DATA WAREHOUSE DEVELOPMENT FRAMEWORK WITH DATA QUALITY CONSIDERATION

MUNAWAR

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Computer Science)

Faculty of Computing
Universiti Teknologi Malaysia

SEPTEMBER 2016

UNIVERSITI TEKNOLOGI MALAYSIA

DECLARATION OF THESIS AND COPYRIGHT

Author's full name : **MUNAWAR**

Date of birth : **24 JUNE 1969**

Title : **DATA WAREHOUSE DEVELOPMENT FRAMEWORK WITH DATA QUALITY CONSIDERATION**

Academic Session : **2015/2016-2**

I declare that this thesis is classified as:

☐	**CONFIDENTIAL**	(Contains confidential information under the Official Secret Act 1972)*
☐	**RESTRICTED**	(Contains restricted information as specified by the organization where research was done)*
☑	**OPEN ACCESS**	I agree that my thesis to be published as online open access (full text)

I acknowledged that Universiti Teknologi Malaysia reserves the right as follows:

1. The thesis is the property of Universiti Teknologi Malaysia.
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
3. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

SIGNATURE

200907M10165
(NEW IC NO. /PASSPORT NO.)

Date: 11/9/2016

SIGNATURE OF SUPERVISOR

PROF. DR. NAOMIE SALIM
NAME OF SUPERVISOR

Date: 13/9/2016

NOTES : * If the thesis is CONFIDENTAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction.

We hereby declare that we have read this thesis and in our opinion this thesis is sufficient in terms of scope and quality for the award of the degree of Doctor of Philosophy (Computer Science)

Signature : .....

Name of Supervisor I : Prof. Dr. Naomie Salim

Date : 13/9/2016

Signature : .....

Name of Supervisor II : Dr. Roliana Ibrahim

Date : 11/9/2016

BAHAGIAN A – Pengesahan Kerjasama*

Adalah disahkan bahawa projek penyelidikan tesis ini telah dilaksanakan melalui kerjasama antara _____ dengan _____

Disahkan oleh:

Tandatangan : ..... Tarikh : .....

Nama : .....

Jawatan : .....

(Cop rasmi)

* *Jika penyediaan tesis/projek melibatkan kerjasama.*

BAHAGIAN B – Untuk Kegunaan Pejabat Sekolah Pengajian Siswazah

Tesis ini telah diperiksa dan diakui oleh:

Nama dan Alamat Pemeriksa Luar : **Prof. Dr. Abdullah bin Embong**
Institute of Postgraduate Studies,
Universiti Malaysia Pahang,
Lebuhraya Tun Razak,
26600 Gambang, Kuantan,
Pahang.

Nama dan Alamat Pemeriksa Dalam 1 : **Assoc. Prof. Dr. Azizah binti Abdul Rahman**
Faculty of Computing,
UTM Johor Bahru.

Disahkan oleh Timbalan Pendaftar di Sekolah Pengajian Siswazah:

Tandatangan : ---------- Tarikh : ----------

Nama : **ASRAM BIN SULAIMAN @ SAIM**

I declare that this thesis entitled "*Data Warehouse Development Framework with Data Quality Consideration*" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : 

Nama : Munawar

Date : 11/9/2016

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor, Professor Dr. Naomie Salim, for encouragement, guidance, and critics. I am also very thankful to my co-supervisor Dr. Roliana Ibrahim for her guidance, advices and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to my colleges Firman, Nuril and Cahyanto for sharing knowledge and experience in data warehouse development. Their views and tips are useful indeed.

My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Unfortunately, it is not possible to list all of them in this limited space

I also owe special thanks to my wife Mira and all my children (Hanifah, Rois and Safiya) who were very patient with the author who always seemed to be working. Very special thanks to my mother, for her pray and support.

.

ABSTRACT

Data quality issues can emerge at any stage of data warehouse development. Rapid growth in data volumes has given rise to new problems such as data quality, which is a critical issue when data is transferred from one system to another. Lack of data quality provided by data warehouse can lead to bad strategic decisions and indicates a significant failure rate. Research on data warehouse development and data quality lacks clearly defined standards that are essential for data warehouse development applicable for various types of organizations. This study formulated a framework for incorporating data quality into data warehouse development to improve data warehouse quality. Data from five organisations was collected through qualitative and quantitative methods. Quadrant analysis was used in the analysis of quantitative data. In addition, confirmation from experts helped in determining specific data quality dimensions that can be correlated with data warehouse development. From these findings, a framework was developed for incorporating data quality into data warehouse development. Five stages of data warehouse development comprising requirements analysis, conceptual design, logical design, extract-transform-load and physical design were identified. Sixteen data quality dimensions, namely comprehensiveness, accuracy, clarity, applicability, conciseness, consistency, correctness, currency, convenience, timeliness, traceability, interactivity, accessibility, security, maintainability and speed were incorporated into the five stages. The framework was evaluated through the implementation of a real case. There was a positive correlation between data quality incorporation into data warehouse development and data warehouse benefits. The framework has shown that incorporating data quality into the entire data warehouse development stages can increase the benefits of data warehouse.

ABSTRAK

Isu kualiti data dapat dilihat di semua peringkat pembangunan gudang data. Pertumbuhan pesat dalam jumlah data telah menimbulkan masalah baru seperti kualiti data, yang merupakan isu kritikal apabila data dipindahkan dari satu sistem ke sistem yang lain. Kekurangan kualiti data yang disediakan oleh gudang data boleh membawa kepada keputusan strategik yang kurang baik dan menunjukkan kadar kegagalan yang ketara. Penyelidikan pembangunan gudang data dan kekurangan kualiti data yang tidak mempunyai piawai yang jelas amat penting untuk pembangunan gudang data di dalam pelbagai jenis organisasi. Kajian ini merumuskan satu rangka kerja untuk menggabungkan kualiti data ke dalam pembangunan gudang data bagi meningkatkan kualiti gudang data. Data daripada lima organisasi telah dikumpulkan melalui kaedah kualitatif dan kuantitatif. Analisis kuadran digunakan dalam analisis data kuantitatif. Di samping itu, pengesahan daripada pakar membantu dalam menentukan dimensi kualiti data tertentu yang boleh dikaitkan dengan pembangunan gudang data. Daripada penemuan ini, sebuah rangka kerja telah dibangunkan untuk menggabungkan kualiti data ke dalam pembangunan gudang data. Lima peringkat pembangunan gudang data yang terdiri daripada analisis keperluan, reka bentuk konsep, reka bentuk logik, ekstrak-mengubah-beban dan reka bentuk fizikal telah dikenal pasti. Enam belas dimensi kualiti data iaitu komprehensif, ketepatan, kejelasan, kebolehgunaan, keringkasan, ketekalan, ketepatan, tidak usang, kemudahan, ketepatan masa, kebolehkesanan, interaktiviti, capaian, keselamatan, penyelenggaraan dan kelajuan telah dimasukkan ke dalam lima peringkat tersebut. Rangka kerja ini telah dinilai melalui perlaksanaan kes yang tulen. Terdapat korelasi positif antara penggabungan kualiti data ke dalam pembangunan gudang data dengan manfaat gudang data. Rangka kerja ini telah menunjukkan bahawa menggabungkan kualiti data ke dalam keseluruhan pembangunan gudang data boleh meningkatkan manfaat gudang data.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Introduction

Organisations are currently highly dependent on data, and the rapid growth in data volumes has given rise to new problems for institutions. Foremost of these problems is data quality (Ranjit and Kawaljeet, 2010), which is a critical issue when data are transferred from one system to another. Data quality (DQ) is crucial to effective business performance; when handled inappropriately, it may be a major problem in many information system (IS) projects, including data warehouse (DW) initiatives.

DQ issues are highly essential matters for consideration in DW projects (English, 1999). Most of the time, however, DW developers neglect the effects of low-quality data on such initiatives (Kimball et al, 2008). Many organizations are often aware of DQ problems, but their improvement efforts are generally focused on only data accuracy; many other equally important DQ attributes and dimensions are disregarded such as correctness, consistency etc (Prakash et al, 2004).

A DW is a special database that stores large amounts of data that are collected from heterogeneous sources to satisfy the information demands of decision makers (Marotta and Ruggia, 2002). Many organisations in numerous sectors use DWs as a foundation for decision-making infrastructure and as critical enablers of current strategic initiatives, such as customer relationship management (CRM),

business performance management and supply chain management (Cooper et al, 2000; Eckerson, 2004; Goodhue et al, 2002; Watson et al, 2001). The advantages that DWs present encourage their increasing use in many organisations.

DWs are challenging projects because of the myriad technical and organisational issues that may be encountered. The fundamental goal in building a DW is to improve the quality of information in an organisation (Watson and Haley, 1997) for decision making process. The greatest benefits of DWs are obtained when they are used to improve business processes, support decisions and strategic business objectives and provide excellent information and time savings for users (Watson et al, 2002).

Although DQ in DWs is essential, but there is no agreement about the quality of data that are important for DW development. A consensus amongst researchers is that quality involves several dimensions (Ballou and Tayi, 1999; Cowie and Burstein, 2007; Jarke et al, 1999; Pipino et al, 2002; Prakash et al, 2004), but no general agreement has been achieved as to which set of dimensions should be accorded priority.

Apart from objective perspectives, subjective viewpoints may also influence the manner by which DQ is approached because different users formulate their own quality criteria, which are based on needs that are specific to them. In such a case, therefore, ‘fitness for use’ (Wang and Strong , 1996) can serve as the best definition of DQ because this description underscores the relative use of data—data types that are appropriate for a given task or project may present sufficient attributes that are useful for another initiative. The concept of ‘fitness for use’ was widely adopted in the quality literature to emphasizes the importance of taking a consumer viewpoint of quality because ultimately it is the consumer who judge whether or not a product is fit for use. (Wei-Liang and Shi-Dong, 2009; Shariat Panahi, et al, 2013)

Quality is a key concept in DW development, in which all the stages are concerned with achieving excellence. Therefore, understanding the primary DQ

dimensions in every stage of DW development is critical. As far as researcher know, however, only a few studies have discussed the DQ dimensions that are essential to the entire DW development stages.

1.2. Research Background

Many studies have been devoted to DQ to solve the problems encountered in DW development. These studies have proposed various methods, amongst which the most widely used are techniques for modelling the DQ consideration in a particular stage of DW development. Despite the progress made in research, however, only a few studies have discussed techniques that are directed towards DQ integration into the entire DW development stages.

One of the simplest techniques for ensuring DQ is data inspection and rework, which involves data cleansing (English, 1999). This approach centres on maintaining data value and providing solutions to two DQ dimension-related problems: accuracy and consistency. Even as these techniques come with advantages, they are nonetheless unsuitable for frequent data modification. The more frequent the use of data cleansing algorithms, the higher the costs involved.

Other technique is process-oriented techniques (Wang, 1998; English, 1999; Redman, 2001; Scannapieco et al, 2002), in which the causes of data errors are identified through modifications to data access and updates activities. Some process-oriented methods are Total Data Quality Management (TDQM) and Total Information Quality Management (Helfert and von Maur, 2001). These techniques emphasise quality management in a static environment, without practical considerations in addressing dynamic changes in data resources and without detailed guidance to capturing such changes. Even though these methods are useful, they require elaboration before they can be applied, yet no specific procedure for customisation is provided. A complete set of metrics is also lacking, and only a few

algorithms have been developed for a subset of dimensions, such as accuracy, completeness, consistency and timeliness (Redman, 2001; Naumann et al, 2004).

Another technique for ensuring DQ is the use of metadata in addressing the DQ issues encountered in DW development. Under this category fall the techniques discussed in (Shankaranarayan, 2005; Fu-cheng et al, 2008; Helfert and Herrmann, 2002; Santana and Moura, 2004). (Metadata can be used to drive every processing units in DW and to trace back a data item step by step since its source system until its complete integration to the DW) These techniques enable decision makers to determine DQ in a context-dependent manner. Nevertheless, these methods cannot be used to evaluate the quality of data in a particular stage in DW development.

The above-mentioned techniques present certain benefits, but they do not provide a complete framework for managing DQ. The methods do not allow for DQ evaluation at a specific phase through measurements of the DQ associated with one or more preceding phases.

Good quality data ensures user's trust in data warehouse system making it more usable and, optimizes the business benefits gained (Kumar, and Thareja, 2013). Many authors denote that successful DW can be measured by the net benefits derived from DW (Hwang and Xu, 2008). However, detecting defects and improving data quality comes with a cost and if the targeted quality level is high, the costs often negate (offset) the benefits. This study aims at providing empirical evidence of benefits derived from DW, thereby extending the body of research regarding the success of DW in term of DW benefits. Furthermore, this research investigates the correlation between DQ dimensions and DW benefits.

The most often cited benefit of DW construction is improvement in information quality for decision making process (Popovic et al, 2009) through improvement of information processes (Jaklic et al, 2009). In today's era, information plays a critical role in the success of organisations. Any information acquired by decision makers must be in good quality to support decision making. In

decision making, information reduces uncertainty, enables organisations to rapidly react to business events, and support organisations in making changes in corporate strategies, plans and performance indicators (Jaklic et al, 2011).

1.3. Research Motivation

More and more organisations have been building their own DWs. Despite high failure rates, spending on DW projects grow by 43% annually (Winter and Strauch, 2003) and is expected to considerably rise in succeeding years (Agosta, 2004) because of the dramatic drop in storage costs.

The development and maintenance of DWs are highly expensive and time consuming. Generally, a full-scale DW application/system costs more than US$1 million (Hwang at al, 2004). A total annual DW budget can range from US$2 to US$10 million, which includes provisions for a support staff comprising 20 to 50 people; the average first-year cost of a DW project is about US$1.26 million (Amin and Arefin, 2010; Schiefer at al, 2002). The average time for the development of a DW is 12 to 36 months (Amin and Arefin, 2010)

Several surveys indicate that a significant percentage of DWs fail to satisfy expectations or are outright failures. Failure rates vary, but these usually average from 20% to 50% (Agosta, 2004; Conner, 2003; Watson et al, 2001). The failure can be traced to a single cause: the absence of quality (Cowie and Burstein, 2007). These statistics highlight the necessity of developing methods designed to determine the DQ dimensions that are critical to DW development.

Poor DQ costs US businesses more than US$600 billion (Eckerson, 2002) and consumes approximately 10% of a typical organisation’s revenue (Redman, 2001). Because of poor DQ, increasing operating costs and unfavourable effects on customer satisfaction and on the community at large (Nemoni and Konda, 2009)

become obstacles to determining appropriate management decisions (English, 1999); these consequent problems also diminish data value and the ability of executives to align organisational goals and objectives (Redman, 2001). Furthermore, poor DQ gives rise to the need for information scrap and rework measures, which lead to resource wastage in terms of manpower, finances, materials and facilities (Pighin and Leronutti, 2008); it causes organisational inefficiency and capital losses (Redman, 2001) and is a leading cause of failure rates (Haug et al, 2011, Chenoweth et al, 2006; Hayen et al, 2007; Johnson, 2004; Ramamurthy et al, 2008).

The quality of DWs should be guaranteed beginning at the early phase of a DW project (Rizzi et al, 2006). The earlier the problems are identified, the sooner specific recommendations can be made to ensure that data quality are correctly incorporated (Giblett, 2002). Waiting until the testing stage to incorporate DQ measures means no time would be devoted to addressing the DQ issues that arise across all the phases of DW development. Consequently, organisations cannot guarantee the timely delivery of DW projects (Giblett, 2002), which in turn, results in a high failure rate (Loshin, 2008). Hence, incorporating DQ into the entire DW development stages determines the success of DW projects (Ballou and Tayi, 1999).

1.4. Problem Statement

Awareness of data and information quality issues has grown rapidly in this information era. Even though there has been no consensus about the distinction between data quality and information quality (Zhu, et al, 2014), but many approaches do not distinguish between data and information and define data quality and information quality equal (Jaklic et al., 2009; Ranjit and Kawaljeet, 2010). Therefore, data quality and information quality in this study are treated equal.

Eventhough most researchers agree that DW development phases encompass the following tasks: requirements analysis, conceptual design, logical design and

physical design (Inmon, 2005; Kimball et al, 2008; Rizzi, 2009), little empirical research has been conducted so far. Such claims need to be tested empirically. It is essential, therefore, that researcher investigate the common practiced DW development stages that is applicable to all types of organizations particularly in Indonesia.

DQ issues can emerge at any stage of DW development (Kumar and Thareja, 2013). Data are influenced by various processes for converting data into DW data; to a certain extent, therefore, these processes also affect DQ. However, disagreement remains over which DQ dimensions should be maximised for DW development. Given this backdrop, finding DQ dimensions in the entire DW development stages is necessary to guarantee that high quality levels are maintained (Nemoni and Konda, 2009).

Maintaining high-quality data in a DW is a key success factor for IT data warehousing professionals (Loshin, 2008) because inaccurate data significantly diminish the strategic value of a DW (Evans, 2005). Better informed, more reliable decisions come from high quality of data during the process of loading a data warehouse (Verma et al, 2014). Lack of DQ provided by DW can lead to bad strategic decisions. Thus, DQ in DW needs to be assured (Gosain and Heena, 2015). Improving DQ in DW is important because it is used in the process decision support which requires accurate data (Akbar et al., 2013).

System quality and service quality are general issues from the perspective of IS success through DeLone and McLean IS success model (Shin, 2003), to affect both use and user satisfaction. However, there is no much research done to the DeLone and McLean model in DW context (AlMabhouh and Ahmad, 2010). The IS success model on DW, the original context cannot fits nicely into the DW environment. User satisfaction as an important part of DeLone and McLean model is not considered a good indicator of success for multiple-user applications such as DW (Wixom and Watson, 2001; Lu and Wang, 1997). In addition, it is unlikely that

organisations would spend millions of dollars on DW projects just to make their users happy (AlMabhouh and Ahmad, 2010).

One of the most costly and time-consuming tasks to be done in DW projects is to ensure high DQ (Nemoni and Konda, 2009). Lack of DQ cause a significant failure rate of DW (Haug et al, 2011). Therefore, DQ consideration in all the stages of DW development has a great impact on overall quality of DW (Nemoni and Konda, 2009) which has a crucial effect on the managerial strategic decision to be made (Gosain and Heena 2015).

There is no academic research that defines the quality of DW or the success of DW (Amornbuth, 2015), because of difficult to measure (Bilal Ali, 2014; Hwang and Xu, 2008). Nonetheless, many authors claim that success of DW can be measured by DW benefits obtained (Hwang and Xu, 2008; Mukherjee, 2003; Wixom and Watson, 2001) through DQ consideration in DW development (Kumar, and Thareja, 2013).

Relatively few studies have been conducted on the integration of DQ into the development of DW. The DWQ technique for incorporating DQ to DW design and usage (Jarke et al, 1999), DQ consideration in non-functional requirements (Paim and Castro, 2003), DQ-based procedure for eliciting functional (queries) and non-functional requirements (Vaisman, 2007), meta data based quality model to enforce quality in DW (Kumar and Thareja, 2013) and automated ETL testing on DQ of DW (Dakrory et al, 2015) are a few examples of DQ consideration in DW development. The above-mentioned studies have provided valuable knowledge on DQ and DW issues, but a drawback to these is that they analysed DQ dimensions in a single phase of DW development, thus integrating all DQ consideration in all phases of DW development to be a single framework is difficult. As far as researcher know, no study has developed or proposed a method for DQ consideration in the entire DW development stages. The contribution of the current research is its consideration of DQ in the entire DW development stages.

Although DQ issues have a direct economic and social impact (Wang and Strong , 1996) yet little work is done for formulating a framework for incorporating DQ in DW development (Kumar, and Thareja, 2013). Effective DQ management can be achieved only with a framework characterised by DQ consideration into the framework for identification of anomalies in data management (Nemoni and Konda, 2009).

1.5. Research Problems and Objectives

DW development is considerably more costly than IS projects, and failure rates in the former are high. This study contributes to the literature in the following ways: It provides insights into how corporations or organisations can determine the critical factors that affect the quality of DW projects. It is a timely initiative because research is rarely conducted on the DQ aspects of DW development. It is hoped that the findings of the current work will complement or supplement the results of previous investigations into this matter.

This study features a detailed analysis designed to determine specific DQ dimensions that may be correlated with DW development for the purpose of increasing DW benefits gained. The research problem undertaken in this study can be defined in the question, ‘How can organisations build a framework for the design and development process of DW projects to effectively incorporate DQ issues and accordingly improve data warehouse quality? The aim articulated in the research question is achieved by fulfilling the following objectives:

Objective 1: To investigate the common practices of DW development stages in some organizations;

Objective 2: To identify DQ dimensions related to DW development stages that is applicable to all types of organization;

Objective 3: To identify the correlation between DQ dimensions and DW benefits;

Objective 4: To develop a DQ-oriented framework for the design and development process of DWs to improve DW quality.

1.6. Scope of the Study

According to Rudra and Yeo (2000) there is a scarcity of empirical studies that examine the DW success. As far as researcher know, there is not a shred of empirical evidence that the DW has been investigated or showed the degree of DW development in Indonesia. Therefore, this study investigates the extent to which Indonesian firms developed DW and its benefits. It is aimed at providing empirical evidence, thereby extending the body of research regarding the development of DW.

One of the most important ideas in this study is the unit of analysis. The unit of analysis is the major entity that should be analysed to refer to the hierarchy of aggregation (Davidsson and Wiklund, 2001). Inappropriate unit of analysis may lead to results that are erroneous conclusion (Silverman and Solmon, 1998). Based on primary research questions posed and analysed where data is collected, unit of analysis in this study is organisations.

The multi case studies (five cases) approach was adopted for the current research. The multi case studies strengthen the results by replicating the pattern-matching and eliminate chance associations (Eisenhardt, 1991), thus increasing confidence in the robustness of the theory (Tellis, 1997; Herriott and Firestone, 1983). The multi case studies also increase the external validity and reliability of the research compared to a single-case study (Yin, 2003). The multi case studies also expand the generalizability of the findings compared to a single case alone.

Eventhough Eppler's information quality is popular, there is no much research done to test it in DW context. Jaklic et al (2011) used Eppler's information quality to show information quality impact on different uses of information in DW

especially in process driven stage. However, as far as researcher know, there is no research done to test Eppler's information quality in the entire DW development stages.

As indicated in previous section, the development and maintenance of DWs are highly expensive and time consuming. Only medium and upper-sized organisation are able to build full scale of DW. Hence, the proposed framework were implemented to one organization only.

1.7. Significance of Study and Expected Contributions

As previously discussed, building a DW is no small endeavor, considerable resources from the corporation in term both of time and money. Due to the strategic importance of DWs, it is absolutely crucial to guarantee their DQ from the early stages of a DW project (Rizzi et al, 2006). These DQ aspects should be identified at various level of DW development (Gosain and Singh, 2008) to guarantee high quality levels (Nemoni and Konda, 2009). For this reasons, this study attempts to distinguish all DQ dimensions in the entire DW development stage.

Despite incredible growth, a large number of DW initiatives end up as failures. Over 50% of DW projects would experience limited acceptance, if not outright failure because of poor DQ (Friedman, 2004). Poor DQ may have a significant negative impact on the efficiency of an organization such as reduce the organization's revenue between 8 – 12 % (Haug, 2011). Accordingly, this study investigates correlation between DQ dimensions in DW and DW success as indicated in its benefits gained.

For an effective DW system, the quality aspects should be incorporated properly at various levels of DW development (Gosain and Singh, 2008) and should cover all of DW development layer (Munos et al, 2010). Therefore, the use of a

framework for managing DQ in DW is very much needed. Within a framework, the DQ at one stage can be ensured as a result of the DQ measurements conducted in previous stages (Nemoni and Konda, 2009). This study attempts to compare effect of DQ consideration to DW quality using the proposed framework (as represented by the university) and without the proposed framework (as represented by the case studies in five organizations)

1.8. Thesis Structure

The thesis is organised as follows.

Chapter 1 presents the introduction, which highlights the importance of DQ in DW development, research background and scope of study. It then presents the research problems and objectives , that is, to explore the need to develop a framework that integrates DQ into the entire DW development.

Chapter 2 is the literature review, which details existing techniques for DW development and ensuring DQ in DW initiatives, as well as the benefits that organisations can derive from building DWs.

Chapter 3 explains the methodology adopted in solving DQ issues in DW development, including the research model, data collection methods and justification for adopting the case study.

Chapter 4 presents an overview of the organisations treated as case studies and the approaches common to the different DW development cases. The results of a survey on DW development and the findings and discussion of recent literature were used as a foundation for determining common practices of DW development stages that is applicable to all types of organisations.

Chapter 5 presents the commonly encountered DQ dimensions in DW development and explores the benefits of DW implementation, as determined from the case studies. The results of a survey on treating DQ dimensions in the entire DW development stages and the findings and discussions were used as bases in comprehensively examining the DQ dimensions commonly encountered in the entire DW development.

Chapter 6 presents confirmation of DW benefits from reviewed literature in five case studies and three consultants. Identifications of relationship amongst DQ dimensions in DW development stages and DW benefits gained is also discussed in this chapter.

Chapter 7 defines the framework for integrating commonly adopted DW development practices and DQ dimensions in the entire DW initiatives. This framework is designed to identity the benefits of DW implementation in the studied organisations.

Chapter 8 summarises the study's main findings and contributions, provides the concluding remarks and highlights future research directions.