source i.e., Elasticsearch, the data is indexed as it lands into Elasticsearch. Elasticsearch provides a data visualization plug-in called Kibana which enables the near-real time visualization of IoT data. The updated architecture is presented in Figure 4.12. It shows how the five conceptual elements shown in Figure 4.2 are related to the physical elements of the framework. People are at the top-most level, which represent the stakeholders of the smart building environment such as building developers, building management, IT professionals and residents of the building. Process element defines data-driven processes, which are relevant to the smart building. This includes monitoring via sensors, data sourcing, ingesting, storing, analysis, visualization, decision making and finally actuation. The technology element includes the technology stack including Flume, R, Elasticsearch, HDFS, Kibana, Spark and Power BI. The information includes the near real-time data visualizations in Kibana, Power BI dashboards for the IoT data visualization and the output of the decision-making process using Spark. Finally, the facility element represents the facilities in the smart building including HVAC systems, fire alarms, lights, parking spaces and garbage spaces.

Table 4.1 summarizes the details of various big data tools used in the development of the IBDMA. It lists the processes in which each of these tools are used and the purpose of each of these tools in the IBDMA reference architecture implementation.

Elements in the IBDMA architecture and their purpose

Sr. No	Element	Process	Purpose
1	Flume	Ingestion	For ingesting streaming IoT data in Elasticsearch and HDFS.

2	Elasticsearch	Storage	Indexing streaming data to be visualized in Kibana.
3	Kibana	Visualization	Visualizing streaming IoT data in near real-time.
4	HDFS	Storage	Storing both batched and streaming data.
5	Spark	Analysis/Decision Making	Analysing IoT data in near real-time to enable decision making and actuation of smart facility.
6	Power BI	Visualization	Visualizing batched and forecasted data.
7	R	Ingestion/Decision Making	Web scrapping and predictive analytics.



Iteration 5 – Updated and improved near real-time data analysis and actuation architecture

Appendix J: IBDMA reference architecture evaluation:

Iteration 1 use case and evaluation

The details of use case to evaluate iteration 1 are:

Test Case 01 - Detection of oxygen levels

Test Case 01	**Detection of low oxygen level in smart building.**
Context	Oxygen levels fall below human comfort levels at a specified location in the building.
Problem	Low oxygen levels in the smart building may remain unnoticed by smart building management which could result in the discomfort of the residents and could potentially prove fatal.
Description	Oxygen levels in the smart building falls below a certain threshold, the oxygen sensor senses this. Value is recorded and is reflected into the web system. Data is scrapped from web using R.

The evaluation results for iteration 1 (Figure 4.8) against the thirteen evaluation criteria (as presented in Figure 3.8) are shown in Table below.

Evaluation results for evaluating each test case against the thirteen ECs

Test Case	**Evaluation Criteria**	**Pass✓/Fail✗**
01		
	Efficacy (EC1)	✗
	Validity (EC2)	✗
	Generality (EC3)	✓
	Consistency with people (EC4)	✓
	Consistency with technology (EC5)	✗
	Completeness (EC6)	✗
	Simplicity (EC7)	✓
	Clarity (EC8)	✗
	Accuracy (EC9)	✗
	Performance (EC10)	✗

	Efficiency (EC11)	✗
	Robustness (EC12)	✗
	Scalability (EC13)	✗

Iteration 2 use case and evaluation

Use case:

The details of use case to evaluate iteration 2 are:

Test Case 02 - Detection of smoke levels

Test Case 02	**Detection of temperature and smoke in smart building.**
Context	Temperature in the smart building rises and smoke is detected indicating a fire scenario in a specific location in the smart building.
Problem	Smoke and fire in the smart building may remain unnoticed by the smart building management that could result in a serious incident and may prove fatal.
Description	Temperature and smoke levels in the smart building rise above the threshold, the temperature and smoke detection sensors sense this. Values are analysed and stored which are used for actuating the fire extinguisher at the location where fire is detected.

The evaluation results for iteration 2 (Figure 4.9) against the thirteen evaluation criteria (as presented in Figure 3.8) are shown in Table below.

Evaluation results for evaluating each test case against the thirteen ECs

Test Case	**Evaluation Criteria**	**Pass✓/Fail✗**
02		
	Efficacy (EC1)	✗
	Validity (EC2)	✓
	Generality (EC3)	✓
	Consistency with people (EC4)	✓
	Consistency with technology (EC5)	✗
	Completeness (EC6)	✗
	Simplicity (EC7)	✓

	Clarity (EC8)	✓
	Accuracy (EC9)	✗
	Performance (EC10)	✗
	Efficiency (EC11)	✗
	Robustness (EC12)	✗
	Scalability (EC13)	✗

Iteration 3 use case and evaluation

The details of use case to evaluate iteration 3 are:

Test Case 03 - Detection of oxygen levels

Test Case 03	**Detection of low oxygen level in smart building.**
Context	Oxygen levels fall below human comfort levels at a specified location in the building.
Problem	Low oxygen levels in the smart building may remain unnoticed by smart building management which could result in the discomfort of the residents and could potentially prove fatal.
Description	The oxygen levels in smart building falls below a certain threshold, the oxygen sensor senses this. Value is analysed in near-real time and the HVAC system is activated.

The evaluation results for iteration 3 (Figure 4.10) against the thirteen evaluation criteria (as presented in Figure 3.8) are shown in Table below.

Evaluation results for evaluating each test case against the thirteen ECs

Test Case	**Evaluation Criteria**	**Pass✓/Fail✗**
03		
	Efficacy (EC1)	✗
	Validity (EC2)	✓
	Generality (EC3)	✓
	Consistency with people (EC4)	✓
	Consistency with technology (EC5)	✗

	Completeness (EC6)	✗
	Simplicity (EC7)	✓
	Clarity (EC8)	✓
	Accuracy (EC9)	✗
	Performance (EC10)	✗
	Efficiency (EC11)	✗
	Robustness (EC12)	✗
	Scalability (EC13)	✗

Iteration 4 use case and evaluation

The details of use case to evaluate iteration 4 are:

Test Case 04 - Detection of oxygen levels

Test Case 03	**Detection of low oxygen level in smart building.**
Context	Oxygen levels fall below human comfort levels at a specified location in the building.
Problem	Low oxygen levels in the smart building may remain unnoticed by smart building management that could result in the discomfort of the residents and could potentially prove fatal.
Description	The oxygen levels in the smart building fall below a certain threshold, the oxygen sensor senses this. The value is analysed in near-real time and the HVAC system is activated. The historical data can be visualised.

The evaluation results for iteration 4 (Figure 4.11) against the thirteen evaluation criteria (as presented in Figure 3.8) are shown in Table below.

Evaluation results for evaluating each test case against the thirteen ECs

Test Case	**Evaluation Criteria**	**Pass✓/Fail✗**
04		
	Efficacy (EC1)	✓
	Validity (EC2)	✓
	Generality (EC3)	✓
	Consistency with people (EC4)	✓

	Consistency with technology (EC5)	✗
	Completeness (EC6)	✗
	Simplicity (EC7)	✓
	Clarity (EC8)	✓
	Accuracy (EC9)	✓
	Performance (EC10)	✗
	Efficiency (EC11)	✗
	Robustness (EC12)	✗
	Scalability (EC13)	✗