Investigating Pluralistic Data Architectures in Data Warehousing

A thesis submitted for the degree of

Doctor of Philosophy

By

Kazeem Oladele

College of Engineering, Design and Physical Sciences

Brunel University

October 2014

ABSTRACT

Understanding and managing change is a strategic objective for many organisations to successfully compete in a market place; as a result, organisations are leveraging their data asset and implementing data warehouses to gain business intelligence necessary to improve their businesses. Data warehouses are expensive initiatives, one-half to two-thirds of most data warehousing efforts end in failure. In the absence of well-formalised design methodology in the industry and in the context of the debate on data architecture in data warehousing, this thesis examines why multidimensional and relational data models define the data architecture landscape in the industry. The study develops a number of propositions from the literature and empirical data to understand the factors impacting the choice of logical data model in data warehousing. Using a comparative case study method as the mean of collecting empirical data from the case organisations, the research proposes a conceptual model for logical data model adoption. The model provides a framework that guides decision making for adopting a logical data model for a data warehouse. The research conceptual model identifies the characteristics of business requirements and decision pathways for multidimensional and relational data warehouses. The conceptual model adds value by identifying the business requirements which a multidimensional and relational logical data model is empirically applicable.

ACKNOWLEDGEMENTS

I would like to extend my sincere appreciation to everyone that have helped in making this thesis possible

- First, I would like to thank Professor Mark Lycett for his relentless advice, critique and support of my work during the course of this study. I'm particularly grateful to Professor Lycett for his professionalism, scholarly guidance and those stimulating thoughts, I am immensely grateful to have him as my supervisor.

- Second, I would like to thank my second supervisor Dr. Sergio de Cesare for review and critique of my work, I am immensely grateful for his kindness and conversations while at St Johns.

- Third, I would like to thank Dr. Thomas Ilvento, Department of Applied Economics and Statistics, University of Delaware for reviewing my work and offering useful advice and tips in our discussion sessions.

- I would like to thank my kids, Ella and Louis for been so understanding and being there for Dad through the course of this study.

- Finally, I would like to thank the teams at GWealth and ICapital for taking time out of their busy schedule to take part in this study.

PUBLICATION

The work in this thesis has led to the following publication:

Oladele, K. and Lycett, M. 2013, "*Investigating Alternating Data Architectures in Data Warehousing*", Proceedings of the 14^{th} International Conference on Informatics and Semiotics in Organisations, Stockholm, Sweden, March 25-27, 2013

Table of Contents

ABSTRACT .... **II**

ACKNOWLEDGEMENTS .... **III**

PUBLICATION .... **IV**

LIST OF FIGURES .... **XI**

LIST OF TABLES .... **XIII**

ACRONYMS .... **XV**

CHAPTER 1 - INTRODUCTION **13**

1.1 Background to the Problem 13

1.2 Aim and Objectives 15

1.3 Method of Research 16

1.4 Structure of the Thesis 17

CHAPTER 2 – LITERATURE REVIEW **20**

2.1 Introduction 20

2.2 Overview of a Data Warehouse System 21

2.3 The Key Components of a Data Warehouse System 23

2.4 Implementing a Data Warehouse 27

2.5 Decision Theories 33

2.5.1 Normative Decision Theory 35

2.5.2 Descriptive Decision Theory 37

2.5.3 Prescriptive Decision Theory 39

2.5.4 Naturalistic Decision Making 40

2.6 Proposed factors influencing the choice of logical data model in data warehousing 43

2.6.1 High performance of query execution is a factor influencing the choice of logical data model for a data warehouse 44

2.6.2 The degree of focus on specific or generic functionalities is a factor influencing the choice of logical data model for a data warehouse 45

2.6.3 The goal and the scope of a data warehouse is a factor influencing the choice of logical data model for a data warehouse 48

2.6.4 The implementation orientation of available resources is a factor influencing the choice of logical data model for a data warehouse 49
2.6.5 Enterprise Data consolidation to meet heterogeneous reporting and analytics requirements is a factor influencing the choice of logical data model for a data warehouse 50
2.6.6 Proposed Conceptual Model for Logical Data Model Selection 52
2.7 Summary 54
CHAPTER 3 - RESEARCH METHODOLOGY 55
3.1 Introduction 55
3.2 Overview of Research Methodologies 56
3.2.1 Research Philosophy – The Positivist Paradigm 58
3.2.2 Research Philosophy – The Post Positivist Paradigm 60
3.3 Qualitative Research Methods 61
3.4 Reasons for using case study research method 64
3.5 Research Design Framework 66
3.5.1 Components of Research Design Framework 66
3.5.2 Research Topic, Aim and Objectives 67
3.5.3 Research Propositions and Theory Development 68
3.5.4 Unit of Analysis and Data Collection 69
3.5.5 Presentation of Results and Linking Data to Propositions 73
3.5.6 Transcribe Data 74
3.5.7 Apply Code to Text 75
3.5.8 Align Analytic Categories (Analytic Codes) with Propositions 76
3.5.9 Data Analysis 77
3.5.10 Description of Units of Analysis 78
3.6 Summary 79
CHAPTER 4 - CASE STUDY: GWEALTH 80
4.1 Introduction 80
4.2 The GWealth Management 81
4.2.1 Background of the Data Warehouse Project 82
4.2.2 Project Definition and Objectives 83
4.2.3 The Data Warehouse Architecture 84

4.3 Implementation of Research Design for GWealth 86
4.3.1 Transcription of GWealth Data and Applying Code to Text 87
4.3.2 Convert Descriptive Codes to Analytic Codes and Compare Codes for Saturation 89
4.3.3 Identify Sub Category Codes and Main Category Codes 91
4.3.4 Identify Properties of GWealth Category Codes 93
4.3.5 Analytic Codes and Frequency Count of Analytic Codes 94
4.4 Presentation of the Results – GWealth Case Study 95
4.4.1 Assessing the impact of high performance of query execution 98
4.4.2 Assessing the impact of specific business requirement 103
4.4.3 Assessing the impact of goal and scope of a data warehouse 108
4.4.4 Assessing the impact of implementation orientation of available 114
4.4.5 Assessing the impact of enterprise data consolidation 120
4.5 Summary of the Findings 123
4.5.1 Finding 1: *High Query Performance is a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse* 123
4.5.2 Finding 2: *Specific Requirement is a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse* 124
4.5.3 Finding 3: *Goal and Scope is a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse* 126
4.5.4 Finding 4: *Staff Experience is a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse* 127
4.5.5 Finding 5: *Enterprise Data Consolidation is Not a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse* 128
CHAPTER 5 - CASE STUDY: ICAPITAL........ 130
5.1 Introduction 130
5.2 Overview of ICapital 131
5.3 Background of the Data Warehouse Project 133
5.3.1 Project Definition and Objectives 135
5.3.2 The Data Warehouse Architecture 136
5.3.3 RFDW Data Sourcing Layer 137
5.3.4 RFDW Information Model 137
5.3.5 RFDW Event Manager 139

5.3.6 RFDW Client Interface 140
5.4 Implementation of Research Design for ICapital 140
5.4.1 Transcription of ICapital Data and Applying Code to Text 140
5.4.2 Convert Descriptive Codes to Analytic Codes and Compare Codes for Saturation.. 142
5.4.3 Identify Sub Category Codes and Main Category Codes 145
5.4.4 Identify Property of ICapital Category Codes 147
5.4.5 Analytic Codes and Frequency Count of Analytic Codes 148
5.5 Presentation of the Results - ICapital Case Study 149
5.5.1 Assessing the impact of heterogeneous reporting and analytics requirements........ 152
5.5.2 Assessing the impact of generic business requirements 155
5.5.3 Assessing impact of goal and scope of data warehouse 160
5.5.4 Assessing the impact of implementation orientation of available resources 165
5.5.5 Assessing the impact of high performance of query execution 169
5.6 Summary of the Findings........ 174
5.6.1 Finding 1: *Enterprise Data Consolidation is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse* 174
5.6.2 Finding 2: *Generic Requirements is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse* 175
5.6.3 Finding 3: *Goal and Scope is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse*........ 177
5.6.4 Finding 4: *Staff Experience is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse*........ 178
5.6.5 Finding 5: *High Query Performance is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse* 179

CHAPTER 6 - DISCUSSION OF THE CASE STUDIES........181
6.1 Introduction........ 181
6.2 Discussion of the Research Findings 182
6.2.1 High Performance of Query Execution........ 183
6.2.2 Focus on Specific and Generic Business Requirements 188
6.2.3 The Goal and Scope of the Data Warehouse 192
6.2.4 Implementation Orientation of Available Resources........ 196

6.2.5 Enterprise data consolidation for heterogeneous reporting and analytics requirements 200
6.3 Test of Relationship Between Analytic Codes and Logical Data Models 202
6.3.1 ACV1 and the Significant Relationship Test 208
6.3.2 ACV2 and the Significant Relationship Test 209
6.3.3 ACV3 and the Significant Relationship Test 210
6.3.4 ACV4 and the Significant Relationship Test 211
6.3.5 ACV5 and the Significant Relationship Test 212
6.4 Decision Matrix for Logical Data Model Adoption 213
6.5 The Revised Conceptual Model for Data Model Decision 217
6.5.1 Difference Between Proposed and Revised Conceptual Models 218
6.5.2 The Revised Conceptual Model 220
6.6 Implications for Practice 225
6.6.1 Requirements and Funding Considerations 226
6.6.2 Data Integration Considerations 228
6.7 Implications for Theory 229
6.8 Summary 231
CHAPTER 7 - CONCLUSION 233
7.1 Research Summary 233
7.2 Research Contributions 236
7.3 Research Limitations 238
7.4 Areas of Future Research 240
BIBLIOGRAPHY 241
APPENDIX 1 – DESIGN OF RESEARCH INSTRUMENTS 255
APPENDIX 2 – DESIGN OF RESEARCH INSTRUMENT FOR GWEALTH 257
APPENDIX 2 – DESIGN OF RESEARCH INSTRUMENT FOR ICAPITAL 263
APPENDIX 3 – GWEALTH OPEN CODES 269
APPENDIX 4 – ICAPITAL OPEN CODES 275
APPENDIX 5 –TEST OF SIGNIFICANT RELATIONSHIP 281

LIST OF FIGURES

Figure 2-1: End-to-end component of corporate data warehouse system (*Adapted from* Inmon, 2005; Inmon, Imhoff and Sousa, 2001; Sperley, 1999) 23
Figure 2-2: The Model of Decision Process (*Adapted from (Dillon, 1998)*) 37
Figure 2-3: Recognition-Primed Decision Making (*Adapted from Klien, G., 2008*) 42
Figure 2-4: Proposed Decision Model - Logical Data Model Selection Decision 53
Figure 3-1: Research Design Framework (*Author Proposed*) 67
Figure 3-2: Presentation of Results and Linking Empirical Data to Propositions (*Adapted from Graham R. Gibbs (2007)*) 74
Figure 3-3: Data Analysis (*Adapted from Graham R. Gibbs (2007)*) 77
Figure 4-1: End-to-End Architecture of GWealth Data Warehouse System 84
Figure 4-2: Client Reporting Conformed Dimension Data Architecture *(Adapted from GWealth architecture artifacts)* 86
Figure 4-3: Transcribed Empirical Data from GWealth Research Participant 87
Figure 4-4: Line-By-Line Coding of Transcribed Empirical Data 88
Figure 4-5: Transformation of Descriptive Codes to Analytic Codes 91
Figure 4-6: Sub-Category Code (*High Query Performance)* for the Analytic Codes 92
Figure 4-7: Main Category Code for High Query Performance Sub-Category Code 93
Figure 4-8: Property of Main Category Code 94
Figure 4-9: Frequency Occurrence of Analytic Code 95
Figure 4-10: Frequency Analysis of the Analytic Codes Using the Pareto Chart 97
Figure 4-11: Impact of Performance on Logical Data Model Decision 100
Figure 4-12: Impact of Specific Requirement on Logical Data Model Decision 105
Figure 4-13: Impact of Defined Goal and Scope on Logical Data Model Decision 110
Figure 4-14: Impact of Staff Experience on Logical Data Model Decision 116
Figure 4-15: Impact of Enterprise Data Consolidation on Logical Data Model Decision 121
Figure 5-1: Event Manager Dynamic Messaging Processing *(Adapted from ICapital Artifacts)* 139
Figure 5-2: Transcribed Empirical Data from ICapital Research Participant 141
Figure 5-3: Line-By-Line Coding of Transcribed Empirical Data 142
Figure 5-4: Transformation of Descriptive Codes to Analytic Codes 144
Figure 5-5: Sub-Category Codes (*Generic Requirements*, *Specific Requirement*) for the Analytic Codes 146
Figure 5-6: Main Category Codes for Generic Requirements Sub-Category Codes 147
Figure 5-7: Property of ICapital Category Codes 148
Figure 5-8: Frequency Occurrence of Analytic Code 149
Figure 5-9: Frequency Analysis of the Analytical Codes Using the Pareto Chart 151

Figure 5-10: Impact of Enterprise Data Consolidation for Multiple Reporting and Analytic Needs on Logical Data Model Decision 153
Figure 5-11: Impact of Generic Requirements on Logical Data Model Decision 157
Figure 5-12: Impact of Defined Goal and Scope on Logical Model Decision 162
Figure 5-13: Impact of Staff Experience on Logical Data Model Decision 167
Figure 5-14: Impact of Performance in Data Warehouse Logical Data Model Decision........... 171
Figure 6-1: Revised Conceptual Decision Model for Selecting Logical Data Model 219

LIST OF TABLES

Table 2-1: Differences between Data Warehouse and Operational Database (*Adapted from Todman, 2001; Inmon, 2005*).. 22
Table 2-2: The Key Architecture Components of Corporate Data Warehouse System (*Adapted from Sperley, 1999*).. 24
Table 3-1: Characteristics of Qualitative and Quantitative Research, Walliman (2006) 57
Table 3-2: Qualitative Research Methods in Social Science (*Adapted from Beinum, 1999; Gravetter et al, 2009; Czaja and Blair, 2005; Speziale, Streubert and Carpenter, 2007*)........... 63
Table 3-3: Outline of the Research Propositions 69
Table 3-4: Research Participants from GWealth and ICapital Case Organisations...................... 71
Table 3-5: Steps Linking Empirical Data to the Research Propositions...................................... 75
Table 3-6: Steps Linking Analytic Codes with the Research Propositions 76
Table 4-1: Descriptive Codes for Line-By-Line Coding 89
Table 4-2: Analytic Codes for Descriptive Codes 90
Table 4-3: Frequency Analysis of GWealth Analytic Codes *(See Appendix 3 for GWealth Open Codes)* 96
Table 4-4: Percentage Code Frequency of Query Performance vs. Other Analytic Codes.......... 99
Table 4-5: Percentage Code Frequency of Specific Requirement vs. Other Analytic Codes..... 104
Table 4-6: Percentage Code Frequency of Defined Goal & Scope vs. Other Analytic Codes... 109
Table 4-7: Percentage Code Frequency of Staff Experience vs. Other Analytic Codes............. 115
Table 4-8: Percentage Code Frequency of Enterprise Data Consolidation vs. Other Analytic Codes.. 120
Table 5-1: Risk and Finance Data Warehouse Common Business Data *(Adapted from ICapital Artifacts)* 138
Table 5-2: Descriptive Codes for Line-By-Line Coding 143
Table 5-3: Analytic Codes for Descriptive Codes 144
Table 5-2: Frequency Analysis of the ICapital Analytic Codes *(See Appendix 4 for ICapital Open Codes)* 150
Table 5-3: Percentage Code Frequency of Enterprise Data Integration vs. Other Analytic Codes 152
Table 5-4: Percentage Code Frequency of Generic Requirements vs. Other Analytic Codes ... 156
Table 5-5: Percentage Code Frequency of Defined Goal & Scope vs. Other Analytic Codes... 161
Table 5-6: Percentage Code Frequency of Staff Experience vs. Other Analytic Codes............. 166
Table 5-7: Percentage Code Frequency of Query Performance vs. Other Analytic Codes 170
Table 6-1: Summary of the Findings at GWealth and ICapital Case Studies............................ 182
Table 6-2: Number of Frequency of GWealth Analytic Codes 203
Table 6-3: Number of Frequency of ICapital Analytic Codes.. 203

Table 6-4: Contingency table of Logical Data Models and Observed Frequency of Analytic Codes.. 204
Table 6-5: Observed and Expected Frequency of Analytic Codes and Logical Data Models ... 204
Table 6-6: Chi square distribution table, Source: *https://people.richland.edu/james/lecture/m170/tbl-chi.html, [Date Accessed: 26/05/2014; 30/09/2014]*.. 207
Table 6-7: Percentage Contribution of the Analytic Code Variables 208
Table 6-8: Derivation and Translation of Relative Weight and Rating Scale for Decision Matrix 215
Table 6-9: Decision Matrix for Logical Data Model Adoption in Data Warehousing.............. 216

ACRONYMS

• **BA**	Business Analyst - New
• **BI**	Business Intelligence
• **BPODM**	Business Process-Oriented Data Mart
• **BPODW**	Business Process-Oriented Data Warehouse
• **CBDSS**	Communication Based DSS
• **CBIS**	Computer Based Information System
• **CIF**	Corporate Information Factory
• **CDW**	Classic Data Warehouse
• **CRAE**	Compliance Rule Analytics Engine
• **CDW**	Corporate Data Warehouse
• **DA**	Data Architect
• **DASD**	Direct Access Storage Device
• **DBDSS**	Document Based DSS
• **DBODSS**	Database Oriented DSS
• **DD**	Data-Driven
• **DDDSS**	Data-Driven DSS
• **DODSS**	Document-Oriented DSS
• **DDW**	Department Data Warehouse
• **DDDW**	Distributed Dimensional Data Warehouse
• **DDWBA**	Dimensional Data Warehouse and Bus Architecture
• **DFM**	Dimensional Fact Model
• **DM**	Data Mart
• **DML**	Data Mart Layer
• **DMS**	Data Mining System
• **DPL**	Data Processing Layer
• **DSS**	Decision Support Systems
• **DW**	Data Warehouse
• **DWL**	Data Warehouse Layer
• **DVDA**	Data Vault Data Architecture
• **EDW**	Enterprise Data Warehouse
• **EII**	Enterprise Information Integration
• **EIS**	Executive Information System
• **FOA**	First Order Abilities
• **FSA**	Financial Services Authority
• **GD**	Goal-Driven
• **GDW**	Galactic Data Warehouse
• **HODSS**	Hypertext Oriented DSS
• **IDSM**	Intelligent Decision Support Methods
• **KBDSS**	Knowledge Based DSS

- **KD** Knowledge Discovery
- KDDSS Knowledge Oriented DSS
- KS Knowledge System
- LDM Logical Data Model
- LS Language System
- MBDSS Model Based DSS
- MDM Multidimensional Data Model
- MIS Management of Information Systems
- MODSS Model Oriented DSS
- NEA Naturally Evolving Architecture
- NDM Naturalistic Decision Making
- ODS Operational Data Store
- PC Personal Computer
- PM Project Manager
- PPS Problem Processing System
- PS Presentation System
- PTS Pivot Table Services
- ODBC Open Database Connectivity Layer
- OLAP Online Analytical Processing
- OLTP Online Transaction Processing
- OWB Oracle Warehouse Builder
- RAE Risk Analytics Engine
- RFDW Risk and Finance Data Warehouse
- RDBMS Relational Database Management System
- RDM Relational Data Model
- RFDW Risk and Finance Data Warehouse
- SAL Solver Access Layer
- SBDSS Spreadsheet Based DSS
- SODSS Spreadsheet Oriented DSS
- SSIS SQL Server Integration Services
- TA Technical Analyst
- TDW Transactional Data Warehouse
- UAL User Application Layer
- UD User-Driven

CHAPTER 1 - INTRODUCTION

1.1 Background to the Problem

Over time, organisations have accumulated vast amounts of data in the course of conducting day-to-day business activities. In their drive to increase market share and to create greater shareholder value, organisations have come to rely on their most prized asset, data, in order to obtain the necessary intelligence to understand and improve their businesses (Russom, 2009). The need to manage that data has led to increased number of organisations initiating data warehousing projects to support decision making to drive growth. A data warehouse is primarily used in strategic decision making; it is a collection of decision support technologies, aimed at enabling knowledge workers (executives, managers, and analysts) to make better and faster decisions (Chaudhuri and Dayal, 1997). However, data warehouses are expensive initiatives, one-half to two-thirds of most data warehousing efforts end in failure (Hayen et al, 2007; Mark, 2005).

While the need for and the benefits of data warehouses are understood, there is no consensus on the logical data model on which to implement a data warehouse (Rizzi, Abelló, Lechtenbörger and Trujillo, 2006; Kimball and Ross, 2002; Imhoff, Galemmo and Geiger, 2003; Inmon, 2005; Kimball 1997). Two main data models based on multidimensional or relational paradigm exist in data warehousing, they are commonly referred to as multidimensional and relational data models respectively (Kimball and Ross, 2002; Inmon, Imhoff, Sousa, 2001).

In its basic form, a multidimensional data model enables a data warehouse to cater for a specific reporting or analytics need of a business function rather than the need of an entire organisation (Kimball, 1997; Kimball and Ross, 2002). This way, a multidimensional data warehouse is engineered and driven from the perspective of the business users (Kimball, 1997; Kimball and Ross, 2002; Bruckner, List and Schiefer, 2001; Prakash and Gosain, 2003; Sperley, 1999; Golfarelli and Rizzi, 1998; Husemann, Lechtenbörger and Vossen, 2002). In order to use a multidimensional data model to develop a data warehouse that cater for the needs of an organisation, '*bus architecture*' is often recommended as the basis for multidimensional enterprise data warehouses (Kimball, 1997; Kimball and Ross, 2002). A bus architecture framework uses conformed dimensions in multidimensional approaches to provide a broader view of an enterprise. A dimension is said to be conformed having been transformed through a process of standardization that enables its attributes to have similar column names and domain contents (Kimball, 1997; Kimball and Ross, 2002).

A relational data model is often portrayed through a Corporate Information Factory (CIF) framework (Inmon, Imhoff and Sousa, 2001). A corporate information factory is an '*organisation information ecosystem*' consisting of an external world, applications, data integration, operational data stores, data warehouse, data marts, intranet and internet, data mining, alternative storage and decision support systems (Inmon, Imhoff and Sousa, 2001). In a corporate information factory approach, data within an enterprise is extracted from various operational upstream systems and consolidated in a data warehouse.

The logical data model for a data warehouse in corporate information factory framework is a relational data model; this data model is oriented to an enterprise view of an organisation (Inmon, Imhoff and Sousa, 2001; Imhoff, Galemmo and Geiger, 2003; Tryfona, Busborg and Christiansen, 1999). The outcome of these superficially opposing approaches has resulted in a practice in the industry where most data warehouses are largely implemented using either multidimensional or relational approaches and the reasons why one is chosen over the other are unclear empirically.

1.2 Aim and Objectives

The aim of this research is to examine the decision factors influencing the choice of logical data model for data warehouses in the industry.

The objectives of the research are to:

1. Review the state-of-the-art in data warehouse development in order to distill the factors affecting the choice of data model in data warehousing.
2. Synthesise the literature in order to try and develop a framework impacting the choice of logical data model in data warehousing.
3. Undertake a comparative case study in the organisations using the multidimensional and relational data models to test the factors impacting the choice of logical data model for a data warehouse.
4. Evaluate the outcome of the study to determine the alignment of the decision factors in practice and contributions to theory.

1.3 Method of Research

The work in this thesis uses case study as the primary method as it is appropriate for an empirical investigation of contemporary phenomenon in its natural context using a number of sources of evidence (Hancock and Algozzine, 2006). Research that asks why and how, where a researcher requires no control and does not attempt to influence the behaviour of participants is suited to case study (Yin, 2009). Case study method is used to achieve various aims; it is used to provide description, to generate propositions, hypothesis or theory (Hancock and Algozzine, 2006; Yin, 2009; Eisenhardt, 1989; Cavaye, 1996). Case study is also appropriate for situations where the focus of a study is broad and complex and there is not a lot of theory available (Dul and Hak, 2008). Case study focuses on understanding the dynamics within cases and researchers that utilize case method routinely use a variety of techniques such as interviews, questionnaires and observations to collect evidence (Eisenhardt, 1989; Cavaye, 1996).

This study is comparative in its nature, using data from various divisions covering wealth management (GWealth) and investment banking (ICapital) divisions of a Tier 1 global banking institution (TBank). The case organisations employed multidimensional and relational data models for their data warehouses respectively. The research design for this study uses literature to distill the key propositions influencing the choice of logical data model for a data warehouse. These propositions inform a set of interview questions that were administered to GWealth and ICapital research participants.

The interviews of research participants were conducted in parallel stream enabling the investigator to collect empirical data specific to each case. The GWealth serves affluent and high net worth of clients with offices in 25 countries across key business areas including private and international banking, wealth advisory, research, economics and strategy. The ICapital is a major player in global financial markets offering investors a range of security products including equities, government and corporate debt instruments, commodities, securitization, credit, credit derivative and interest rate products.

1.4 Structure of the Thesis

In addressing the research aim and objectives, the thesis is organized as follows.

Chapter 2: Drawing on literature, the chapter presents a review of literature in data warehousing and examines the key components of a data warehouse system. A broad view of the phases of implementing a data warehouse system is presented leading to examination of decision theories. The chapter proceeds by exploring the discourse in the industry regarding the adoption of multidimensional or relational data model for a data warehouse. The research propositions are then presented and discussed. Finally, based on decision theory, a conceptual model for adopting a logical data model for a data warehouse is presented.

Chapter 3: This chapter explores research philosophies and presents an overview of research methodologies. The chapter examines why a case study method is suitable for this study and compares and contrasts single and multiple case methods. The chapter proceeds by presenting the research design for this study, discusses the components of the research framework and describes the process of linking the empirical data with the research propositions.

Chapter 4: This chapter provides an overview of GWealth case study and describes its data warehouse project including the project definition and objective. The data architecture of GWealth data warehouse is introduced and the chapter proceeds by presenting the results of the case study at GWealth. Finally, the chapter concludes by providing a summary of the research findings.

Chapter 5: The chapter presents an overview of ICapital data warehouse project including the project definition and the objectives of the data warehouse. The data architecture of ICapital data warehouse is introduced including a description of the key components of its data warehouse. The chapter proceeds to present the results of ICapital case study. Finally, the chapter concludes by providing a summary of the research findings.

Chapter 6: This chapter provides a comprehensive discussion of Chapters 4 and 5. The chapter provides in-depth analysis of the research findings in relation to the literature. The chapter proceeds to provide a broader discussion of each of the research propositions drawing on literature and the empirical data to drive the discussion of the research findings, leading to a statistical examination of the significant relationship between the research analytical code variables and the logical data models. The chapter proceeds by presenting a decision matrix for adopting logical data model and represents the revised conceptual model presented in chapter 2. Finally, the chapter discusses the implications of this study for practice and theory and concludes with a summary.

Chapter 7: The chapter concludes the study in this thesis; it presents a summary of the research chapters and the contributions of the research to industry and research sectors. The chapter outlines the limitations of this research and provides the areas for future research.

CHAPTER 2 – Literature Review

2.1 Introduction

Data warehousing has been around for over two decades and has become an important part of an organisation's information technology infrastructure. To build a data warehouse, data from heterogeneous operational systems are extracted, de-duplicated, cleansed, transformed and consolidated in a data warehouse. The key components of a data warehouse represent the components that must be integrated and function together for the efficient running of a data warehouse system. The architecture of a data warehouse describes the components and services that make up a data warehouse system, how they fit together and how they scale for future growth of a data warehouse system. A body of literature exists on data warehousing; however, there is no formalised standard for developing a data warehouse. The literature in data warehousing reflects the debate in the industry in the absence of consensus on the logical data model with which to implement a data warehouse. In data warehousing, the observed trend in the industry and the research sector is that, in general, two approaches based on the multidimensional and relational data models are commonly used for developing a data warehouse.

This chapter is divided into 8 Sections. Section 2.2 looks at the overview of a data warehouse system. Section 2.3 looks at the key components of a data warehouse system from an end-to-end perspective and Section 2.4 explores different phases of implementing a data warehouse. Section 2.5 examines decision theory and explores the theoretical prescription for decision making.

The Section looks at normative, descriptive and prescriptive decision making theories. Additionally, Section 2.5 examines how decisions are made under risk and uncertainties and discusses the naturalistic decision-making. Section 2.6 engages the literature to discuss the criteria influencing the choice of logical data model. Finally, Section 2.7 concludes with a summary.

2.2 Overview of a Data Warehouse System

Data warehousing has been around for over two decades and has become an important part of an organisation information technology infrastructure (Inmon, Strauss and Neushloss, 2008; List, Bruckner, Machaczek, and Schiefer, 2002; Golfarelli and Rizzi, 1998; Bonifati, Cattaneo, Ceri, Fuggetta and Papaboschi, 2001; Böhnlein and Ulbrich-vom Ende, 2000; Herden, 2000). Data warehousing grew out of organisations need for information, a data warehouse contains data sourced from multiple heterogeneous systems within an organisation. The data from heterogeneous sources are extracted, de-duplicated, cleansed, transformed and consolidated in a data warehouse (Inmon, Imhoff and Sousa, 2001; Luján-Mora and Trujillo, 2003, Kimball and Ross, 2002, Imhoff, Galemmo and Geiger, 2003; Inmon, 2005; Reeves, Ross and Thornthwaite, Kimball, 1998). A data warehouse is therefore, in essence, a repository that enables data from multiple operational systems within an organisation to co-exist in a single database. At an enterprise level, the advantage of a data warehouse to an organisation is that different business areas of an organisation are able to subscribe to and obtain operational data from a data warehouse.

Additionally, the data in a data warehouse has one view that is consistent throughout an enterprise (Inmon, 2001; Inmon, 2005; Inmon, 2008; Inmon, Imhoff and Sousa, 2001). A data warehouse exhibits a number of characteristics, which distinguishes it from other operational databases. Table 2-1 below outlines the differences between a data warehouse and an operational database.

Distinguishing Characteristics	Data Warehouse	Operational Database
Subject Oriented	Business data organized around the key subjects: sales, orders, locations	Business data organized around the business application systems it supports
Non-Volatile	Data not subjected to editing once is in the data warehouse. A Query returns the same results no matter and how often the query is run	Data is volatile because the applications they support frequently update the database
Integrated	Data derived from multiple sources within and outside the organisation merged an ETL process	Data derived from the operational system it supports
Time Variant	Every data structure in a data warehouse has a time element to track historical and current data values	Time elements are frequently updated

Table 2-1: Differences between Data Warehouse and Operational Database (*Adapted from Todman, 2001; Inmon, 2005*)

A data warehouse system is a collection of decision support technologies, aimed at enabling knowledge workers (typically executives, managers, and analysts) to make better and faster decisions (Chaudhuri and Dayal, 1997). It is postulated from the above that the main reason why an organisation engages in implementing a data warehouse is to support timely management decision-making (Kimball and Ross, 2002; Inmon, 2005; Inmon, Imhoff and Sousa, 2001;

Sperley, 1999). Figure 2-1 below illustrates end-to-end components of a corporate data warehouse system.

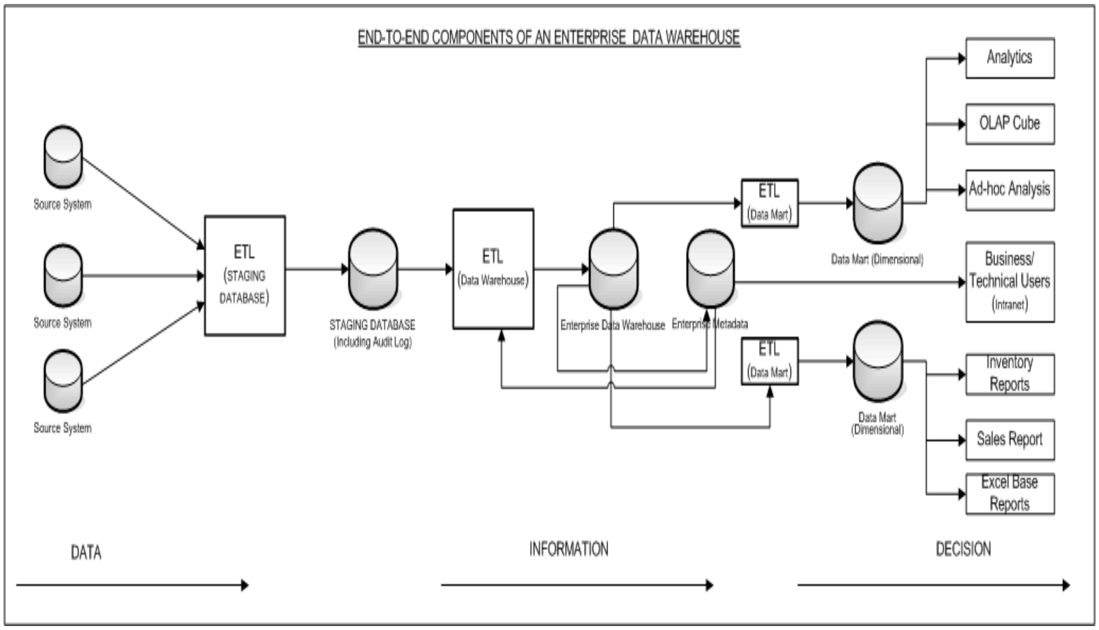

Figure 2-1: End-to-end component of corporate data warehouse system (*Adapted from* Inmon, 2005; Inmon, Imhoff and Sousa, 2001; Sperley, 1999)

2.3 The Key Components of a Data Warehouse System

The key components of a data warehouse system represent the components that must be integrated, co-exist and function together for efficient operation of a data warehouse system. The architecture of a data warehouse describes the components and services that make up a data warehouse system in relation to how they fit together and how they scale for future growth (Inmon, Imhoff and Sousa, 2001; Inmon, Strauss and Neushloss, 2008). The corporate data flow architecture defines the components that are required to meet the various data objectives and requirements of the business across an enterprise. The components include standards, principles and methodologies (Hashmi, 2000).

The benefit of a well architected data warehouse is that it is in alignment with a corporate information factory framework as it adapts to new and changing business landscape (Moody et al, 2000). A corporate information factory is the physical embodiment of the notion of information ecosystem. It consists of external world, applications, integration, operational data stores, data warehouse, data marts, intranet and internet, exploration and data mining, alternative storage and decision support systems (Inmon et al, 2001). A corporate information factory is therefore a landscape of organisation systems. Table 2-2 outlines the key components of a corporate data warehouse system.

Component Layer	Description
Data Acquisition	Describes the acquisition of data warehouse source data
Data Processing	Focuses on data extraction, transformation and loading
Data Warehouse	Describes the data warehouse repository
Metadata	Describes attributes definitions, transformation and data quality rules
User Applications	Provides information to DSS users to make decisions

Table 2-2: The Key Architecture Components of Corporate Data Warehouse System (*Adapted from Sperley, 1999*)

- A data acquisition layer (DAL) provides the source data that is integrated into a data warehouse. The data in the data acquisition layer is sourced from a number of operational, legacy and external systems and are processed into a data warehouse through the components of a data processing layer (Imhoff, Galemmo and Geiger, 2003; Kimball and Ross, 2002; Inmon, 2005; Inmon, Imhoff and Sousa, 2001; Inmon, Strauss and Neushloss, 2008).

- A data processing layer (DPL) encapsulates the activities associated with loading source data into a data warehouse. The components in a data-processing layer are responsible for staging the source data extracts. The main component of a data processing layer is an ETL processing engine. An ETL engine is responsible for all the back room data management processes and delivering data into a data warehouse (Caserta and Kimball, 2004; Mrunalini, Kumar, Geetha and Rajanikanth, 2006; Vassiliadis, Simitsis and Baikousi, 2009).

- A data warehouse layer (DWL) is the destination database for storing data sourced from multiple sources within and outside an organisation. A data warehouse contains subject-oriented data entities normalized as a relational model (*in corporate information factory framework*) or architected as a collection of fact tables with conformed dimensions (*in bus architecture framework*). This way, a data warehouse delivers common view of corporate data for an enterprise; it is also a data provider to other downstream user applications within an organisation (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003; Kimball and Ross, 2002; Kimball, 1997; Reeves, Ross, Thornthwaite and Kimball, 1998; Inmon, Imhoff and Sousa, 2001).

- A metadata component contains information that describes the details of the objects that make up a data warehouse; this includes for instance, the definition of objects, entities and attributes of a data warehouse. Metadata is described in the literature as data about data and it is important in resource discovery. Metadata take on different aliases in many organisations; terms such as information library, data dictionary and information directory have been used to describe metadata repository (Inmon, Imhoff and Sousa, 2001; Mallach, 2000; Malaxa and Douglas, 2005; Marco, 2000; NISO, 2004).

- User application layer (UAL) describes the types of applications accessing a data mart layer in end-to-end data warehouse system. A number of applications in this layer include applications deployed to write ad-hoc queries, standard end-user applications including reporting and analytics applications and OLAP interfaces (Inmon, Imhoff and Sousa, 2001; Imhoff, Galemmo and Geiger, 2003).

The next Section looks at the phases of implementing a data warehouse

2.4 Implementing a Data Warehouse

While the need for a data warehouse is clear, there is no clear consensus on the logical data model on which to build a data warehouse (Rizzi, Abelló, Lechtenbörger and Trujillo, 2006; Kimball and Ross, 2002; Imhoff, Galemmo and Geiger, 2003; Inmon, 2005; Kimball, 2005; Kimball 1997), however, two approaches based on multidimensional and relational data models are commonly used to implement a data warehouse. In the literature, data warehouses using a multidimensional data model are primary based on star schemas, here; a collection of dimensions are joined to a central fact table usually through surrogate keys. The attractiveness of using a star schema as the logical data model for a data warehouse is that it is less expensive, easier to build and improves performance (Kimball and Ross, 2002; Kimball, 2005; Kimball 1997). On the other hand, a relational data model is used for a data warehouse where the objective of the data warehouse is to cater for needs of an organisation as a whole. In this approach, a relational data model is used for large data warehouses (Galemmo and Geiger, 2003; Inmon, 2005).

In the literature, the starting point of any design work in implementing a data warehouse is a conceptual design phase (Sperley, 1999; Tryfona, Busborg and Christiansen, 1999; Husemann, Lechtenbörger, Vossen, 2002; Herden, 2000). The main purpose of a data warehouse conceptual phase is to identify the processes and components that are required to build, use and maintain a data warehouse system (Sperley, 1999). A conceptual schema is a map of the key concepts and associated relationships that is transformed into a logical design model; this becomes the basis for the physical implementation of a data warehouse (Husemann, Lechtenbörger and Vossen, 2002).

In the industry, there is a difference between the logical data model of a data warehouse that is oriented to the requirements of a business unit (*within an organisation*) and that oriented to the requirements of an entire enterprise. As data warehouse development process progresses from a conceptual phase to a logical phase in the development life cycle, a decision needs to be made to determine the type of logical data model that will underpin a data warehouse. This decision as important as it is has not been given the attention and focus it deserves in the literature. Very little work exists that aims at data architecture in data warehousing, related work in this field is a study that investigated which data warehouse architecture is most successful (Ariyachandra and Watson, 2006). The study largely compared different architecture methods, based on four measures (*information quality*, *system quality*, *individual impact* and *organisation impact*); it determined which method is considered '*Most Successful*' (Ariyachandra and Watson, 2006).

Data warehouses are very expensive initiatives, various failure rates ranging from one-half to possibly two-thirds of data warehousing efforts end in failure (Hayen et al, 2007; Mark, 2005). Hayen et al (2007) notes three factors relating to environmental, project and technical factors are largely responsible for failure of data warehouses. Environmental factors are associated with lack of support from senior management, strategic changes such as mergers and acquisitions, corporate culture, organisation decision making and lack of business drivers. According to Hayen et al (2007), project factors contributing to failure of data warehouses include lack of understanding of complexity, workload and cost involved in such an undertaken, inability to quantify return on investment (ROI), poor selection of tools, failure to understand and manage project scope.

Additionally, Hayen et al (2007) notes that technical factors impacting the success of data warehouses include data quality, lack of understanding of data needed for a data warehouse and lack of common definition of data across an organisation. Barlow (2013) notes building a data warehouse requires careful consideration of the risks associated with building a data warehouse, overtime, implementing a data warehouse may not be successful for a number of other reasons. First, a data warehouse requires a solid business case, data warehouses that primarily emanates from a technical perspective lacks business case because it is difficult to get "after-the-fact by-in" from the business (Barlow, 2013). Second, a data warehouse requires executive sponsorship and continuous engagement of the management for a data warehouse to be successful (Barlow, 2013). Third, lack of understanding of the scope of what can be achieved, rather than capturing every data that an organisation encounters without careful consideration of the cost associated with such an implementation, instead, implementation teams should balance the ideal with reality of what can be achieved (Barlow, 2013). Fourth, Barlow (2013) notes, organisation should not wait for data governance to be perfect before starting to develop a data warehouse and making data governance a pre-requisite for building a data warehouse is costly because it may cause a data warehouse project to collapse under its own weight.

While there are varieties of reasons why a data warehouse may be deemed a failure, the failure rate of most data warehouse implementations is a reminder of the need to ensure that an appropriate logical data model is adopted for a data warehouse. In data warehousing, the most dominant logical data models are multidimensional and relational data models, the decision to engage one logical data model at the expense of the other is influenced by a number of factors that must be carefully considered; this thesis explores and examines the degree to which these factors influences such a decision. Furthermore, the chosen logical data model for a data warehouse influences the overall data warehouse methodology and the way the components of the data warehouse system is architected and integrated. In data warehousing, multidimensional data warehouses are based on a multidimensional logical data model. The rationale underpinning the multidimensional approach is that a data warehouse is implemented to address the need of a given user community, consequently, such a data warehouse is driven from the perspective of the business users expressed in the business requirements document-BRD (Bruckner, List and Schiefer, 2001; Prakash and Gosain, 2003; Sperley, 1999; Golfarelli and Rizzi, 1998; Husemann, Lechtenbörger and Vossen, 2002). In multidimensional data modelling, data elements such as entities are denormalised to provide a structure that consists of dimensions and fact table typically known as the star schema (Kimball and Ross, 2002; Sperley, 1999). A multidimensional data model in most part contain the same level of information as a relational data model, however, a multidimensional data model provides data in way that enhances the user understanding of the data, improve performance and resilience to change (Kimball and Ross, 2002; Herden, 2000). In a multidimensional data model, all the dimensions are symmetrical to the fact table enabling the queries that are based on a multidimensional data model to be processed in a similar fashion (Kimball, 1995, 1997; Kimball and Ross, 2002). In the literature, a number of advantages have been attributed to a multidimensional data model:

- First, multidimensional data model has limited complex data structures; this simplifies the data relationship in a multidimensional data warehouse. The simplicity of the data model allows a data warehouse to be structured in a way that makes it easier for the business users to understand and query (Kimball, 1995, 1997; Kimball and Ross, 2002; Reeves; Ross, Thornthwaite and Kimball, 1998; Sen and Sinha, 2005; Sperley, 1999).

- Second, a multidimensional data model is considered a predictable and standard framework that improves the end-user query performance and facilitates re-usability of components through '*bus architecture*' (Kimball, 1997; Reeves, Ross, Thornthwaite and Kimball, 1998; Weininger, 2002). The bus architecture is a framework that allows dimensions to be re-used across many fact implementations to provide a broader view of the organisation; dimensions are conformed in bus architecture framework. In order to improve performance, a multidimensional data model provides a fixed structure with no alternative join paths; this is considered to greatly simplify queries evaluation and optimization in multidimensional data warehousing (Kimball, 1995; Kimball, 1997; Golfarelli and Rizzi, 1998; Sen and Sinha, 2005).
- Third, the general understanding of a well architected data warehouse is that its logical data model must be in the form of a star schema. The star schema enables the dimensions of a data warehouse to be connected to the central fact table using the surrogate keys (Husemann, Lechtenbörger and Vossen, 2002).

The alternative view of logical data modelling in data warehousing is a relational data model. The relational data modelling approach has its root in the traditional relational database development in online transaction application-processing (OLTP) sphere. The rationale underpinning this alternative approach in data warehousing is based on the assumption that a data warehouse is enterprise-wide in nature, as such; a data warehouse must not be biased towards a given user community or any department within an organisation (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003; Tryfona, Busborg, Christiansen, 1999). This position is based on the premise that different departments within an organisation usually require data in a variety of format to address their business needs. As a result, additional data processing is always necessary to deliver data in the way required by a business area (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003; Inmon, Imhoff and Sousa, 2001). In the literature, a number of advantages are attributed to a data warehouse implemented on a relational data model:

- First, a relational data model provides capabilities such as enterprise data consolidation and diversification, this enables different areas of an organisation to source data from the same centralised data warehouse irrespective of how they want to finally see or use data in their relevant departments (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003).
- Second, data warehouses based on relational data models are versatile, flexible and provides a single source of consistent data within an organisation. A data warehouse using a relational data model is oriented to the major subject areas of an enterprise (Inmon, 2005; Drewek, 2005; Inmon, Imhoff and Sousa, 2001).
- Third, a relational data model enables a data warehouse to address generic and '*unknown*' business requirements because most business users do not know the requirements they want from a data warehouse until the data warehouse is built. Over time, a number of business requirements are later submitted against the data warehouse (Inmon, Imhoff and Sousa, 2001; Inmon, 2005; Inmon, Strauss and Neushloss, 2008; Tryfona, Busborg and Christiansen, 1999).
- Fourth, a relational data model enables a data warehouse to be extended gracefully and allows data elements to be combined seamlessly and cohesively (Inmon, 2005; Drewek, 2005; Imhoff, Galemmo and Geiger, 2003). The value of a relational data model to the design of a data warehouse is that it is not only flexible; the model can be viewed from

multiple different perspectives. The perspectives allow the data warehouse to support different business reporting and analytic requirements within an organisation, thus, reducing the occurrence of silo data marts in an organisation (Inmon, 2005; Drewek, 2005; Imhoff, Galemmo and Geiger, 2003).

In the light of the financial and material costs usually associated with implementing a typical data warehouse and given the high stake nature of most data warehouse implementations in many organisations, this thesis investigates why the case organisations adopted multidimensional and relational data models respectively for their data warehouses. In so doing, the thesis examines decision making specifically the Naturalistic Decision Making (NDM) theory as the theoretical framework for this research. NDM theory looks at decision making in high stake situations characterised by stress and uncertainties. This thesis examines the theoretical prescriptions of decision theories in general and explores how decisions are made in real-life context. The next Section explores decision theories.

2.5 Decision Theories

Today's decision-makers are faced with complex and interrelated problems that are routine and dynamic in nature, as a result, a number of decisions are made rather a single decision; these decisions are interdependent and the environment where they are made changes periodically (Hassan, Spector, Davidsen, 2008; Johnson and Busemeyer, 2010). Decision theory is theory about a decision; it is aimed at goal oriented behaviour given the available alternative options

(Hansson, 1994). In order to understand how specific decisions are made, there is a need to have concrete understanding of the context and knowledge underpinning such decisions including for instance, the details regarding the organisation where the decisions are made, the political, social, economic and historical factors that surround the decisions; not least, the participants taking part in the decision making process (March and Heath, 1994). The classical theories of choice interpret actions as rational choice (March and Heath, 1994; Simon, 1955). This approach assumes people make rational choices based on a procedure for making informed choices. A decision process is consequential and preferential because actions are anticipated and dependent on future effects of today's actions (Simon, 1955; March and Heath, 1994; Dillon, 1998; Jones, 1999). Also, alternatives are considered relative to expected consequences and personal preference of the decision-maker (March and Heath, 1994; Jones, 1999). The theories of rational choice are based on a number of assumptions within which a rational choice must take place, the assumptions include *'(1) the set of alternatives open to choice, (2) the relationship that determine the pay-offs as a function of the alternative that is chosen, (3) the preference-orderings among pay-offs'* (Simon, 1955, pp. 100). Rational decision making engages logic of consequences, in so doing; a number of factors are considered. For instances:

- What actions are possible and what alternative choices are available?
- What are the effects of future expectations of current actions given the alternative choices?
- What are the personal preferences of the decision-maker given the consequences inherent in the alternative choices?

- What decision rule directs how given values are associated with consequences which then informs the choice that is made amongst the alternatives (Simon, 1995, March and Heath, 1994; Jones, 1999; Johnson and Busemeyer, 2005, 2010).

In a pure theory of rational choice, one school of thought assumes decision-maker exhibits common sets of preferences; in this paradigm, the environment determines alternatives and preferences. Also, a decision-maker has good knowledge of the available alternatives and associated consequences (March and Heath, 1994). Another school of thought assumes that all the possible alternatives and consequences of alternatives are known with certainty, that preferences are not only known with precision and consistency, they are stable (March and Heath, 1994). In further exploring decision-making, this thesis examines other aspects of decision theory such as normative models of decision making, descriptive, prescriptive decision making and discusses the naturalistic decision making.

2.5.1 Normative Decision Theory

Normative decision theory is about how decisions should be made in order to be rational, it is not aiming at how people make decisions instead; it focuses on how people should make decision in the absence of good information (Border, 1984; Hansson, 1994). Normative decision models are derived from mathematical models such as utility and probability theories and statistics. The models are termed normative because they use mathematical rules; the practice of using mathematical rules is the norm in evaluating deviation of decision from the models (Baron, 2004). In normative decision theory, constructing a decision model begins with a decision-maker understanding the situation in which a theory is to be built (Brenna, 1995). As humans have limited capacity in absorbing information to process complex tasks especially when faced with value-laden situations surrounded by uncertainty, normative decision theory such as probability provides a solution to lack of human capacity to handle complex tasks (Brenna, 1995). This is done in two ways: First, a normative model allow complex problem to be broken down into manageable parts, thus, reducing the implicit cognitive workload in normative decision making. Second, normative model foster rational choice by linking alternatives or choices to norms that is external to the decision problem thereby ensuring internal consistency (Brenna, 1995).

Normative decision techniques are appealing in situations where there is a need to engage competing choices that must satisfy multiple criteria characterized by risk and uncertainty (Brenna, 1995). Data warehousing decision such as adopting a multidimensional or relational data model presents a dilemma for a decision-maker because the risk of failure is real and costly. The choice of one data model over another is contingent on resolving uncertain future events such as introduction of unknown business requirements to address changing business needs. However, from a practical perspective, it is not immediately clear what are the benefits of engaging normative mathematical models to determine the choice of logical data model for a data warehouse. The next Section examines the descriptive decision theory.

2.5.2 Descriptive Decision Theory

The theory of choice traditionally emphasises decision making as rational choices between alternatives (Johnson and Busemeyer, 2005, 2010). Rational choice is based on expectations given the consequences of action of previous objectives. The descriptive theory of decision making focuses on the way people actually make decisions instead of prescribing an ideal decision for a particular decision situation (Dillon, 1998; Johnson and Busemeyer, 2010). Dillon (1998) identified three phases of decision making process in descriptive theory; the phases correspond to *intelligence*, *design* and *choice*. In the intelligence phase, the need for a decision is identified; this is called searching the environment phase. The design phase begins when the environment search is completed; it involves investigating the domain problem and developing alternative solutions to address the problem domain. In the last phase of descriptive decision making, *choice*; an appropriate action is taken, a choice is made from the alternatives that were generated in the design phase (Dillon, 1998). Figure 2-2 below illustrates the model of the decision process.

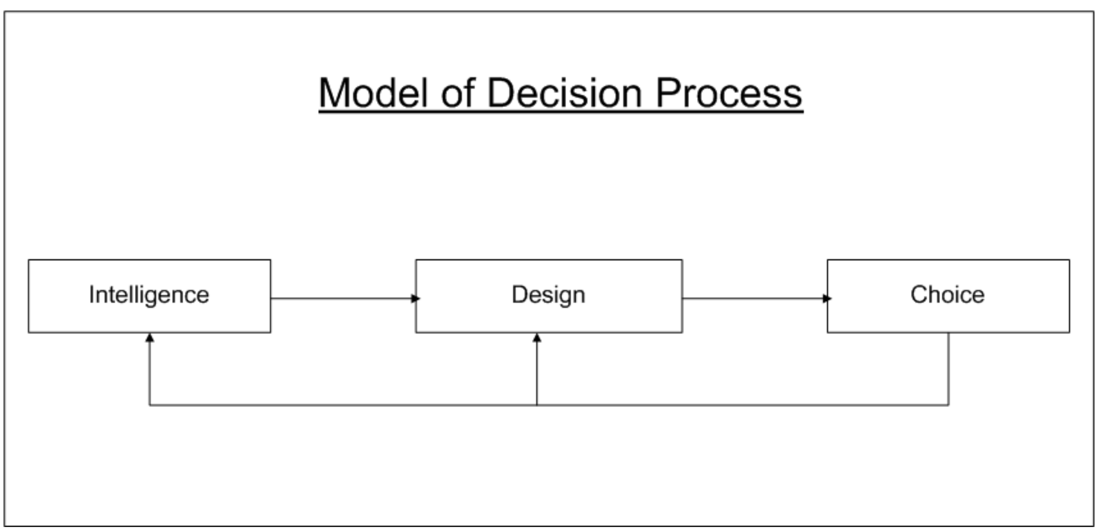

Figure 2-2: The Model of Decision Process (*Adapted from (Dillon, 1998)*)

As illustrated above, each phase of the descriptive decision making is iterative; each phase may call for further intelligence or generate additional problems which can contain implicit *intelligence*, *design* and *choice* phases. The central theme of descriptive theory is bounded rationality; this is also known as limited rationality (Dillon, 1998; Jones, 1999). Limited rationality highlights the impact of cognitive obstacles to decision making in classical decision theory. In bounded rationality, decision-makers intentionally wants to make rational decisions, however, they are not always able to do so because of the inherent mismatch between the decision that a decision-maker makes and the environment where the decision is made (Jones, 1999). In descriptive decision theory, there is an acknowledgement of uncertainty because a decision maker does not know the likely future outcome of a given action. It is unlikely that outcomes will match expectations because the decision-maker does not know all the alternative options and not all consequences associated with an alternative option are evaluated. Instead, decision maker satisfice rather than maximise the expected value; in so doing, the decision maker selects an alternative that exceeds some laid down criteria instead of selecting the best alternative amongst the available alternatives (March and Heath, 1994). The next Section examines prescriptive decision theory.

2.5.3 Prescriptive Decision Theory

The literature has explored the differences between how decision should be made and how they are actually made, the conclusion was, the presence of potentially costly errors has not done much to dissuade the tendency to make decisions using simplistic rule of thumb that sometimes produce decisions that departs from those instructed by normative theories (Kunreuther et al, 2001). A prescription for correcting deviation from normative theories of decision making is referred to as prescriptive model (Baron, 2004). In prescriptive theory, the problem is not really that decision-makers are not necessarily unaware of the need to consider the consequences of costly errors and the likelihood that such errors will occur, but rather, they have limited capacity to process information in an optimal manner. In this case, a prescriptive rule of thumb is capable of enhancing the normative process given that decision-makers have natural limitation to process information. In prescriptive decision theory, the likelihood of risk is translated into a concrete form that gives the decision-maker a clear context for evaluating risks (Kunreuther et al, 2001). Prescriptive decision theory incorporates the framework of normative decision theory; the advantage of prescriptive decision theory is that it is flexible enough to allow it to be tuned to the needs of decision-maker as a prescriptive advice (Bell, Raiffa and Tversky, 1988; Dillon, 1998).

The next Section examines naturalistic decision making.

2.5.4 Naturalistic Decision Making

The major contribution of naturalistic decision making (NDM) over the other theories of decision making is describing how decisions are made in real-world setting. In contrast to the prescriptions of rational theories of choice, in reality, decision makers are not creating alternative options and are not evaluating options based on logic of consequences that is acceptable to a decision maker. The observation in the literature is that in instances where decision-makers compare options, rarely do they engage any form of systematic evaluation techniques (Klein, 2008). The focus of naturalistic decision making is looking at how people actually make decision given constraints such as uncertainty, limited time availability, unstable conditions, high stakes and vague goals. When making decisions, people are accessing and categorising situations using their previous experiences to make judgements and each categorised situation engages a different course of action (Klein, 2008).

In contrast to the model of rational choice which assumes decision-makers passively await the outcome of the alternatives, in NDM, decision-makers actively engages action to change events (Baron, 2004; Klein, 2008). In naturalistic decision-making, a decision-maker commits himself or herself to a given course of action where acceptable alternatives exist although the decision-maker may not necessarily compare those alternatives (Klein, 2008). In naturalistic decision making, a leader engages action to understand his or her world instead of collecting information in a passive manner (Klein, 2008; Azuma et al, 2006). The naturalistic decision making shifts our perception of decision making from the model of rational choice to an approach that is based on knowledge that draws considerable experience on the part of the decision maker (Klein, 2008).

One of the key models of naturalistic decision-making is recognition-primed decision model (RPD). The recognition-primed decision model explores how people make good decisions without comparing options. The recognition-primed decision model seeks to describe how decision-makers use past experience using repository of patterns and mental simulations (Klein, 2008). The pattern recognition provides reference, context and highlights potential cues, acceptable goals and type of reaction to a typical situation. Within this framework, when people make decisions, they match and compare current decision situation with past situations, whenever there is a match, a course of action similar to previous decision situation is taken (Klein, 2008).

Pattern recognition is not the only way people make decisions in recognition-primed decision model; decision-makers also use mental simulations to evaluate possible options without necessarily comparing the options. Mental simulation is engaged to see how a decision situation will play out given current circumstances; a course of action is then initiated if mental simulation is deemed to work. Otherwise, the decision-maker tries to improve or adapt mental simulation or consider another cause of action that is satisficing enough. Satisficing entails choosing an alternative option that exceeds a given target or a set of criteria (March, 1994; Klein, 2008). In satisficing, alternatives are compared with a target until an alternative that is good enough is located; this is in sharp contrast to the rational model of choice which stipulates that decision-makers maximises and chooses the best possible alternative rather than satisfice (March, 1994). The recognition-primed decision model combines intuition with analysis, having pattern matching as the basis for intuition and simulation being the deliberate, conscious and analytical element of the model (Azuma et al, 2006; Klein, 2008).

A widely used decision model in high-stake decision situation is *observation, orientation, decision* and *action*, also referenced as OODA loop. In OODA loop, the observation element entails making an observation of the overall situation or simply, situation awareness. The orientation element of the model involves judging a situation to understand what it means or entails. The decision and action elements involve execution (decision making) and monitoring of the decision (Azuma et al, 2006). Figure 2-3 illustrates an adapted model of recognition-primed decision model.

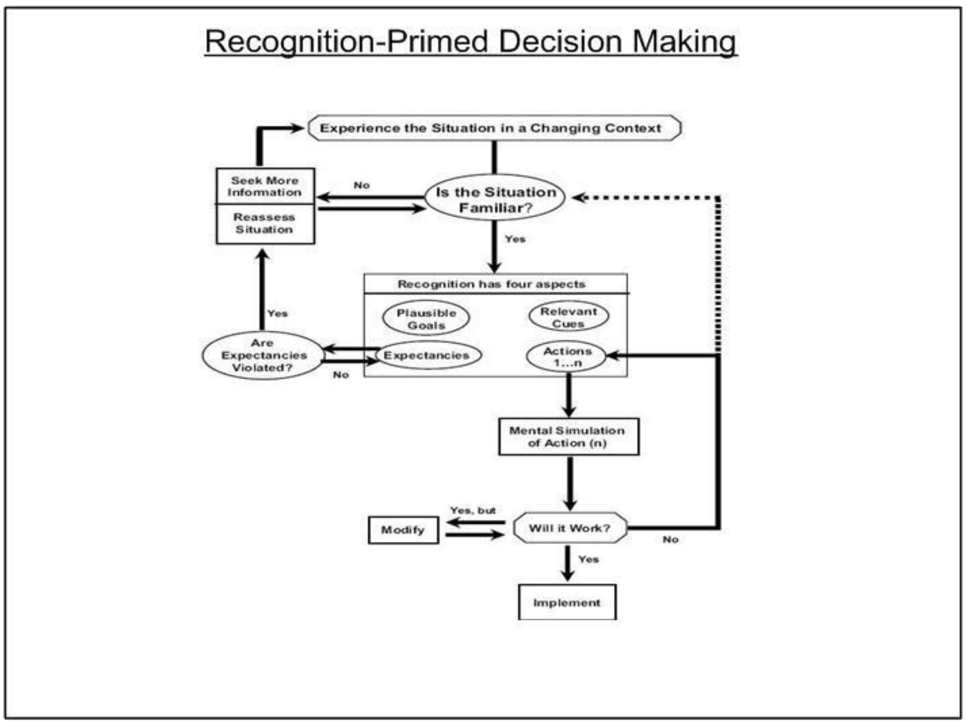

Figure 2-3: Recognition-Primed Decision Making (*Adapted from Klien, G., 2008*)

In the above recognition-primed decision model, as decision-maker experiences a dynamic decision situation, he calls on his or her prior experience to determine if the situation is familiar, if it is not, he or she seeks more information. If the situation is familiar, the pattern present the relevant cues, identify plausible goals, provide expectancies and suggests the appropriate type of reaction for current situation, mental simulation is engaged to see how the situation will play out (Klein, 2008). If the course action is deemed not workable, an alternative course of action is sought, this continues until a workable action is identified. If a workable action requires further modification, it is modified and finally when an acceptable course of action is identified the course of action is implemented and a decision is made (Klein, 2008). In naturalistic decision making, a decision-maker calls on past experience when confronted with a decision situation. In recognition-primed decision model, experience is a source of wisdom in decision making; consequently, a decision-maker adopts and executes the first solution deemed workable instead of spending time evaluating many different options to obtain an optimal decision (Klein, 2008). The next Section examines the propositions for this study.

2.6 Proposed factors influencing the choice of logical data model in data warehousing

In implementing a data warehouse, a logical data model engages an architecture principle that maps a conceptual phase to the real world of available software and processes. The logical data model of a data warehouse reflects the decisions that an organisation has made reflecting its principles and values (Sperley, 1999). The proposed factors influencing the choice of logical data model for a data warehouse are the factors that must be carefully considered before the physical implementation of a data warehouse is embarked upon in an organisation. The proposed factors are the summary of the literature and the preliminary interviews conducted with the research participants in the case organisations. The empirical data from the study is used to examine the degree to which data supports the proposed factors influencing the choice of logical data model for a data warehouse. The proposed factors influencing the choice of logical data model for a data warehouse are:

- High performance of query execution is a factor influencing the choice of logical data model for a data warehouse (Golfarelli and Rizzi, 1998; Kimball, 1995; 1977; Sen and Sinha, 2005)
- Focus on specific or generic functionalities is a factor influencing the choice of logical data model for a data warehouse (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003; Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998; Kimball and Ross, 2004)
- The goal and scope of the data warehouse is a factor influencing the choice of logical data model for a data warehouse (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003; Kimball and Ross, 2002; Kimball and Ross, 2004; Reeves, Ross, Thornthwaite and Kimball, 1998)
- The implementation orientation of available resources is a factor influencing the choice of logical data model for a data warehouse (Hayen et al, 2007; Vassiliadis, 2004)
- Consolidation of enterprise data to meet heterogeneous reporting and analytics requirements is a factor influencing the choice of logical data model for a data warehouse (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003)

2.6.1 High performance of query execution is a factor influencing the choice of logical data model for a data warehouse

One of the decision factors that impact the choice of logical data model for a data warehouse is the performance of a data warehouse. Improving the performance of a data warehouse is discussed in the literature (Golfarelli and Rizzi, 1998; Kimball, 1995; 1977).

It is important to address the performance requirements of the business users and a multidimensional data model is often used as a way to achieve high performance in data warehousing. However, the performance requirement of a multidimensional data warehouse is considerably different from the performance requirement of a relational data warehouse oriented towards an enterprise. High performance of query execution is an important consideration for a data warehouse addressing the generic requirements of an enterprise. In data warehousing, decision support queries requires significant aggregation and joining, in order to improve performance, de-normalisation is usually promoted in multidimensional environment than normalisation of the subject entities (Sen and Sinha, 2005). This is because a multidimensional data model is seen as the technique that presents data in a standard intuitive framework that allows for high performance (Kimball, 1995; 1977). As a result, the most important performance consideration in data warehousing concerns the optimal selection of database objects such as indexes, which is based on the logical data model adopted for a data warehouse (Golfarelli and Rizzi, 1998).

2.6.2 The degree of focus on specific or generic functionalities is a factor influencing the choice of logical data model for a data warehouse

The degree of focus on specific or generic functionalities is another important decision that must be considered that impacts the choice of logical data model for a data warehouse. The degree of focus on functionality is the extent to which a data warehouse is commissioned to engage a particular task or sets of tasks (Inmon, 2005; Inmon, Imhoff and Sousa, 2001; Kimball and Ross, 2002; Sperley, 1999; Reeves, Ross, Thornthwaite and Kimball, 1998).

The CIF framework presupposes an organisation implements a data warehouse spanning the entire enterprise although it is recommended that this be carried out in phases, however, building an enterprise data warehouse is a long and complex process requiring extensive data modelling which may make take several years to succeed. In order to avoid long drawn out implementation, some organisations are settling for data warehouses based on multidimensional data model (Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998). One reason why an organisation may want to implement a multidimensional data warehouse is that it is faster to roll out than a data warehouse based on a relational data model (Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998; Sperley, 1999). Additionally, data warehouses based on multidimensional data models do not require enterprise-wide consensus (Chaudhuri and Dayal, 1997; Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998; Prakash and Gosain, 2003; Bonifati, Cattaneo, Ceri, Fuggetta and Paraboschi; 2001). A multidimensional data model is used to implement a data warehouse that supports the specific functionality of the business, as such, it is considered excellent for data processing and ensures good performance for complex slice and dice operations (Imhoff, Galemmo and Geiger, 2003; Kimball and Ross, 2002). Furthermore, a multidimensional data model is extendable to support wider view of an enterprise through conformed dimensions (Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998). In order to support the enterprise view of an organisation, conformed dimensions are built as persistent master data; this enables consolidation of dimensions in multidimensional data model supporting multiple fact tables within the bus architecture. Thus, bus architecture framework enables multidimensional data models to support the enterprise view of an organisation (Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998; Kimball and Ross, 2004).

On the other hand, a data warehouse built on a relational data model is engineered to supports the business rules of an entire enterprise. A relational data model is not oriented to the need of a particular application requirement or to address the need for specific functional analysis. Instead, it is designed to support an enterprise enabling the possibility of different types of reporting and analytics across the firm (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003). In this respect, where high degree of focus on functionality is expected from a data warehouse, it is considered reasonable to engage a multidimensional data model as the basis for a functional data mart. In this situation, the data warehouse is implemented on a relational data model (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003).

As discussed above, a multidimensional data warehouse is extendable to support the enterprise view of an organisation using conformed dimensions using bus architecture framework (Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998; Kimball and Ross, 2004). While a multidimensional data warehouse is characterised as a data mart in the CIF framework, a multidimensional data mart is considered a functional data warehouse in bus architecture (Inmon, Imhoff and Sousa, 2001; Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998; Kimball and Ross, 2004). In this instance, it is perfectly plausible that having considered all the factors relating to the criterion of functionality, a decision is made to engage a multidimensional or relational data model as the foundation for a data warehouse.

2.6.3 The goal and the scope of a data warehouse is a factor influencing the choice of logical data model for a data warehouse

One of the factors that impact the choice of logical data model for a data warehouse is the goal of a data warehouse. The goal of a data warehouse ultimately defines the scope of a data warehouse. The scope of a data warehouse where the goal of the data warehouse is oriented toward an enterprise is larger than the scope of a data warehouse where its goal is oriented to the need of a particular business area within an enterprise (Imhoff, Galemmo and Geiger, 2003). The multidimensional and relational data models assume the requirements for a data warehouse emerges from a business area and the whole of the enterprise respectively (Kimball and Ross, 2004; Imhoff, Galemmo and Geiger, 2003). However, the scope of requirements for an enterprise is considerable than the scope of requirements from a functional business area (Imhoff, Galemmo and Geiger, 2003). The differences in the scope of a data warehouse oriented to a functional area and that oriented to the entire enterprise has implications for the meaning of simple data concept within an organisation. For instance, different business areas see the business concept *counterparty* in different ways. In one business area it may represent a *customer*, in another, it may represent *an intermediary*; both are legitimate interpretation of counterparty. From enterprise perspective, the meaning of counterparty includes the *types* identified above and more (Imhoff, Galemmo and Geiger, 2003). From a critical analysis perspective, both multidimensional and relational data models support the enterprise view of an organisation (Inmon, 2005; Kimball and Ross, 2002; Reeves, Ross, Thornthwaite and Kimball, 1998). The difference is the stage where each supports the enterprise view of an organisation. For instance, a multidimensional data warehouse implements the enterprise view of an organisation via the bus architecture (Kimball and Ross, 2002). The bus architecture framework allows a multidimensional data model to capture the enterprise view of an organisation by joining conformed dimensions to multiple fact tables through the surrogate keys. However, a relational model is oriented to the enterprise view of an organisation right from the beginning of the logical data model at the inception of a data warehouse project (Inmon, Imhoff and Sousa, 2001; Inmon, 2005). A decision that clearly sets out the goal, hence the scope of a data warehouse is required; this decision directly impacts and influences the choice of logical data model for a data warehouse (Imhoff, Galemmo and Geiger, 2003).

2.6.4 The implementation orientation of available resources is a factor influencing the choice of logical data model for a data warehouse

The implementation orientation of available resources is another criterion that impacts the choice of logical data model for a data warehouse. Resources in this context include the development staff tasked with implementing a data warehouse, Vassiliadis (2004) notes that data warehousing landscape is defined by "*do-it-yourself*" advice from experts and proprietary vendor solutions. In the industry, there is an appreciable increase in the number of adopters of multidimensional data warehouses (Kimball and Ross, 2004). Underlying this phenomenon is a perception that data warehouses based on multidimensional data models are easier to implement and increases the likelihood of quick-wins for the management. In the literature, one-half to two-thirds of data warehousing efforts end in failure, it is not difficult to see why a risk averse implementation team will gravitate towards a logical data model they have previous experience and avoid costly experiments (Hayen et al, 2007; Mark, 2005). Objectively, there are situations where it is appropriate to adopt a multidimensional data model as part of a business intelligence landscape.

The common practice in most enterprise data warehouse implementations is to engage a multidimensional data model as the foundation for a data presentation layer where it is accessed directly by end users using reporting and analytical applications. However, end user reporting and analytical applications in most cases are not permitted to interface directly with an enterprise relational data warehouse (Inmon, Imhoff and Sousa, 2001; Inmon, 2005; Imhoff, Galemmo and Geiger, 2003). Given the high visibility nature of most data warehouses and the risk adverse nature of most implementation teams tasked with such undertaking, it is imperative to determine the type of logical data model that is the foundation of a data warehouse.

2.6.5 Enterprise Data consolidation to meet heterogeneous reporting and analytics requirements is a factor influencing the choice of logical data model for a data warehouse

Another criterion that affects the choice of logical data model for a data warehouse is the degree that a data warehouse is required to address different reporting and analytics requirements of diverse consumer groups within an organisation. One of the arguments often recognised as the basis for implementing a data warehouse is the need to capture diverse set of business data in a data warehouse (Inmon, 2005; Kimball, 2002; Tryfona, Busborg and Christiansen, 1999; Weininger, 2002, Husemann, Lechtenbörger, Vossen, 2002). Integrating complex enterprise data for various reporting and analytics requirements requires flexible data model that enables diverse subject area entities to co-exist within a data warehouse. Large organisations have complex business structures; however, complexities appear not to present a problem for a relational data model. From a practical standpoint, it is far less challenging and demanding to integrate multiple new business lines into a relational data warehouse than into a multidimensional data warehouse. This is because most relational data models are usually oriented to the enterprise view of an organisation and focuses on complex data structures and interactions (Imhoff, Galemmo and Geiger, 2003; Tryfona, Busborg and Christiansen, 1999). From an analytical perspective, multidimensional and relational data models are complementary and in some ways engage a similar approach in addressing the heterogeneous enterprise reporting and analytics requirements of an organisation; For instance, both multidimensional and relational data models support phased implementation approaches in developing large data warehouses. In a multidimensional data warehouse, the phase approach begins with a business unit initiating a data warehouse; an additional schema is layered on top of the existing multidimensional schema as more requirements are introduced using conformed dimensions in bus architecture framework; In this approach, a data warehouse is a 'union of star schemas' that models new business data. By using the bus architecture framework, a data warehouse is able to provide the enterprise view of an organisation enabling a data warehouse to address different consumer reporting and analytics requirements within an organisation (Kimball, 1997; Kimball and Ross 2002; Reeves, Ross, Thornthwaite and Kimball, 1998). Likewise, the CIF approach is iterative; this enables complex data warehouse implementations to be broken down into multiple implementation phases enabling the management and the business to see the benefits of their data warehouse at an early stage. In this approach, an enterprise data warehouse focuses on integrating one line of business at a time (Inmon, 2005). While the CIF approach needs to be iterative to have any chance of success, in practice, there is a degree of integration activities that is required to ensure compatibility between the different phases of the implementation cycles in order for a data warehouse to address different consumer reporting and analytics requirements of an enterprise.

2.6.6 Proposed Conceptual Model for Logical Data Model Selection

The conceptual model in figure 2-4 below is the proposed decision model for adopting a logical data model for a data warehouse. The inputs into the proposed conceptual decision model are the proposed factors influencing the choice of logical data model in data warehousing. These are:

- High performance of query execution
- Focus on specific or generic functionalities
- The goal and scope of the data warehouse
- The implementation orientation of available resources
- Consolidation of enterprise data to meet heterogeneous reporting and analytics requirements

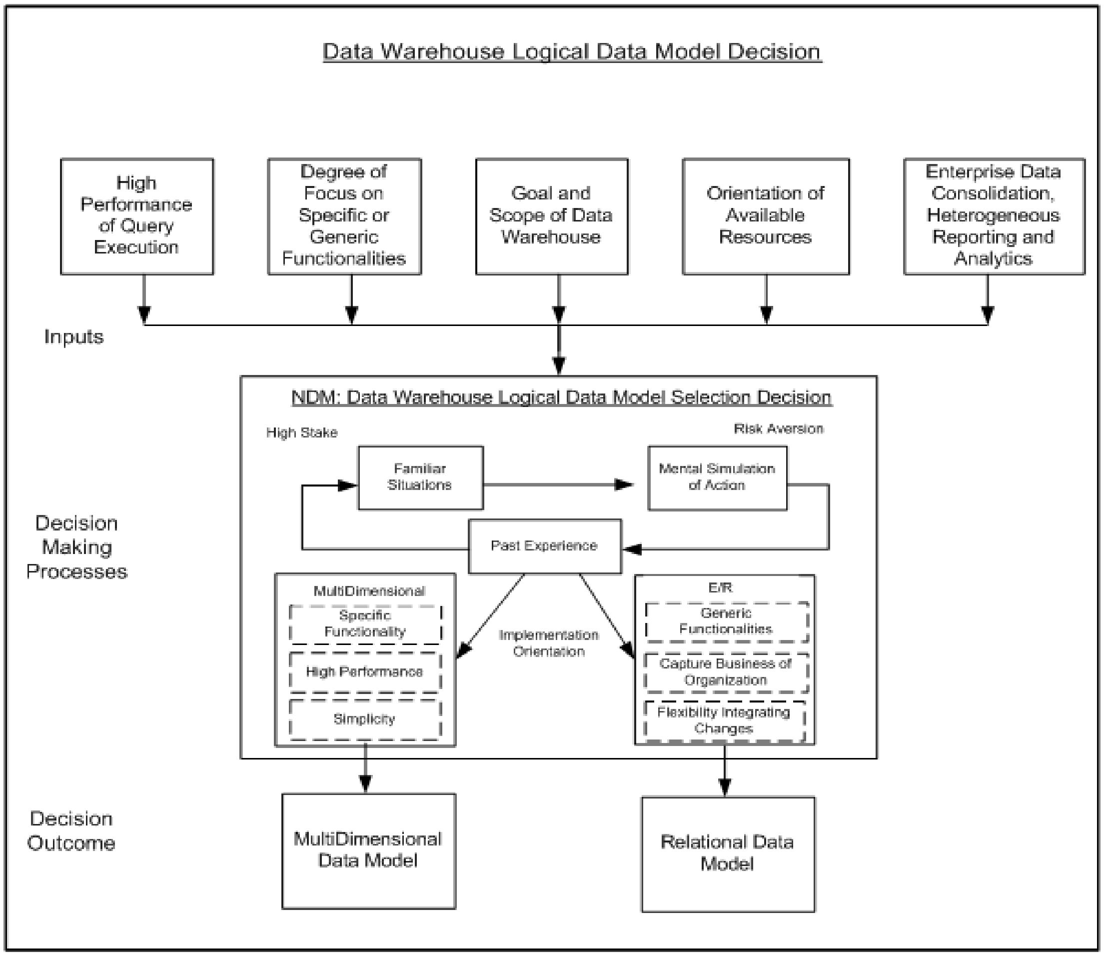

Figure 2-4: Proposed Decision Model - Logical Data Model Selection Decision

The proposed conceptual decision model is based on review of literature distilling the proposed factors influencing the choice of logical data model and the prescriptions of the naturalistic decision making theory. In alignment with the recognition-primed decision making of naturalistic decision making theory, decision-makers in the proposed conceptual decision model relied on their past experiences when making a decision to adopt a logical data model for a data warehouse. As discussed in Section 2.5.4, in this proposed conceptual decision model, mental simulations, situation recognition and categorisation are dependent variables encapsulated within the decision maker OODA loop; the outcome of the model is a decision pathway leading to an adoption of a logical model for a data warehouse.

2.7 Summary

This chapter presented a review of the literature on data architecture in data warehousing. The chapter examined the overview of data warehouse system in general and discussed the key components of a data warehouse system. The chapter identified and discussed the end-to-end components of a data warehouse system including the data acquisition and presentation layers. The chapter further explored the industry approaches for implementing data warehouses and engaged the literature to discuss the implementation phases of a typical data warehouse system. The chapter examined decision-making from a theoretical perspective and explored the classical theories of choice, in so doing, the chapter examined the normative, descriptive and prescriptive theories of decision making and further explored the naturalistic decision making theory.

The chapter further outlined and discussed a number of propositions that are considered the factors influencing the choice of a logical data model in data warehousing. Finally, the chapter presented the proposed conceptual decision model for logical data model adoption in data warehousing. Based on naturalistic decision making and its recognition primed decision model, the proposed conceptual decision model implements a decision pathway for logical data model adoption in data warehousing.

CHAPTER 3 - Research Methodology

3.1 Introduction

Social research methodologies are divided mainly into quantitative and qualitative methods. Quantitative research use statistical methods and is characterized by direct manipulation of independent variables. A research using qualitative methods use words to describe trends and patterns to explain social phenomena. In pursuing the aim and objectives of this study, a number of qualitative research methods are considered including action research, survey and grounded theory, with case study method at the centre of the research effort for this study. Case study method is an empirical enquiry that investigates contemporary phenomenon in its real-life context and focuses on understanding the dynamics in a setting. An explanation of the different types of case study methods is explored and the suitability of case method for this research is examined. The research design for this study together with the method of data collection and the components of the research framework are documented.

This chapter is divided into a number of Sections addressing key topics discussed in the chapter. Section 3.2 looks at the overview of qualitative research and engages the literature to discuss the philosophy underpinning different research paradigms. Section 3.3 looks at various forms of post-positivist research methodologies and Section 3.4 examines the reasons for using case study method for this research. Section 3.5 contrasts single case and multiple case methods and examines when and why it is appropriate to use these methods. Section 3.6 examines the research design framework for the study and Section 3.7 concludes with a summary.

3.2 Overview of Research Methodologies

An exploration of philosophy is significant for three reasons with particular reference to research methodology (Crossan, 2003). First, exploration of research philosophy helps a researcher to specify and refine the research methods to use in a study and clarifies the overall strategy to engage a study. This impacts the type of evidence gathered, its origin, the way in which the evidence is interpreted, and how it helps to answer the research question(s). Second, knowledge of research philosophy enables and assists a researcher in evaluating different methodologies and avoids unnecessary work by identifying the limitations of a particular approach at an early stage. Third, exploration of research philosophy helps a researcher to be creative and innovative in selecting and adapting methods that were previously outside the researcher's experience. A research method can be described, examined and categorised at different levels, the philosophical level focuses on assumptions relating to the most general features of the world, encompassing aspects such as mind, matter, reason and proofs for knowledge. The philosophical aspects underpinning methods and facilitate the classification of research methods into paradigms (Clark, 1998). One of the most pronounced features of contemporary social research is the range of research perspectives that operates concurrently. Social science research is marked by plethora of school of thought, each with its own theoretic assumptions, research methodologies, and adherents (Orlikowski and Baroudi, 1991). The social phenomena studied in social research are complex; the existence of a range of perspectives allows an exploration of phenomena from diverse frames of reference. Orlikowski and Baroudi (1991) points out the dominant perspective in information system research involves discussion of the status of information system research vis-à-vis the norms of what constitutes a scientific discipline.

The indicator of a research tradition is the extent to which there exists a set of dominant philosophical assumptions or a worldview that informs the work of researchers in a discipline. Social research methodologies are divided into two broad categories namely: quantitative and qualitative methods. Quantitative research is characterized by direct manipulations of independent variables combined with random assignment of participants to groups (Hancock and Algozzine, 2006). Quantitative research uses numbers, normally in the form of statistics to explain phenomena. Qualitative research on the other hand does not involve direct manipulation of variables; rather, words are used to describe trends and patterns to explain social phenomena. Apart from making a simple distinction of the use of measurement or description as the main approach in collecting and analysing data, quantitative and qualitative research possess the characteristics outlined in Table 3.1

Key Attributes	Quantitative Research Description	Qualitative Research Description
Orientation	Quantitative research uses inductive approach to test theories	Qualitative research uses deductive approach to generate theories
Epistemology	Quantitative research is based on a positivist approach inherent in the natural sciences	Qualitative research rejects positivism by relying on individual interpretation of social reality
Ontology	Objectivist because social reality is regarded as objective fact	Qualitative research is constructionist because social reality is seen as a constantly shifting product of perception

Table 3-1: Characteristics of Qualitative and Quantitative Research, Walliman (2006)

The degree of control an investigator intends to exert on research processes influences the choice of research methods (Yin, 2009). A method in relation to research at the most general level means epistemology or the study of how we know things. This requires making strategic choices ranging from whether to carry out experiments or perform participant observation fieldwork. At a specific level, method is about technique, that is, what kind of sample to use, whether one decides to engage face-to-face interview, conduct telephone surveys or carry out experiments (Bernard, 2000). The consistency between the aim of a research, the research questions, the chosen methods, and the personal philosophy of the researcher is essential and underpins the rationale for any research project (Clark, 1998). In order to make any decision regarding the method to engage a research, an understanding of the two extremes of research philosophy i.e. *positivism and post positivism* should be explored and understood.

3.2.1 Research Philosophy – The Positivist Paradigm

Ryan (2006) notes using scientific method and language to investigate and write about human experience is meant to keep research free of values, passions, politics and ideology of the researcher, this approach to research is the positivist paradigm. The structure supporting positivism is a modernist worldview. In modernism, only certain and empirical knowledge are valid, and rational is valued over other ways of knowing such as intuition. Positivism seeks to reduce everything to abstract and universal principles, and tends to fragment human experience rather than treat it as a complex whole (Ryan, 2006).

Positivist social researchers look for uniform, precise rules that organised social behaviour. Positivists in the social and behavioural sciences examine simplified models of the social world to see how a small number of variables interact. The language of positivism is a numeric one; the goal is a series of statistical equations that explains and predict human behaviour (Rubin and Rubin, 2005). The natural science approach emphasises universal laws of cause and effect based on explanatory framework, which assumes realist ontology; that is, reality consists of a world of objectively defined facts. The hypothetico-deductive method is the principal means by which causal relationships are established. In this approach, the scientist's ideal strategy is the experimental control of subsets of variables in the service of testing, either verification or falsification of prior theory (Hammersley, 1993).

The basic belief system of positivism is rooted in a belief that there exists a reality out there, driven by immutable natural laws. Positivism is constrained to practice an objectivist epistemology, if there is a real world operating according to natural laws, then the inquirer must behave in ways that put questions directly to nature and allow nature to answer back directly (Guba, 1990). Much of the work of the natural scientist concerns the methodological detailing of operationalisation and measurement. Quantification is the sum of standardization, measurements and numbers are crucial to the natural science approach, because they renders the concepts in theoretical schemes, observable, manipulable and testable (Hammersley, 1993). This is also the necessary, if not always the condition for the findings of research to be replicable and generalisable (Hammersley, 1993).

The critics of positivist epistemologies argue that division between objectivity and subjectivity, public and private knowledge, scientific and emotional knowledge is socially constructed. Dualistic ways of viewing the world are used to control ideas about what knowledge is legitimate; knowledge cannot be divorced from ontology (Ryan, 2006). The proponents of the alternative research philosophies such as critical and post-modern theory argue that the goal of research is learning about contingent truth, truth that seems to hold at a particular time under specified circumstances. Rather than assume neutral researcher, alternative paradigm assumes researchers' ideas and personality affects research and that the aim of research is to bolster social or political agenda (Rubin and Rubin, 2005).

3.2.2 Research Philosophy – The Post Positivist Paradigm

The alternative epistemological position is expressed in the naturalistic or interpretive paradigm. This is the result of the critique of positivist scientific method as the sole basis for understanding human activity. The naturalistic paradigm draws on Dilthey's view (1977) that a clear distinction should be drawn between the disciplines of natural science and human science (Hammersley, 1993). While external observation and explanation of regularities in the physical events could direct the natural science, Dilthey (1977) argue that human sciences should be premised upon a search for meaning or understanding. The naturalistic paradigm is described by a number of characteristics including a commitment to constructivist epistemologies, emphasis upon description, explanation and representation of reality through the eyes of the participants. The importance of viewing meaning of experience and behaviour in the context and full complexity is one of the defining characteristics of naturalistic paradigm.

So also is a view of scientific process generating working hypothesis rather than immutable empirical facts and an attitude towards theorising which emphasises the emergence of concepts from data rather than their imposition in terms of *a priori* theory (Hammersley, 1993). Researchers that follow the naturalistic paradigm are critical of the application of scientific model to the study of social world and have been influenced by different intellectual traditions. Naturalistic researchers share a view that the subject matter of social sciences (*people and their institutions*) is fundamentally different from that of the natural sciences (Bryman and Bell, 2007). This led to methodological pluralism in information research, rather than seeing the social world from strictly positivist and post-positivist interpretive paradigm. Kaplan and Duchon (1998) points out that combining quantitative method and qualitative method introduces testability and context into research. Mingers (2001) suggests that research is richer and more reliable if different research methods preferably from exiting paradigms are routinely combined together, rather than advocating a single discipline be it positivism or post positivism. The next Section examines the various types of qualitative research methods and in particular, the research method for this study.

3.3 Qualitative Research Methods

Case study is a research method for conducting empirical investigation of contemporary phenomenon in its natural context using multiple sources of evidence (Hancock and Algozzine, 2006). It is a study in which cases in real life context are selected and scored in a qualitative manner (Dul and Hak, 2008). A case study is a history of past or current phenomenon, drawn from multiple sources of evidence. It includes data from direct observation and systematic interviewing as well as from public and private archives. Any fact relevant to the stream of events describing the phenomenon is a potential datum in a case study since the context is important (Voss, Tsikriktsis and Frohlich, 2002). Gerring (2007) defines case study as an intensive study of a single or small number of units (*cases*) for the purpose of understanding a larger class of similar units. Dul and Hak (2008) considered case study as a useful research strategy when the topic under investigation is broad and complex, when there is not a lot of theory available and the context is very important. Research questions that ask "*why*" and "*how*", that require no control of behavioural events and focuses on contemporary events are suited to case study method (Yin, 2009). Case study research is associated with description and theory development, where it is used to provide evidence for hypothesis generation and exploration of areas where knowledge is limited (Cavaye, 1996). Apart from case study method, there are other forms of research methods that come under post positivist paradigm. Table 3-2 below presents the different types of post positivist research methods in social research.

Research Method	**Description**
Action Research	Action research is mostly effective in an organisation setting as a practical way of dealing with organisation problems by means of mobilizing and involving social science in a specific manner. In action research, the question presented, which is the object of the study is jointly addressed by the researcher and the researched (Beinum, 1999)
Archival Research	Archival research is recording behaviours from historical records. Archival research strategy allows research questions,

	which focuses upon changes over time to be answered. It involves looking at historical records or archives to measure behaviours and events that occurred in the past (Gravetter et al, 2009)
Content Analysis	Content analysis involves using the techniques of behavioural observation to measure the occurrence of specific events in the literature, movies, television programs or similar media that presents replica of behaviours (Gravetter et al, 2009)
Survey Method	Surveys are based on desire to collect information usually by questionnaire from a sample of respondents from a well-defined population (Czaja and Blair, 2005)
Grounded Theory	Grounded theory explores social processes present in human interactions. This research method is different from other qualitative research methods because the primary purpose of using grounded theory is to develop a theory about dominant social processes rather than to describe a particular phenomenon (Speziale, Streubert and Carpenter, 2007)

Table 3-2: Qualitative Research Methods in Social Science (*Adapted from Beinum, 1999; Gravetter et al, 2009; Czaja and Blair, 2005; Speziale, Streubert and Carpenter, 2007*)

The next Section looks at the reasons for using case study for this research.

3.4 Reasons for using case study research method

Case study research is associated with description and theory development, it is used to provide evidence for hypothesis generation and exploration of areas where knowledge is limited (Cavaye, 1996). Case study is an empirical enquiry that investigates contemporary phenomenon in its real-life context, especially when the boundaries between phenomenon and context is not clearly evident and relies on multiple sources of evidence (Shanks, 2002). Case study is a research strategy that focuses on understanding the dynamics in a setting. It combines data collection methods such as archives, interviews, questionnaires, and observations. Case studies is used to accomplish various aims including providing description, testing or generating theory (Eisenhardt, 1989).

Stake (1995) points out the real business of case studies is particularization, not generalization, a case is chosen and the researcher come to know it well, not primarily to know how it is different from others, but what it is and what it does. The emphasis is on understanding the case itself. A number of studies have used case study method to engage research, this included Allison (1971) study of Cuban missile crisis; Yin (2003) description of case study of neighbourhood organisation and Selznick (1949) description of Tennessee Valley Authority (TVA). Case study research is an intensive study of a single or small number of units (*cases*) for the purpose of understanding a larger class of similar units (Gerring, 2007). Although natural sciences and naturalistic research are legitimate research methods of enquiry, it is important to determine the relevance and appropriateness of a research method to the phenomenon under study (Gillham, 2000).

The choice of research method is affected by the type of research question being investigated, it is also affected by the degree of control a researcher has on the process and the degree of focus on contemporary as opposed to historical events (Yin, 2009). Applying the above prescriptions against this research produces the following observations:

- The focus of this research is to study an empirical phenomenon. This research intends to study, explore and explain why pluralistic data architecture exists in data warehousing. In so doing, the research aims to examine the reasons behind this phenomenon based on the research propositions.
- Case study method is appropriate for a research where the investigator is not aiming to influence the research. The focus of this research is not to influence the phenomenon under study; rather, it aims to gain an understanding and examine the reasons behind the phenomenon under study. This research is not attempting to influence the decision to adopt a particular logical data model for a data warehouse, be it multidimensional or relational data model. Rather, the research aims to explore the factors behind the reasons for adopting the existing data models in the case organisations.

- The choice of research method to conduct an empirical study is affected by the degree of focus on contemporary as opposed to historical events (Punch, 2006). This research focuses on empirical phenomenon in the observed increase in the number of organisations adopting both the multidimensional and relational data models for their data warehouses. The research aims to explore and examine the factors behind the decision to adopt these data models in the research units of analysis.

- Case study method is appropriate for studies where the phenomenon being researched is studied in its natural setting (Hancock and Algozzine, 2006). This research is conducted within the organisations that implemented the multidimensional and relational data model respectively for their data warehouses. This allows the investigator to engage with the actors and stakeholders that took part in the decision to adopt the data models under investigation.

The next Section looks at the research design for this study.

3.5 Research Design Framework

Research design is the blueprint of a case study research (McCoy et al, 1993). It is a plan for assembling, organising and evaluating information according to the problem definition and specific goals for how to use a research finding. This Section focuses on the research design for this study; in particular, it discusses the research methodology to engage this study and in particular the components of the research design framework.

3.5.1 Components of Research Design Framework

Figure 3-1 below outlines the components of this research framework, each of the research components aligns with a chapter in this thesis.

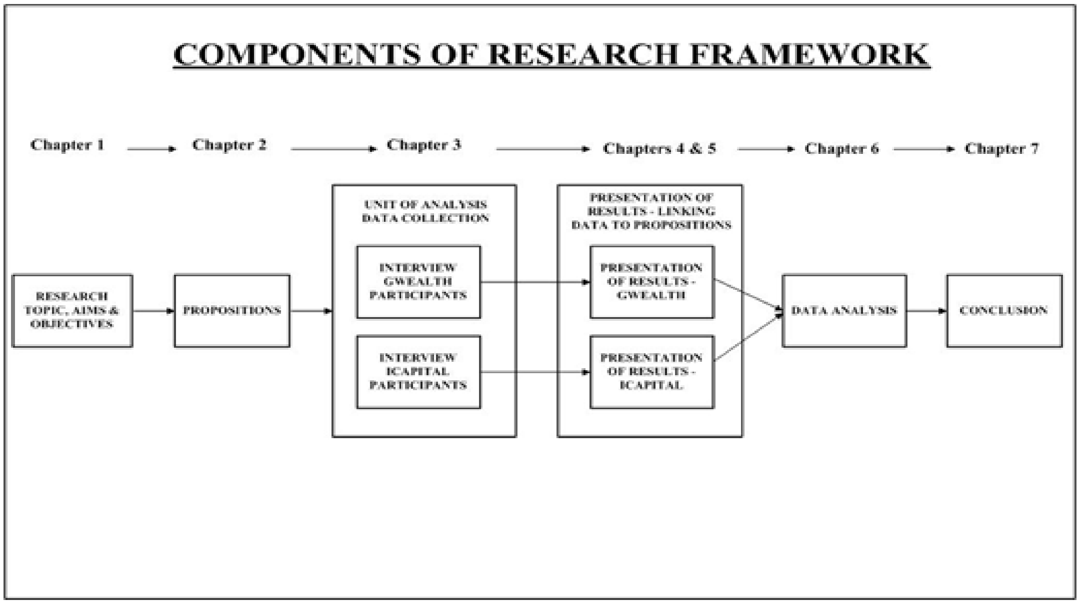

Figure 3-1: Research Design Framework (*Author Proposed*)

3.5.2 Research Topic, Aim and Objectives

The research design framework provides the pathways for conducting this study. The research topic provides the basis for the study and is the link between the research framework and the research question. The framework is organised around the research question; the research question determines how the participant interviews are organised and conducted within the study unit of analysis. This research investigates pluralistic data architectures in data warehousing. The aim of this research is to examine the decision factors influencing the choice of logical data model for data warehouses in the industry. The objectives of the research are to:

- Review the state-of-the-art in data warehouse development in order to distill the factors affecting the choice of data model in data warehousing.

- Synthesise the literature in order to try and develop a framework impacting the choice of logical data model in data warehousing.
- Undertake comparative case study in the organisations using the multidimensional and relational data models to test the factors impacting the adoption of logical data model for a data warehouse
- Evaluate the outcome of the study to determine the alignment of the decision factors in practice and contributions to theory

3.5.3 Research Propositions and Theory Development

This thesis make a number of propositions which are developed from the literature and data collected from preliminary interviews of research participants. The research topic guides the literature review on data architecture in data warehousing, the research method and data collection instruments to examine the research propositions. The research propositions are the factors that the investigator considered influenced the decision to adopt multidimensional and relational data models respectively in the case organisations. This research collects empirical data to test the validity of the proposition by critically examining the degree to which the empirical data supports or disproves the research propositions. Table 3-3 outlines the research proposition for this study.

Research Proposition (Abbreviated)	Research Propositions - Factors affecting the choice of logical data model for a data warehouse
RP1	High performance of query execution is a factor influencing the choice of logical data model for a data warehouse
RP2	Focus on specific or generic business requirements is a factor influencing the choice of logical data model for a data warehouse
RP3	The goal and scope of the data warehouse is a factor influencing the choice of logical data model for a data warehouse
RP4	Implementation orientation of available resources is a factor influencing the choice of logical data model for a data warehouse
RP5	Consolidation of enterprise data to meet heterogeneous reporting and analytics requirements is a factor influencing the choice of logical data model for a data warehouse

Table 3-3: Outline of the Research Propositions

3.5.4 Unit of Analysis and Data Collection

This research is a comparative case study; it adopts two-case approach to address the research question. A research designed to test propositions requires more than one unit of analysis (Bernard, 2006). The use of multiple cases provides the basis for collecting and matching evidence that ultimately increases the confidence in the outcome of a research (Yin, 2009). The units of analysis for this research are GWealth and ICapital. These organisations are wealth management and investment banking divisions of tier 1 European global banking institution, TBank. The research question focuses on these organisations and provides the basis for data collection instruments for the study.

The data model for the GWealth reporting data warehouse is based on a multidimensional data model, this is often called the Kimball approach. A multidimensional data model provides the foundation for Online Analytical Processing (OLAP) and its derivatives including ROLAP and MOLAP. The primary data model for ICapital Risk and Finance data warehouse (RFDW) is a relational data model. The data model aligns with the so-called Inmon approach. The primary mode of data collection for this study is through the interview of research participants at GWealth and ICapital. The interview questions were derived from the research question, literature and the preliminary interviews conducted with the research participants. The research question provides the basis for collecting the empirical data to test the validity of the research propositions. For this study, the interview questions were addressed to manager grade staff at GWealth and ICapital. In total, twenty-one participants from GWealth and ICapital took part in the study. Table 3-4 presents the breakdown of the research participants for this study.

Organisation	Title of Research Participant	Area of Responsibility	Number of Participants
GWealth	Director	Reporting Architecture	2
	Program Manager	Program Management	1
	Project Manager	Program Management	1
	Development Manager	Development	2
	Senior Business Analyst	Development	1
	Senior Developers	Development	3
ICapital	Director	Enterprise Architecture	2
	Program Manager	Program Management	1
	Project Manager	Program Management	1
	Development Manager	Development	2

	Senior Business Analyst	Development	2
	Senior Developers	Development	3
Total Research Participants from GWealth & ICapital			**21**
Total Research Participants from GWealth			**10**
Total Research Participants from ICapital			**11**

Table 3-4: Research Participants from GWealth and ICapital Case Organisations

The key stakeholders outlined above are the source of the empirical data for this research, their roles in the context of the organisation structure is briefly described below:

- Director Architects: Both GWealth and ICapital operates an organisation structure where director grade staff heads a department or technology function; a director holds and controls spending activities within his/her function and is part of the funding group. All technology development initiatives that meets the need of the business requires a director's approval, therefore, a director has a stake in any project requiring his or her approval. Four director architects were part of this study, two each from GWealth and ICapital.
- Program Managers: Two program managers, one each from GWealth and ICapital were part of this study. The program managers were part of the funding group at GWealth and ICapital
- Project Managers: Two project managers, one each from GWealth and ICapital were part of this study. The project managers were part of the execution team responsible for developing the data warehouses at GWealth and ICapital

- Development Managers: Four development managers, two each from GWealth and ICapital were part of this study. The development managers were part of the execution team responsible for developing the data warehouses at GWealth and ICapital
- Business Analysts: Three senior business analysts, one from GWealth, two from ICapital were part of this study. The business analysts were part of the execution responsible for developing the data warehouses at GWealth and ICapital
- Developers: Six senior developers, three each from GWealth and ICapital were part of this study. The developers were part of the execution team responsible for developing the data warehouses at GWealth and ICapital

The primary focus of this research is logical data model adoption for a data warehouse, a decision was made not include the end-users in this research based on the preliminary response from the user community. This was because the users indicated '*they prefer to leave the structural design issues to the development team and that it is up to the development team to deliver the data required by the business'*. On that basis, the focus of this research is purely on the technical community. The interviews of GWealth and ICapital staff were conducted in parallel stream enabling the investigator to collect empirical data specific to each case, based on work schedule and availability of the research participants. This approach enabled the investigator to engage in early analysis of the data to determine if additional follow-up interviews are required to collect more data.

3.5.5 Presentation of Results and Linking Data to Propositions

Chapters 4 and 5 of the thesis are devoted to presentation of results and linking of empirical data from GWealth and ICapital cases to the research propositions. The main logic for linking the empirical evidence to the research propositions is through analysis and interpretation of the field data. Data analysis consists of transcribing; coding, categorising, tabulating, comparing and contrasting the evidence to draw empirically based conclusions. This research engages grounded theory approach as the method of analysing the empirical data for this study; there are two schools of thought in grounded theory. One school of thought takes the position that a researcher should have an empty mind – *Glaser school of thought*. The other school of thought permits a researcher to have a general idea of the area under study –*Strauss school of thought* (Jones and Alony, 2011). This study aligned with the Strauss school of thought that permits a researcher to have a general idea of the area under study; these ideas are expressed as the research propositions. Figure 3-2 illustrates the steps to link the empirical data to the research propositions.

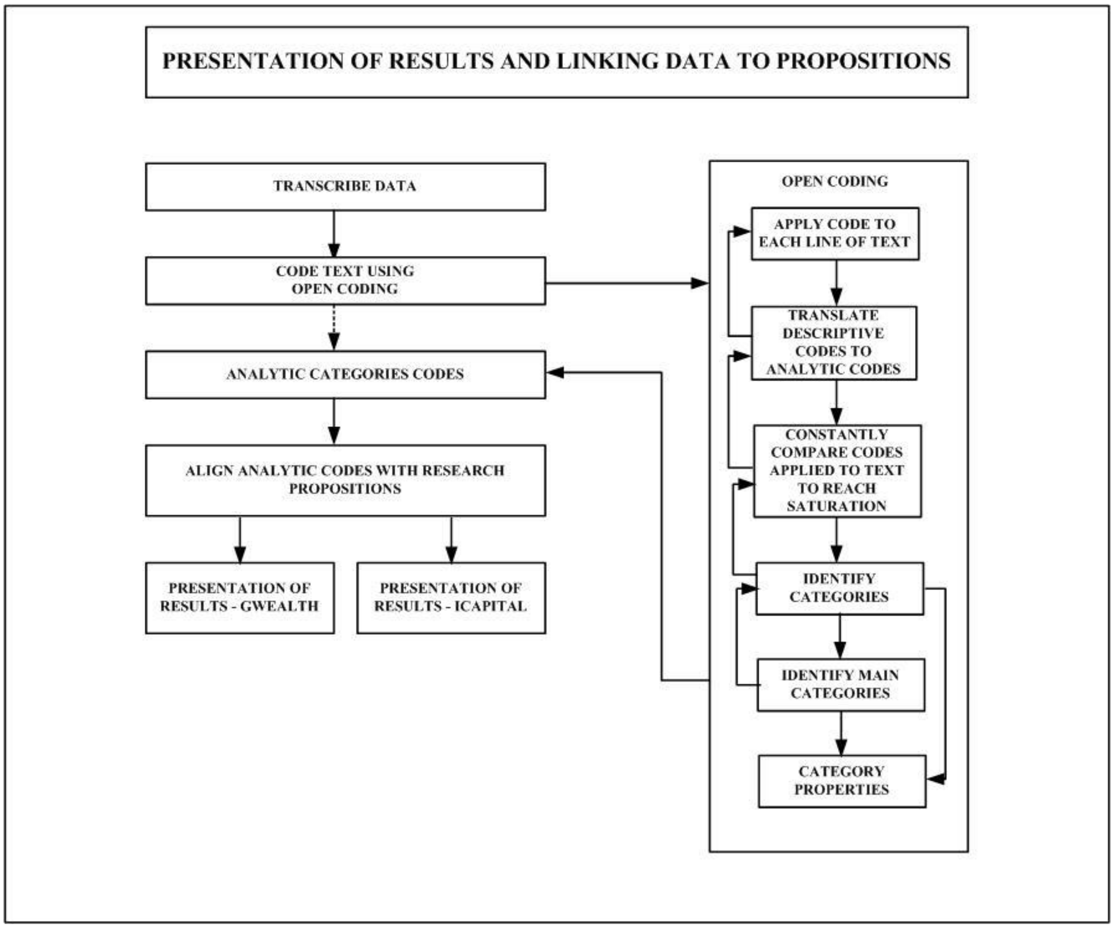

Figure 3-2: Presentation of Results and Linking Empirical Data to Propositions (*Adapted from Graham R. Gibbs (2007)*)

3.5.6 Transcribe Data

A recording device was used to capture the interview sessions with a research participant. Following each interview, empirical data was transcribed into a readable text using Microsoft Word. The investigator carried out the manual work of transcribing the interview data, this ensured data quality is maintained and the investigator is familiar with the data at early stage. Microsoft Excel and Microsoft Word were used as the data analysis software for this study.

3.5.7 Apply Code to Text

Table 3-5 outlines the steps for linking data to propositions using grounded theory open coding method.

Step	Description of Open Coding Activities
1	Applied descriptive codes to every line of the text – This step involved assigning codes for every row of transcribed data; the code for a line of text describes and captures the essence of the line of text from the perspective of the research participants.
2	Translated the descriptive codes into analytical codes – This step involved transforming the descriptive codes in step 1 with codes that introduced analytical context to each coded data.
3	Compared codes to ensure similar texts and passages were applied similar codes for consistency and saturation – This step involved going over the empirical data and ensuring data with similar themes are coded in the same way.
4	Identified categories amongst the analytical codes in MS Excel – This step involved collating the analytic codes in MS Excel and identified a category which a set of analytic codes aligned.
5	Identified main analytic category codes in MS Excel – This step involved creating top level or main category codes under which the category codes (developed in step 4 - above) rolls up, thus, the category codes in step 4 became sub-category codes and rolls up to main category codes.
6	Identified category properties – In this step, sub category codes in step 4 were transformed as properties of the analytic main category codes in step 5 - above.
7	Created data frequency count of the analytical codes in MS Excel –This step involved counting and aggregating the number of times an analytic code appeared in the empirical data in MS Excel.

Table 3-5: Steps Linking Empirical Data to the Research Propositions

3.5.8 Align Analytic Categories (Analytic Codes) with Propositions

Table 3.6 outlines the steps used to align the analytical codes with the research propositions

Step	Description of Activities Linking Analytic Codes with Propositions
1	Created aggregated count of frequency for the analytic codes
2	Created frequency analysis table for the analytic codes
3	Calculated the cumulative numbers and cumulative percentages of each of the analytic codes and presented the details in tabulated frequency analysis table
4	Created *Pareto* chart of the frequency analysis table and opened discussion on presentation of result in the result chapters – PRESENTATION OF RESULTS (GWEALTH), PRESENTATION OF RESULTS (ICAPITAL) chapters 4 and 5 respectively
5	Presented and discussed the propositions in the result chapters. Comparative frequency analysis of the analytic codes is presented and discussed
6	Each analytic code is discussed in relation to the proposition it supports. *An analytic code with similar theme as a proposition is judged to align with the proposition. If an analytic code aligns with a proposition and the analytic code has the highest percentage relative frequency compared to other analytic codes that proposition is judged as supported by data*

Table 3-6: Steps Linking Analytic Codes with the Research Propositions

3.5.9 Data Analysis

Data analysis is a process of moving from raw interviews to evidence-based interpretations that are the foundation of published report. Data analysis entails classifying, comparing, weighing and combining materials from interviews to extract meaning and implications, to reveal patterns or to align the descriptions of events into a coherent narrative (Rubin and Rubin, 2005).

Figure 3-3 below outlines the data analysis steps for this study.

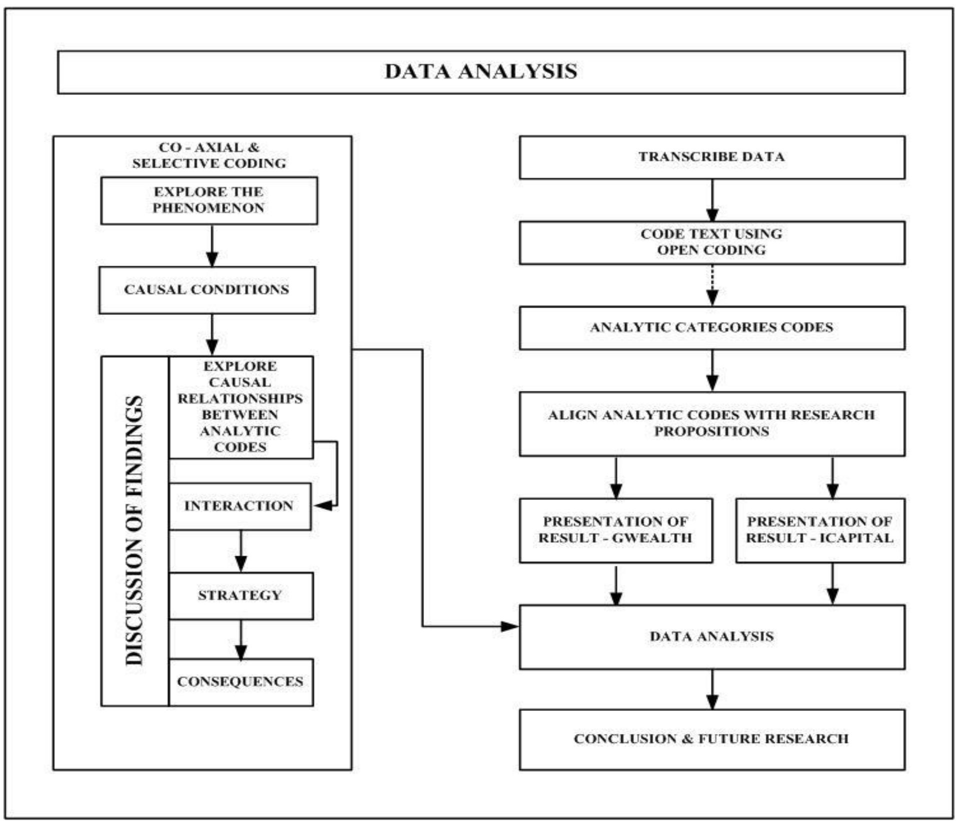

Figure 3-3: Data Analysis (*Adapted from Graham R. Gibbs (2007)*)

The investigator explored and discussed the causal relationships between the analytic codes presented in chapters 4 and 5. The discussion addressed how the analytic codes connected to each other. Sub categories and category properties were used to drive the discussions of the relationships between the analytic category codes. As part of the data analysis, the investigator examined each of the analytic code variables in detailed and presented the ratings of the each of the analytic codes for GWealth and ICapital cases. Additionally, the findings of the research are discussed; using Chi Square statistical method, the degree of significant relationship between the research code variables and the logical data models is explored and discussed. Furthermore, the degree of relationship between the analytic code variables is explored and discussed.

3.5.10 Description of Units of Analysis

The GWealth is the wealth management business of TBank serving affluent and high net worth clients with offices in 25 countries across key business areas including Private and International Banking, Wealth Advisory, Research, Economics and Strategy. The ICapital is the investment banking arm of TBank and a major player in the global financial markets offering investors a range of security products including Equities, Government and Corporate Debt Instruments, Commodities, Securitization, Credit, Credit Derivatives and Interest Rate products. The next two chapters, Chapter 4 and Chapter 5 are devoted to presentation of results from GWealth and ICapital case studies.

3.6 Summary

This chapter examined the research method for this study and provides an overview of research methodologies in social science. The chapter explored the underlying philosophy underpinning social science research and looked at the positivist and post-positivist research paradigms. The chapter acknowledged the scaffolding supporting positivism is modernism and outlined the alternative view to the natural science research, the naturalistic approach. The chapter further outlined and discussed the different types of research methodologies within the post-positivist paradigm and provided the justification for using the case study method for this research. Furthermore, the chapter examined the research design for this study and explored the means of data collection and data analysis for this research. Finally, the chapter outlined and discussed the components of the research framework for this study.

CHAPTER 4 - Case Study: GWealth

4.1 Introduction

This chapter presents the outcome of the case study at GWealth, the wealth management division of global financial institution TBank. The chapter focuses on presentation of interview results and alignment of the findings with the research propositions. The data collection method for this research is through interview of research participants. In presenting the result of this case study, the interview data was transcribed and transformed through grounded theory open coding methods involving line-by-line coding, translation of the descriptive codes into analytic codes, repeated comparison of the descriptive codes and the analytic codes for saturation, and identification of core themes from the analytic categories codes.

This chapter is divided into three Sections; Section 4.2 presents an overview of the GWealth management including its key business areas. The Section looks at the background of the GWealth reporting project, the project definition and the objectives including the architecture of the GWealth reporting data warehouse. Section 4.3 presents the findings of the case study and alignment of interview results with the research propositions. Lastly, Section 4.4 concludes the chapter with a summary.

4.2 The GWealth Management

GWealth is the wealth management arm of a major European financial institution TBank. It is one of the leading global wealth managers, and the UK's largest, with total client assets of £134.1bn, as at 30 June 2011. GWealth has offices in 25 countries and serves affluent, high net worth and intermediary clients world-wide. The company is divided into a number of key businesses including national and International Private Banking, Brokerage and Online Services. Other businesses include Wealth Intermediary and Advisory, Economics, Research and Strategy divisions. The national and international banking arm of the business offers bespoke banking and investment solutions through its network of global regional offices. GWealth offers diverse services to its global clients' including Financial Planning and Advisory, Discretionary Investment Advice, Bespoke Banking and Investment Solutions to high net worth individuals. GWealth also offers investment advice across the asset classes including Wealth Structuring. GWealth international banking business operates in onshore and offshore markets and covers around 8, 400 clients in over 144 countries.

GWealth intermediary arm of the business provides financial advice, products and services to wealth intermediaries and corporate organisations seeking offshore and cross border banking. The arm also offers a range of services to the financial needs of expatriates and non-UK nationals. Typically, GWealth intermediary clients are required to have between £10,000 and £100,000 of asset to invest.

International premier banking clients are required to have between £100,000 and £500,000 of asset to invest. The wealth advisory is the fiduciary and tax-led business of GWealth. It provides clients with a range of holistic wealth planning and structuring services and acts as the center of excellence for tax and fiduciary related issues. GWealth also has business operations in the Americas offering investment to ultra-high net worth clients; services offered to these clients are asset management, capital markets execution and trust services. GWealth operates in 12 regional office locations in the Americas.

4.2.1 Background of the Data Warehouse Project

In wealth management, client reporting is a key service differentiator and an area in which GWealth lagged behind its competitors. Although the service is considered a utility within the sector, a number of problems exist with current processes, technology and procedures within the organisation. Pre-data warehouse, client reporting process was manually intensive requiring two full-time teams dedicated to producing the client reports. The manual nature of producing the reports meant that significant efforts and time were expended each month creating the client reports and that process can take up to twenty five business days before reports were fully created and issued. Although the general data quality of the reports were considered to be good, the process to ensure this is the case was equally manually intensive and required series of checks and validation steps to ensure that correct data is presented in client reports. Additionally, the reports were difficult to customise to meet the client needs and often branded inconsistently.

Furthermore, there were distribution issues associated with delivering the reports to the clients; there are no electronic delivery mechanisms and consolidated clients mailing system in place.

The reporting data warehouse project was commissioned to implement a solution to address the problems described above and to deliver a strategic client reporting capability for GWealth. An outside vendor was engaged in an eight week exercise to investigate the use of a centralised information store (a data warehouse) for use within the wealth management. In addition, analysis was carried out to establish the importance and the benefits of such strategic information store for client reporting and GWealth. This work has since been completed and recommendations accepted by GWealth. Accordingly, the data architecture team was subsequently charged with implementing and delivering a strategic client reporting data warehouse for the organisation.

4.2.2 Project Definition and Objectives

The strategic goal of the reporting data warehouse is to facilitate the delivery of best of breed of client reporting solutions across wealth management at GWealth. In particular, the implementation focuses on affluent and high net worth UK booked clients and international clients. The clients reporting project was initiated to streamline the production of high quality reports that is capable of being customised to individual client preferences and segments; including performance against model asset allocation. The reporting project also aims to provide a consistent global delivery of reports on multi-channel basis across the group as well as providing flexible fulfillment options, supporting increased report production throughputs and reduced client attrition.

4.2.3 The Data Warehouse Architecture

The reporting data warehouse provides the data platform for client reporting within GWealth. The data warehouse obtains data feeds from a variety of locations onshore and offshore and from array of systems from these locations. The source files from the operational systems are transferred using a file transfer utility, the utility is used to move source feeds into a file share server where they are loaded into a staging database. Figure 4-1 illustrates the end-to-end architecture of GWealth data warehouse system.

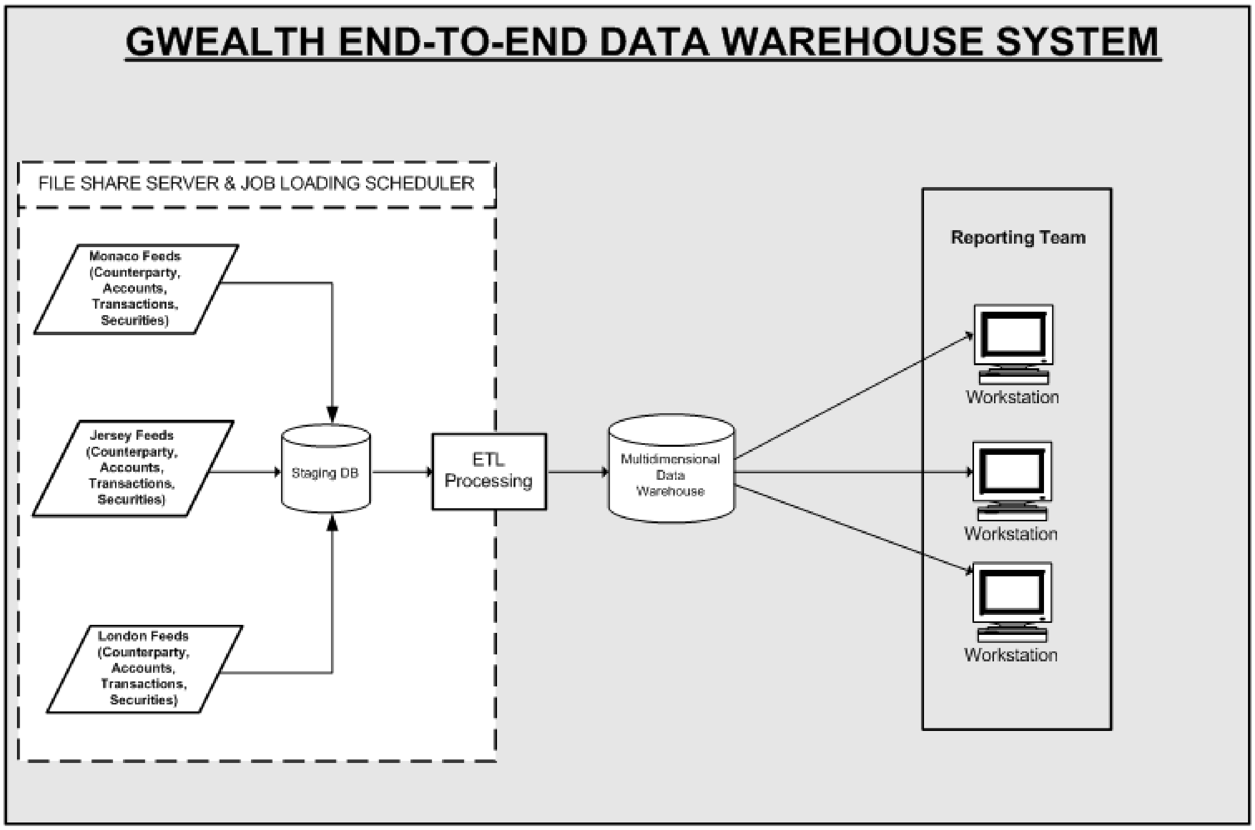

Figure 4-1: End-to-End Architecture of GWealth Data Warehouse System

Using an ETL tool, the data in the staging database is subsequently transferred into the multidimensional data warehouse illustrated in Figure 4-1 above. The GWealth reporting data warehouse is populated in alignment with the transformation rules in the technical design document (TDD). The ETL tool for the reporting data warehouse is high-end and sophisticated, the ETL tool is used to read data from the staging database and apply the transformation logic to populate the data warehouse model, which implements a single schema. At high-level, the tasks performed by the ETL tool include:

- Implementation of feed dependency logic required to load data into the dimensions and facts table
- Extraction of data required by each transformation process from the staging database
- Application of transformation logic to staging data in relation to the rules defined in mapping and transformation specification document.

Finally, a workload scheduler is used to manage the job-loading schedules and to orchestrate each of the ETL processes to load data into the reporting data warehouse. The scheduler is used for tasks such as initiating the batch load run and to manage the job dependencies. Figure 4-2 illustrates the GWealth star schema data model.

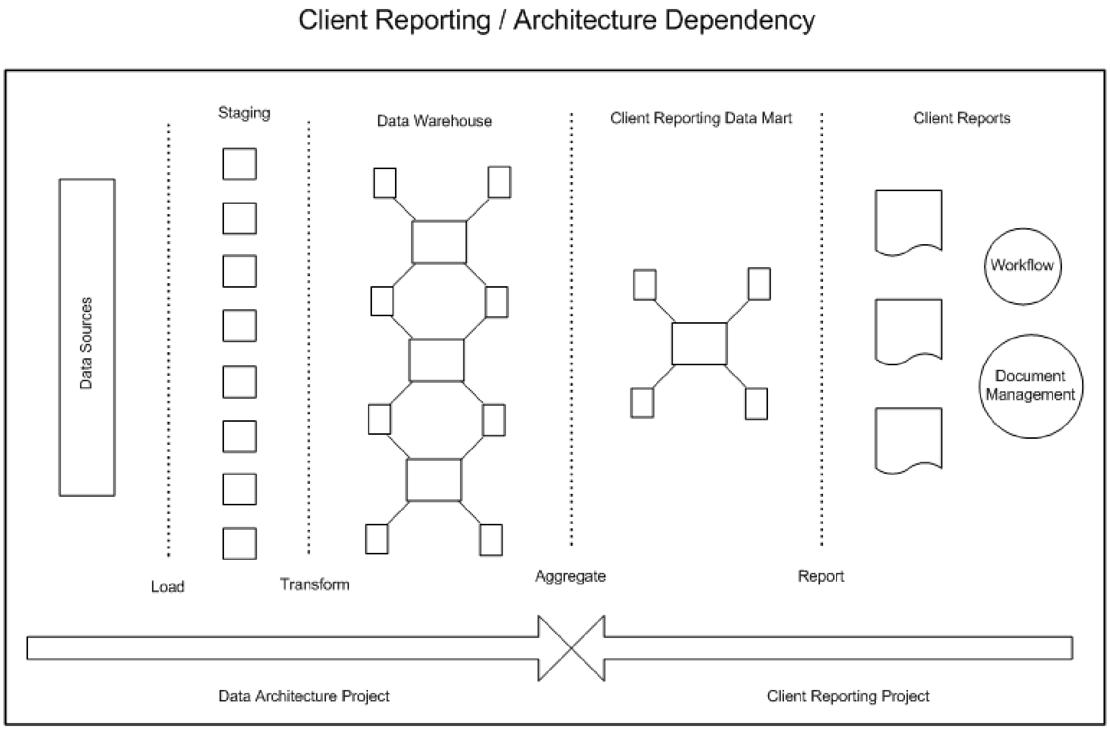

Figure 4-2: Client Reporting Conformed Dimension Data Architecture *(Adapted from GWealth architecture artifacts)*

The next Section presents the implementation of research design for this study.

4.3 Implementation of Research Design for GWealth

As indicated in Section 3.5, research design is the blueprint of case study research (McCoy et al, 1993). Research design is also a plan for assembling, organising and evaluating information according to the problem definition and specific goals for how to use a research finding. This Section describes the open coding activities carried out to identify the GWealth analytic category codes by transforming the empirical data collected from the research participants.

4.3.1 Transcription of GWealth Data and Applying Code to Text

As indicated in Section 3.5.6, a recording device was used to capture the interview sessions with the research participants at GWealth. Following each interview, empirical data was transcribed into a readable text using Microsoft Word. A sample of transcribed interview data is illustrated below in Figure 4-3.

TRANSCRIBED EMPIRICAL DATA

KO

What is the impact of ensuring high performance of query execution on the choice of the multidimensional data model for the reporting data warehouse?

SMMMBW

I think it is extremely important, I think you need to kind of consider what type of queries the user would want to run before you choose a particular model and, here our data warehouse wasn't that big to be honest. At [redacted], we have millions of transactions coming through every single day and again we found performance wise a dimensional model served the purpose of our needs so we built [redacted] on top of that in [redacted] and again the same we had it here; it seems to work a lot better. That said, the performance is also impacted by the database that you choose to use so, I think the moment that you choose and the hardware that you choosing to implement that model on, very very tightly coupled so we did have problems; I'm not sure if you are aware of the whole [redacted] side of things and so you could choose the best model in the world and you could design it to perfection, but if your hardware is not there to support that model, you're still going to have those problems. If your hardware is not up to the scratch, you will struggle and extending that; how you choose to load it would be another thing. Here we use [redacted] and it's a very high performance tool, ETL tool, I did have previous experience in [redacted] previously and this would not want to say anything about [redacted] as such, but at [redacted] we moved from [redacted] to [redacted] so to give you a real life kind of example, it was taking 22hrs with [redacted] to load all the data that we need into [redacted] warehouse and we have 24 hours in a day, so any problem that came up, you got just 2hrs to deal with it and obviously that was extremely difficult to maintain. We moved it to [redacted] and we got it down to 4 to 6 hrs, So it was quite a significant saving – we found the version we were on in [redacted], the version of [redacted] that they gave us at the time comparable with what [redacted] could do for us was massively different. So all those things, I just want to retreat the point, it is very important to choose the right model, but then if you don't have the tool to stand behind that model, it's kind of worthless; you lost all of the benefit that you might got with it.

Figure 4-3: Transcribed Empirical Data from GWealth Research Participant

As indicated in the research design for this study, the next step carried out by the investigator as part of the open coding process to identify the thematic categories for the study was line-by-line coding of transcribed data. Line-by-line coding for this study involved applying descriptive code to each row of text of transcribed data. A descriptive code is a description of what is happening in a particular row of text that is interpreted by the investigator; therefore, a descriptive code captures the essence of a row of text (Gibbs, 2007). For this research, the investigator imported transcribed data from MS Word into MS Excel where line-by-line coding was carried out. A sample of line-by-line coding in MS Excel is illustrated in Figure 4-4 below.

EMPIRICAL DATA: LINE-BY-LINE CODING

Transcribed Data	Apply Code to Text: Descriptive
The main reason why I chose dimensional structure for a warehouse would be	Performance is the main reason why I chose dimensional structure
performance to actually to bring back data. I always find that works a lot better wh	I always find that works a lot better for high performance
got more complex queries to get across to more tables and I find it that sometime:	when you get complex query to get across
you are inputting things in a very, very fast structure when that table gets too big i	Works better if you are imputing things into fast structure
coming through every single day and again we found performance wisea dimensic	Millions of transactions comes in daily
model served the purpose of our needs so we built ... on top of that in ...	Dimensional model served our needs performance wise
and again the same we had it here; it seems to work a lot better. That said, the	Dimensional model seem to work better for performance
performance is also impacted by the database that you choose to use so, I think th	Performance is impacted by the database that you choose
moment that you choose and the hardware that you choosing to implement that n	Performance is impacted by the data model that you choose
very, very tightly coupled so we did have problems; I'm not sure if you are aware c	Performance is impacted by the hardware you choose for your data m
whole ... side of things and so you could choose the best model in the world and y	You could choose the best model in the world
could design it to perfection, but if your hardware is not there to support that moc	Your hardware must be able to support your data model
you're still going to have those problems. If your hardware is not up to the scratch	You will have problems if hardware cannot support your data model
will struggle and extending that; how you choose to load it would be another thin;	It's another thing how you choose to load your data
we use ... and it's a very high performance tool, ETL tool, I did have previous	Used high performance ETL tool
experience in ... previously and this would not want to say anything about	Would not say anything about previous ETL tool
... as such, but at ... we moved from ... to ... so to give	Moved from one ETL to another
you a real life kind of example, it was taking 22hrs with ... to load all the data	Taking 22hrs to load data into a data warehouse
that we need into ... warehouse and we have 24 hours in a day, so any problem th:	Data loading taking almost all the time in a day
came up, you got just 2hrs to deal with it and obviously that was extremely difficu	only 2 hrs to deal with problems
maintain. We moved it to ... and we got it down to 4 to 6 hrs, So it was quite a	Got data loading dow to 6 hrs by moving into another tool
the performance of the micros that we needed for client reporting,	Help with performance of micros for client reporting
It has a massive role, you have to define what data people are looking for, what al	Performance is one of the greatest consideration to take into account
different sources are, what their end goal is, what are you trying to achieve with tl	Performance is very important when choosing your data architecture
business you have to get that very, very clear, is it to provide them with a warehoι	I've not worked in a place where users don't care about performance

Figure 4-4: Line-By-Line Coding of Transcribed Empirical Data

4.3.2 Convert Descriptive Codes to Analytic Codes and Compare Codes for Saturation

The next step carried out by the investigator as part of the open coding process to identify the thematic categories for this research was transforming the descriptive codes into analytical codes. Gibbs (2007) points out investigators should avoid descriptive codes; instead, investigators should formulate analytic codes, therefore, an analytic code is a conceptualization of the descriptive codes by the investigator. The following descriptive codes in Table 4-1 were transformed into the analytic codes in Table 4-2 as part of the process of identifying the category code *High Query Performance*.

	Line-By-Line Coding - Descriptive Code of Empirical Data
1	Performance is the main reason why I chose dimensional structure
2	I always find that works a lot better for high performance
3	when you get complex query to get across
4	Works better if you are imputing things into fast structure
5	Millions of transactions comes in daily
6	Dimensional model served our needs performance wise
7	Dimensional model seem to work better for performance
8	Performance is impacted by the database that you choose
9	Performance is impacted by the data model that you choose
10	Performance is impacted by the hardware you choose for your data model

Table 4-1: Descriptive Codes for Line-By-Line Coding

	Analytic Codes for Descriptive Codes in Figure 4-1
1	Chose multidimensional model for high performance
2	Multidimensional model works better for high performance
3	Multidimensional model works for complex queries
4	Multidimensional model works better for fast structures
5	Client Reporting processes large amount of transactions daily
6	Multidimensional model addressed business performance requirements
7	Multidimensional model works better for high performance
8	Performance is impacted by software and hardware
9	Performance is impacted by choice of data model
10	Performance is impacted by software and hardware

Table 4-2: Analytic Codes for Descriptive Codes

As indicated in the research design for this study, the next step engaged by the investigator was ensuring texts with similar themes were assigned the same code; coded passages were further reviewed and the codes compared for saturation. Figure 4.5 illustrates the process of transforming the descriptive codes in Table 4-1 to the analytic codes in Table 4-2; the process ensured texts with similar themes were assigned the same code using MS Excel.

TRANSFORMING DESCRIPTIVE CODES TO ANALYTIC CODES

Apply Code to Text: Descriptive	Descriptive Code -> Analytic Code	Compare Codes for Saturation
Works better if you are imputing things into fast structure	Multidimensional model works better for fast structures	Multidimensional model works better for high performance
Millions of transactions comes in daily	Client Reporting processes large amount of transactions daily	Multidimensional model works better for high performance
Dimensional model served our needs performance wise	Multidimensional model addressed business performance requirem	Multidimensional model works better for high performance
Dimensional model seem to work better for performance	Multidimensional model works better for high performance	Multidimensional model works better for high performance
Performance is impacted by the database that you choose	Performance is impacted by software and hardware	Performance is impacted by software and hardware
Performance is impacted by the data model that you choose	Performance is impacted by choice of data model	Performance is impacted by choice of data model
Performance is impacted by the hardware you choose for your data model	Performance is impacted by software and hardware	Performance is impacted by software and hardware
You could choose the best model in the world	Performance is impacted by software and hardware	Performance is impacted by software and hardware
Your hardware must be able to support your data model	Performance is impacted by software and hardware	Performance is impacted by software and hardware
You will have problems if hardware cannot support your data model	Performance is impacted by software and hardware	Performance is impacted by software and hardware
It's another thing how you choose to load your data	Performance is impacted by software and hardware	Performance is impacted by software and hardware
Used high performance ETL tool	Performance is impacted by software and hardware	Performance is impacted by software and hardware
Would not say anything about previous ETL tool	There is nothing to say about previous ETL software	Performance is impacted by software and hardware
Moved from one ETL to another	Replaced previous ETL tool with another tool	Performance is impacted by software and hardware
Taking 22hrs to load data into a data warehouse	It took nearly a day to load data into client reporting	Performance is impacted by software and hardware
Data loading taking almost all the time in a day	It took nearly a day to load data into client reporting	Performance is impacted by software and hardware
only 2 hrs to deal with problems	There is not enough time to address data loading issues	Performance is impacted by software and hardware
Got data loading dow to 6 hrs by moving into another tool	Replaced previous ETL tool with another	Performance is impacted by software and hardware
Help with performance of micros for client reporting	Gained improved performance for client reporting	Gained improved performance for client reporting
Performance is one of the greatest consideration to take into account	Performance is an important consideration	Multidimensional model works better for performance
Performance is very important when choosing your data architecture	Performance is an important consideration	Multidimensional model works better for performance
I've not worked in a place where users don't care about performance	Performance is an important consideration	Multidimensional model works better for performance
The answer is as fast as possible	Performance is an important consideration	Multidimensional model works better for performance

Figure 4-5: Transformation of Descriptive Codes to Analytic Codes

4.3.3 Identify Sub Category Codes and Main Category Codes

Category codes are logical groupings of the thematic analytic codes; a category code facilitates the retrieval of the underlying codes (Gibbs, 2007). As indicated in Section 3.5.5, the next step carried out by the investigator was to identify a category which a set of analytic codes aligned. The following categories were identified as the initial set of category codes for this case study (*High Query Performance*, *Specific Requirement*, *Generic Requirements*, *Defined Goal & Scope*, *Staff Experience*, *Data Consolidation* and *Diverse Consumers*). Analytic codes which did not aligned with the category codes outlined above were regarded as standalone category codes. The initial category codes were further distilled to identify top level category codes or main category codes which the initial set of category codes aligned; this ensured that category codes rolls up to a top level category code. As a result, the initial category codes (*High Query Performance*, *Specific Requirement*, *Generic Requirements*, *Defined Goal & Scope*, *Staff Experience*, *Data*

Consolidation and *Diverse Consumers*) became sub-category codes. As part of the open coding process for this research, the following main analytic category codes (*High Performance of Query Execution*, *Business Requirement*, *Firm Objective*, *Employee Experience*, *Enterprise Data Hub*) were identified for this case study. In addition, the following analytic codes (*Model Inflexibility*, *Limited Complexity & Diversity*, *Conceptualization Consideration*, *Diversification*) were standalone category codes, as indicated above; these analytic codes could not be aligned with any of the sub category codes or main analytic category codes. Figure 4.6 and Figure 4.7 respectively illustrates a sample of sub category code (***High Query Performance***) and main analytic category code (***High Performance of Query Execution***) for GWealth case study.

IDENTIFICATION OF SUB-CATEGORY CODES

Descriptive Code -> Analytic Code	Compare Codes for Saturation	Sub Category Code
Chose multidimensional model for high performance	Multidimensional model works better for high performance	High Query Performance
Multidimensional model works better for high performance	Multidimensional model works better for high performance	High Query Performance
Multidimensional model works for complex queries	Multidimensional model works better for high performance	High Query Performance
Multidimensional model works better for fast structures	Multidimensional model works better for high performance	High Query Performance
Client Reporting processes large amount of transactions daily	Multidimensional model works better for high performance	High Query Performance
Multidimensional model addressed business performance req	Multidimensional model works better for high performance	High Query Performance
Multidimensional model works better for high performance	Multidimensional model works better for high performance	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by choice of data model	Performance is impacted by choice of data model	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
There is nothing to say about previous ETL software	Performance is impacted by software and hardware	High Query Performance
Replaced previous ETL tool with another tool	Performance is impacted by software and hardware	High Query Performance
It took nearly a day to load data into client reporting	Performance is impacted by software and hardware	High Query Performance
It took nearly a day to load data into client reporting	Performance is impacted by software and hardware	High Query Performance
There is not enough time to address data loading issues	Performance is impacted by software and hardware	High Query Performance
Replaced previous ETL tool with another	Performance is impacted by software and hardware	High Query Performance
Gained improved performance for client reporting	Gained improved performance for client reporting	High Query Performance
Performance is an important consideration	Multidimensional model works better for performance	High Query Performance

Figure 4-6: Sub-Category Code (*High Query Performance)* for the Analytic Codes

MAIN CATEGORY CODES FOR SUB-CATEGORY CODES

Sub Category Code	Main Category Code
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution

Figure 4-7: Main Category Code for High Query Performance Sub-Category Code

4.3.4 Identify Properties of GWealth Category Codes

As indicated in the research design framework presented in Section 3.5.5, categories have properties, properties enable analytic category codes to be viewed from multiple perspectives (Gibbs, 2007). As part of the open coding process for this study, sub category codes developed in Section 4.3.3 rolled up to the main category codes, thus, the sub category codes became the properties of top level category codes for this case study. This enables the main category codes to be viewed from multiple perspectives. For instance, the category code *High Performance of Query Execution* has a property *High Query Performance* as illustrated in Figure 4.8.

MAIN CATEGORY CODE: PROPERTY

Main Category Code	Property of Main Category Code
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance

Figure 4-8: Property of Main Category Code

4.3.5 Analytic Codes and Frequency Count of Analytic Codes

As part of the open coding process for the study, a count of frequency occurrence of the analytic category codes was developed, the count of frequency of an analytic code is the number of times an analytic category code is noted in the empirical data. This was done by counting the number of times each instance of an analytic category code appeared in the empirical data, the counts were then aggregated in MS Excel to provide total frequency occurrence for an analytic category code; this process is repeated for all the analytic codes. As indicated in Section 3.5.8, analysis of frequency count of analytic codes is an important step in linking the empirical data with the research propositions. Figure 4.9 illustrates the build of frequency occurrence of an analytic code (*High Query Performance*) for this study

FREQUENCY OCCURRENCE OF ANALYTIC CODE

Analytic Code	Frequency Occurrence
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
....	....
High Query Performance	1
High Query Performance	1
High Query Performance	1
Total	72

Figure 4-9: Frequency Occurrence of Analytic Code

4.4 Presentation of the Results – GWealth Case Study

GWealth interviews were conducted with ten participants, two of whom were senior data architects and directors, three senior developers, one program manager, one project manager, two development managers and one business analyst. The directors and the program manager were part of the funding group at GWealth, the data architects, developers; project manager and the IT development managers were part of the execution team responsible for implementing and developing the reporting data warehouse for the organisation. Table 4-1 below presents the frequency occurrence of the analytic codes for GWealth.

Analytic Codes	**No of Frequency**	**Cumulative Number**	**Cumulative Percentage**
High Query Performance	72	72	21.0
Specific Requirement	84	156	45.5
Defined Goal & Scope	65	221	64.4
Staff Experience	81	302	88.0
Enterprise Data Consolidation	22	324	94.5
Model Inflexibility	5	329	95.9
Limited Complexity & Diversity	5	334	97.4
Generic Requirements	4	338	98.5
Conceptualization Consideration	2	340	99.1
Enterprise Consumer Base	2	342	99.7
Diversification	1	343	100.0
Total	343		

Table 4-3: Frequency Analysis of GWealth Analytic Codes *(See Appendix 3 for GWealth Open Codes)*

The frequency occurrence of each of the analytic codes for GWealth is presented in Table 4-1 above. The column "*No of Frequency*" contains the number of times an analytic code is noted as a decision factor in adopting a multidimensional data model for GWealth reporting data warehouse. The column "*Cumulative Number*" is the aggregation of each of the frequency number or frequency count of each of the analytic codes for GWealth. The column "*Cumulative Percentage*" provides the percentage value for the "*Cumulative Number*". Figure 4-10 below presents the data in Table 4-1 in a graphical format using the *Pareto* chart.

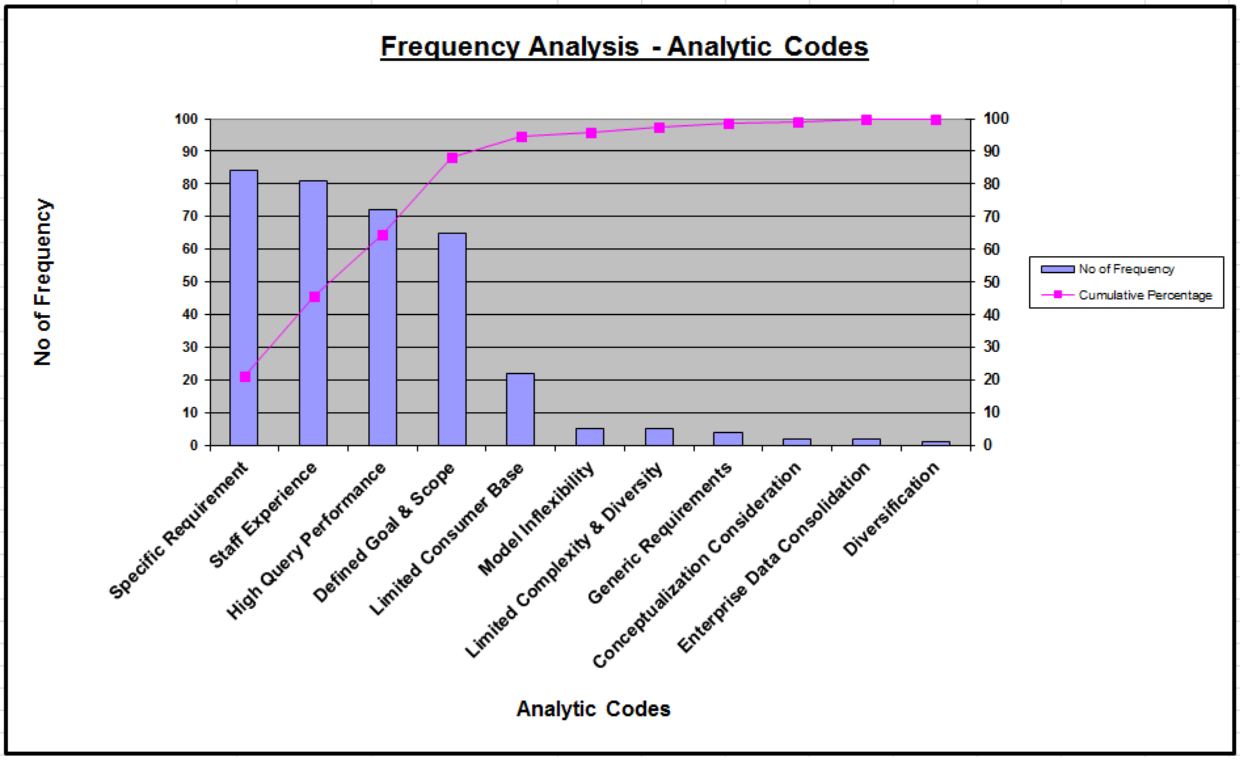

Figure 4-10: Frequency Analysis of the Analytic Codes Using the Pareto Chart

The above Pareto chart presents the frequency analysis of the analytic codes for GWealth. As illustrated in the graph, the analytic codes related to: *High Query Performance*, *Specific Requirement*, *Defined Goal and Scope*, *Past Experience* and *Limited Consumer Base* accounted for 86 per cent of the total factors that influenced the decision to adopt a multidimensional data model for GWealth reporting data warehouse. The rest of this chapter is devoted to presenting the results for the analytic codes.

4.4.1 Assessing the impact of high performance of query execution

In order to understand the impact of high performance of query execution on the choice of logical data model for a data warehouse, the questions below were addressed to GWealth research participants.

- What non-functional requirements impacted the choice of reporting warehouse logical data model?
- What other non-functional factors influenced the decision to adopt the multidimensional data model for the reporting data warehouse?
- How would you characterise the advantages of reporting warehouse data model?

Table 4-2 below outlines the percentage code frequency of the analytic codes related to: *High Query Performance*, *Inflexibility of Model*, *Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration*, *Enterprise Consumer Base*, and *Diversification* analytic codes extracted from Table 4-1.

Analytic Codes	**% Code Frequency**
High Query Performance	79.12087912
Inflexibility of Model	5.494505495
Limited Complexity & Diversity	5.494505495
Generic Requirements	4.395604396
Conceptualization Consideration	2.197802198
Enterprise Consumer Base	2.197802198
Diversification	1.098901099
Total of % Code Frequency	100

Table 4-4: Percentage Code Frequency of Query Performance vs. Other Analytic Codes

Figure 4-11 below illustrates the frequency analysis of the analytic codes (*Table 4-2*) in relation to questions to assess the impact of high performance of query execution on the choice of logical data model for a data warehouse.

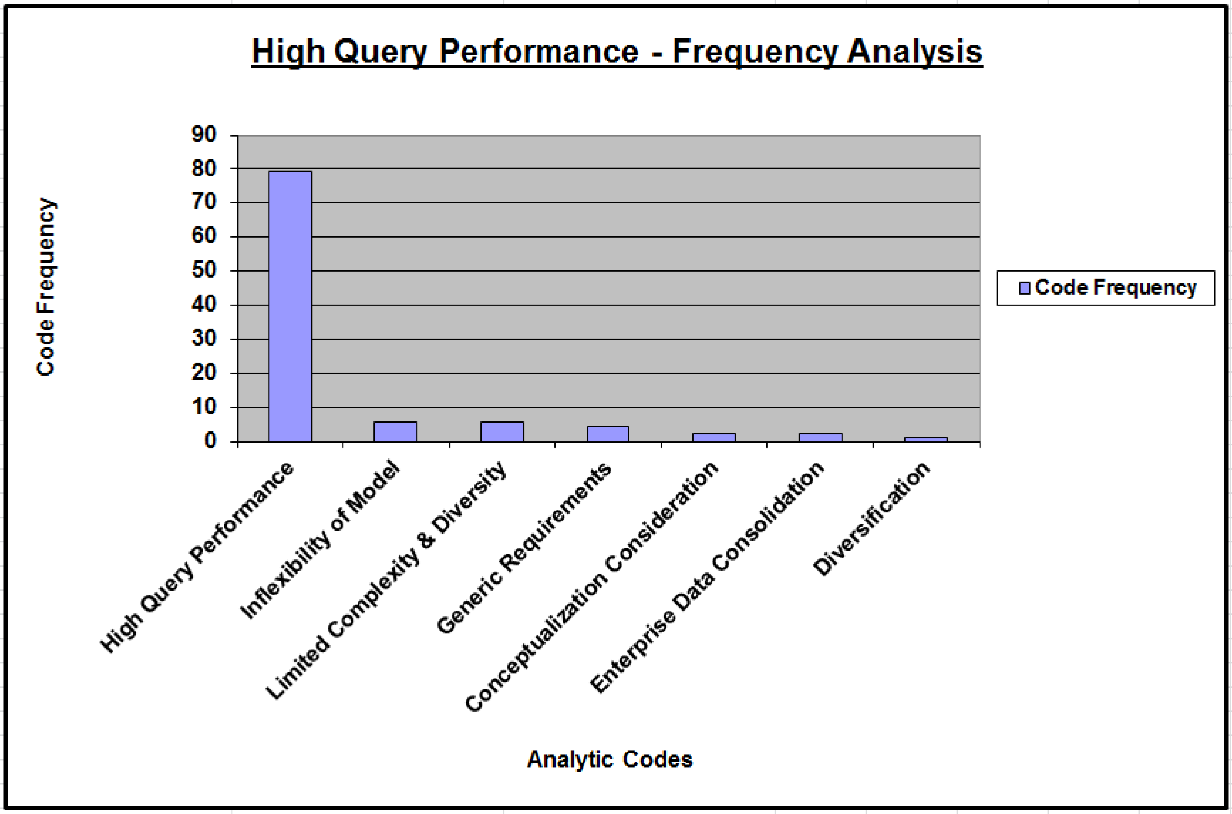

Figure 4-11: Impact of Performance on Logical Data Model Decision

In Figure 4-11, the responses from the research participants were coded along the x-axis, the analytic codes; the frequency analysis of the analytic codes is on the y-axis. The research participants were asked about the impact of performance as a measure to assess the impact of performance of query execution on the choice of data model for the reporting data warehouse. The research participants stated 79 per cent of the time that performance is a key decision factor in using the multidimensional data model for GWealth reporting data warehouse, this percentage number is considered high as a decision factor.

The percentage code frequency of performance as a key decision factor for implementing the reporting data warehouse on a multidimensional data model is more than double the percentage code frequency of all the other analytic codes combined. In contrast, the percentage code frequency for other decision factors is quite low, collectively; they represented 21 percent of the total decision factors. The responses from the research participants indicated that performance is the most important non-functional requirement that influenced the choice of data model for the reporting data warehouse. One of the non-functional requirements of the reporting warehouse that featured prominently in the participant responses is the need to quickly process client reports from the reporting data warehouse. An important element of ensuring the client reports were quickly processed is the need to ensure the reporting warehouse data model is portable and optimized for high performance query execution. All the research participants stated the importance of performance to the business and the expectation of getting high performance out of the reporting data warehouse. The research participants believed the multidimensional data model played a key role in achieving the requirement of performance, having considered the infrastructure supporting the reporting data warehouse. In order to shed light on the other non-functional factors, which influenced the choice of reporting data warehouse logical data model, the research participants reaffirmed that the business requirements was a factor in adopting a multidimensional data model for the reporting data warehouse. The research participants outlined a number of advantages of the reporting data warehouse data model. One of the advantages outlined by the research participants was that the reporting warehouse data model is easier to understand. For instance:

Project Manager	*I would say for me, it is easier to understand and easier to isolate issues and of course, it allows you to focus on what*

	you initially set out to do which is to implement a reporting warehouse
Lead Data Architect	*The main one for me is around the query performance and the end-user being able to understand exactly what you build*

The respondents at GWealth were able to articulate the benefits of the reporting data warehouse data model; they stated it allowed them to focus on what they set out to accomplish which was to build a reporting data warehouse. The respondents also stated their multidimensional data model enabled them to achieve the required non-functional performance requirement of the reporting data warehouse. Furthermore, the respondents were emphatic in emphasising that it is possible to experience an appreciable improvement in performance of a data warehouse based on how the underlying data is structured. This is an indication the respondents were using their experience in situation awareness while making judgement to engage a course of action to address the performance issues of client reporting data warehouse. This is in alignment with the prescription of naturalistic decision making. The frequency occurrence of the analytic code for performance in the narrative provides a justification to support the proposition that high performance of query execution is a decision factor influencing the choice of multidimensional data model for GWealth reporting data warehouse.

4.4.2 Assessing the impact of specific business requirement

In order to assess the impact of specific business requirement on the choice of logical data model for a data warehouse, the questions below were addressed to the research participants

- What is the impact of the business requirements on the choice of reporting data warehouse logical data model?
- How many areas of the business have been on-boarded onto the reporting data warehouse?

The responses from the research participants provided the evidence to support the proposition that the degree of focus on specific business requirement influences the choice of logical data model for a data warehouse. All the research participants at GWealth agreed the business commissioned the reporting data warehouse to address a major requirement for their organisation, to enable the production of clients' statements from the reporting data warehouse. Table 4-3 below outlines the percentage code frequency of the analytic codes related to: *Specific Requirement*, *Inflexibility of Model*, *Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration*, *Enterprise Consumer Base*, and *Diversification* analytic codes extracted from Table 4-1.

Analytic Codes	% Code Frequency
Specific Requirement	81.55339806
Inflexibility of Model	4.854368932
Limited Complexity & Diversity	4.854368932
Generic Requirements	3.883495146
Conceptualization Consideration	1.941747573
Enterprise Consumer Base	1.941747573
Diversification	0.970873786
Total of % Code Frequency	100

Table 4-5: Percentage Code Frequency of Specific Requirement vs. Other Analytic Codes

Figure 4-12 below illustrates the frequency analysis of the analytic codes (*Table 4-3*) in relation to questions to assess the impact of specific business requirement on the choice of data warehouse logical data model.

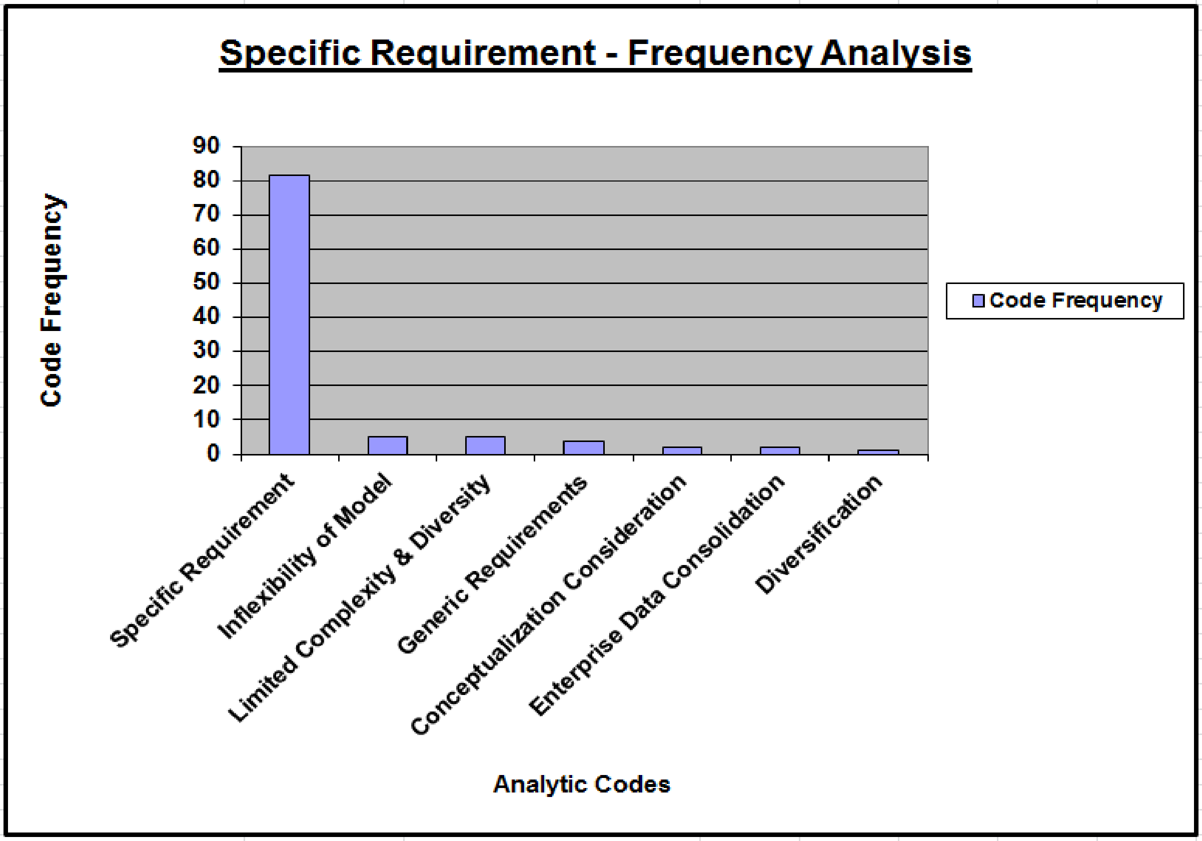

Figure 4-12: Impact of Specific Requirement on Logical Data Model Decision

In Figure 4-12, the responses from the research participants were coded along the x-axis, the analytic codes; the percentage code frequency for each of the analytic codes is on the y-axis. The research participants were asked about the impact of the need to address a specific business requirement on the choice of logical data model for the reporting data warehouse. The research participants at GWealth stated 81 per cent of the time that the reporting data warehouse data model was adopted to address the specific requirement of the business for client reporting, this percentage number is considered high as a decision factor. In contrast, the percentage code frequency for other decision factors is quite low, collectively; they represented 19 percent of the total decision factors for using multidimensional data model for the reporting data warehouse.

The percentage of code frequency of the analytic code to address specific business requirement is almost double the percentage code frequency of other analytic codes combined and 68 per cent more than the percentage code frequency of the analytic codes related to *inflexibility of model enhancement*, *complexity* and *diversity*. Additionally, the percentage code frequency of the analytic code to address specific business requirement is 78 times higher than the percentage code frequency of the analytic code to address the generic requirements of other business groups within GWealth. The research participants agreed that the business requirement had a big impact in influencing the choice of logical data model for reporting data warehouse. The research participants at GWealth believed the requirement from the business to create a reporting data warehouse rather an enterprise data warehouse influenced the decision to adopt a multidimensional data model for the reporting data warehouse. It was recognised by the respondents that data warehouses are requirements based, as such, the reporting warehouse was not created because the respondents thought it would be useful for the firm to create it. Rather, the requirement to create the reporting data warehouse was based on explicit requirements from the business and appropriately funded. The respondents believed their multidimensional data model was in alignment with the requirement from the business for client reporting. For instance:

Data Architect	*The requirement was to develop a data warehouse for creating reporting statements for clients of the firm. I found multidimensional model to be perfect for that type of requirement*
Lead Data Architect	*For reporting warehouse, we have to focus on the universe of data we need for the statement reports and how best to structure it in an efficient manner. The way we have done that is by using the dimensional model.*

Program Manager *In our own case, we are reporting on performance data, so the architects have structured the business requirements around the dimensional model; that has worked for us*

In addressing the question related to the number of other areas of the business "*on-boarded*" onto the reporting data warehouse, the research participants responded the reporting data warehouse has limited consumer user base due to the goal of the reporting warehouse. The reporting data warehouse was commissioned by the business solely to create report of statements for the clients of the firm. Other consumer groups within GWealth were not intended to use the reporting data warehouse, except for the purpose for which it was created and neither is the reporting data warehouse expected to cater for the needs of the other consumer groups within GWealth. The data from the respondents provided the evidence to support the naturalistic decision making, the research participants were able to recognise the requirements from the business was best implemented on a multidimensional data model. For instance, the respondents indicated a multidimensional data model is a perfect fit for the type of requirement requested by the business. This not only suggests a level of situation awareness on the part of the respondent in alignment with the prescription of naturalistic decision making, it also indicated mastery of the subject matter, exercising judgement on when and when not to engage a multidimensional data model. For instance:

Data Architect *For reporting warehouse, the requirement was to develop a data warehouse for creating reporting statements for clients of the firm. I found the multidimensional model to be a perfect fit for that type of requirement*

The data from GWealth appear to support the proposition that the degree of focus on specific business requirement influences the choice of logical data model for a data warehouse. However, the data provided no evidence of how the business requirement was mapped to the individual dimension in the multidimensional data model. The result presented above provides justification to support the proposition that the need to focus on the requirement to address a specific business problem is a decision factor influencing the choice of multidimensional data model for a data warehouse.

4.4.3 Assessing the impact of goal and scope of a data warehouse

In order to understand the impact of the goal and scope of a data warehouse on the choice of a logical data model, the research participants were asked the questions below:

- What is the role of the goal of reporting warehouse on the choice of its data model?
- What is the impact of the scope of reporting warehouse on the choice of its data model?
- What is end-users' ability to understand data relationship in your data warehouse in relation to the goal of the data warehouse?

Table 4-4 below outlines the percentage code frequency of the analytic codes related to: *Defined Goal & Scope*, *Inflexibility of Model*, *Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration*, *Enterprise Consumer Base*, and *Diversification* analytic codes extracted from Table 4-1.

Analytic Codes	**% Code Frequency**
Defined Goal & Scope	77.38095238
Inflexibility of Model	5.952380952
Limited Complexity & Diversity	5.952380952
Generic Requirements	4.761904762
Conceptualization Consideration	2.380952381
Enterprise Consumer Base	2.380952381
Diversification	1.19047619
Total of % Code Frequency	100

Table 4-6: Percentage Code Frequency of Defined Goal & Scope vs. Other Analytic Codes

Figure 4-13 below illustrates the frequency analysis of the analytic codes (*Table 4-4*) in relation to questions to assess the impact of defined goal and scope on the choice of logical data model for a data warehouse.

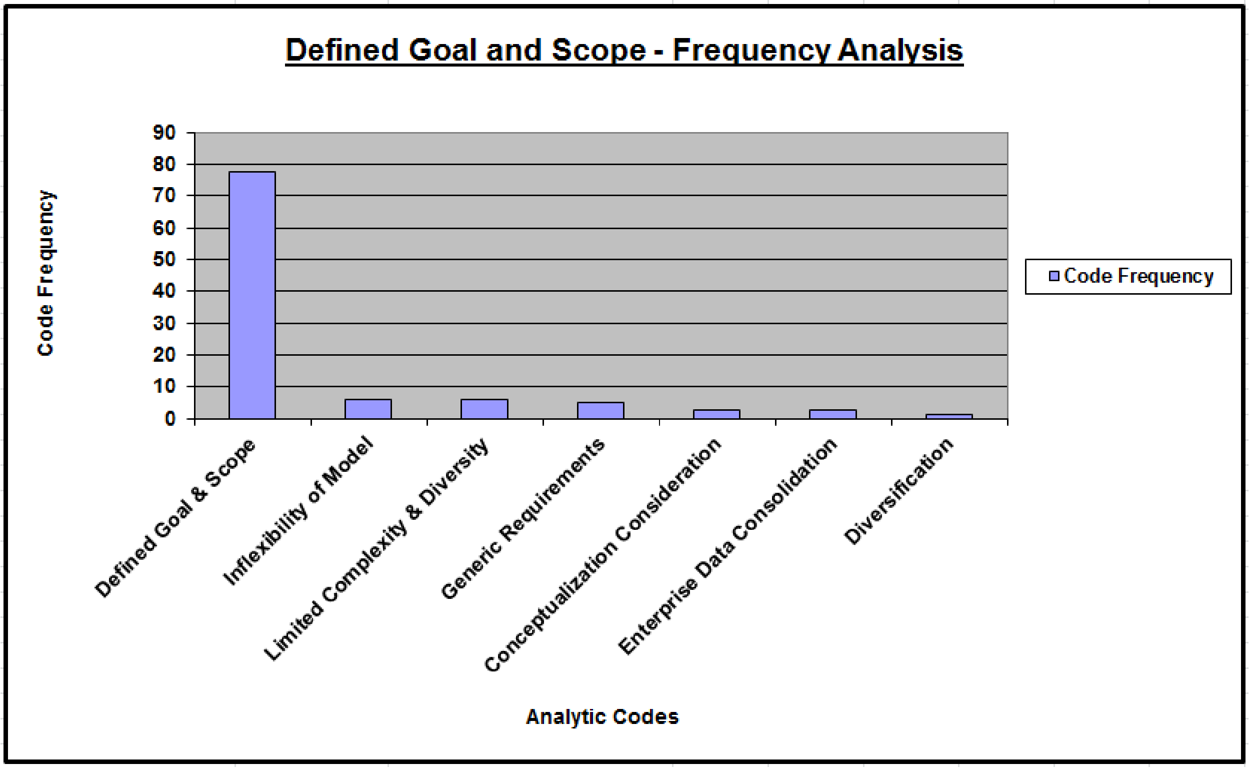

Figure 4-13: Impact of Defined Goal and Scope on Logical Data Model Decision

In Figure 4-13, the responses from the research participants were coded along the x-axis, the analytic codes; the percentage code frequency of each of the analytic codes is on the y-axis. The research participants were asked about the goal of the reporting warehouse as a measure to assess the impact of defined goal and scope of a data warehouse in a decision to adopt a logical data model for the reporting data warehouse. The research participants stated 77 per cent of the time that the goal and scope of the reporting data warehouse was a factor in adopting a multidimensional data model for the reporting data warehouse, this figure is considered high when contrasted with the percentage total for the other decision factors combined. In comparison, the percentage code frequency for the rest of the decision factors is considered low. Collectively, they accounted for 23% of the total decision factors for using multidimensional data model for GWealth reporting data warehouse. Additionally, the percentage code frequency of defined goal and scope is more than 60 per cent higher than the percentage of code frequency related to *inflexibility of model enhancement*, *complexity* and *diversity* decision factors. Further examination of the data indicated the percentage code frequency of the analytic code related to business defined goal and scope is greater than the combined percentage code frequencies of all the analytic codes in Figure 4-13. The data collected at GWealth indicated the respondents considered the business stated goal and scope of the reporting warehouse in selecting its data model. The data also indicated the respondents were aware of selecting an appropriate data model for the reporting data warehouse. This suggested the respondents were cognizant of the risks and potential consequences associated with selecting a data model that is not in alignment with the business-defined goal of the reporting data warehouse. For example:

Data Architect	*You want to use the data model that is best fit for the goal of your data warehouse, otherwise, you may find your initial goals are not fully realised by using an inappropriate data model; that would be unacceptable to the business*
Lead Data Architect	*It is important to consider the goal of your data warehouse before you select your data model, this enables you to select the appropriate architecture having considered all other factors that are critical to achieving the goal of the data warehouse*
Project Manager	*It is important to take the goal of your data warehouse into consideration when you are choosing the data model for your data warehouse. You don't want to find out later in the project that you have selected a wrong data model for your warehouse; that would be an awful and expensive mistake*

Developer	*I think it is a risk to select a data model for your warehouse without due consideration for the goal of your data warehouse*

Additionally, the data from GWealth indicated the reporting data warehouse was intended for no other purposes than for creating reports for the clients of the firm. GWealth reporting data warehouse was not created nor intended as an enterprise data distribution repository serving the needs of the heterogeneous business groups within GWealth, to this extent, adopting a multidimensional data model is judged to be in alignment with the scope of the reporting data warehouse. For instance:

Program Manager	*The sole objective of the reporting warehouse is to use it for reporting purposes, our data model has allowed us to meet that objective*

The data from GWealth indicated the respondents considered the scope of the reporting data warehouse as decision factor in adopting the multidimensional data model for client reporting data warehouse. In particular, the data showed the respondents considered that data warehouses that are small in scope are particularly suitable for a multidimensional data model considering the remarks below:

Data Architect	*If the scope is very small as in the case of the reporting warehouse, a multidimensional data architecture lend itself to a much better data warehouse*
Lead Data Architect	*The scope of the data warehouse is important as the goal of the data warehouse. In most cases, the goal of the data warehouse will determine the scope and size of the data warehouse, both are*

	important factors in determining the choice of your data warehouse data model
Program Manager	*The scope of the reporting warehouse is rather small when compared to other initiatives within the bank; that has contributed to reasons why the architects have used the data model that we have for the reporting warehouse*

The data from GWealth further indicated the respondents were risk averse and were not particularly willing to experiment or try out alternative data modelling approaches that may also be suitable for the client reporting data warehouse. For instance:

Program Manager	*you go with tried and tested methods especially when you are a new team and you want to proof yourself. You don't want to go down the path of, you know; let's try the new design because if you don't deliver you are in trouble*

The respondents at GWealth acknowledged the business defined goal and scope of the reporting data warehouse was a decision factor in adopting the multidimensional data model for the reporting warehouse. The data indicated the respondents implicitly engaged in situation awareness and consequently model categorisation based on recognising the goals and scope that multidimensional data model is suited. The frequency analysis of the analytic code to address the defined goal and scope in the narrative provides the justification to support the proposition that business defined goal and scope is a decision factor influencing the choice of multidimensional data model for a data warehouse.

4.4.4 Assessing the impact of implementation orientation of available

In order to assess the impact of the implementation orientation of available resources on the choice of logical data model for a data warehouse, the questions below were addressed to GWealth research participants.

- What is the impact of staff experience in the decision to adopt a multidimensional data model for the reporting data warehouse?
- Would you agree or disagree that your staff experience influenced the choice of the data model for your data warehouse?

The implementation orientation of the available resources is defined by the experience of the employees tasked with building a data warehouse. The employee experience is a major factor in choosing and adopting a multidimensional data model for the reporting data warehouse. Table 4-5 below outlines the percentage code frequency of the analytic codes related to: *Staff Experience*, *Inflexibility of Model*, *Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration*, *Enterprise Consumer Base*, and *Diversification* analytic codes extracted from Table 4-1.

Analytic Codes	% Code Frequency
Staff Experience	81
Inflexibility of Model	5
Limited Complexity & Diversity	5
Generic Requirements	4
Conceptualization Consideration	2
Enterprise Consumer Base	2
Diversification	1
Total of % Code Frequency	100

Table 4-7: Percentage Code Frequency of Staff Experience vs. Other Analytic Codes

Figure 4-14 below illustrates the frequency analysis of the analytic codes (*Table 4-5*) in relation to questions to assess the impact of implementation orientation of the available resources on the choice of logical data model for a data warehouse.

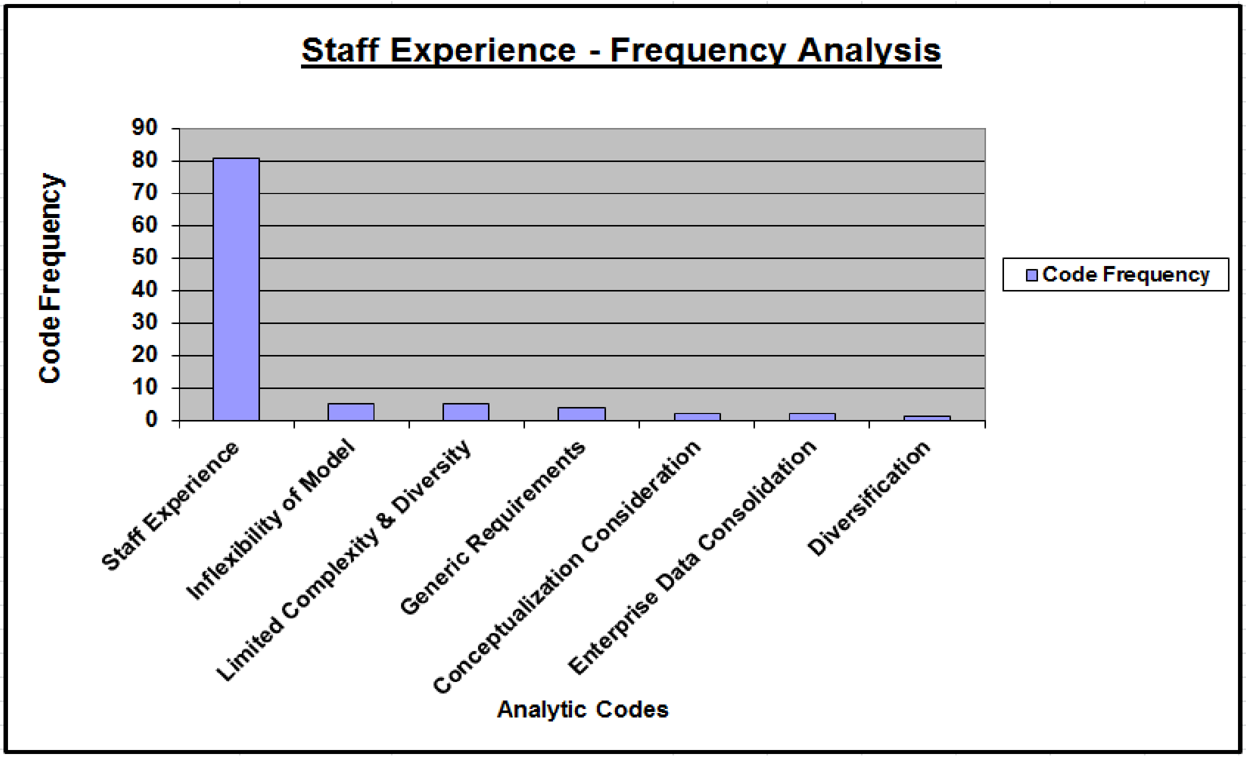

Figure 4-14: Impact of Staff Experience on Logical Data Model Decision

In Figure 4-14, the responses from the research participants were coded along the x-axis, the analytic codes; the percentage code frequency of each of the analytic codes is on the y-axis. The research participants were asked about the impact of staff experience as a measure to assess the impact of the implementation orientation of available resources in adopting a logical data model for a data warehouse. The research participants at GWealth stated 81 per cent of the time that their staff experience is a decision factor in adopting a multidimensional data model for the reporting data warehouse, this number is considered high as a decision factor. In contrast, the percentage code frequency for other decision factors is quite low, collectively; they represented 19 percent of the total decision factors. Additionally, the percentage code frequency of staff experience as a decision factor is greater than the percentage code frequency of all the analytic codes combined. The percentage code frequency of the analytic code for staff experience is more than 5 times the percentage code frequency of the analytic code for *model inflexibility*. The same is true for the percentage code frequency for *complexity* and *diversity* decision factors. The research participants at GWealth are highly experienced and skilled in multidimensional data warehousing; the research participants combined experiences in multidimensional data architecture spanned several decades, they expressed their preference for the multidimensional data architecture and are keen advocates of the data model. The participants at GWealth were risk averse, one of the striking remark made by GWealth respondents was that they would rather stay with tried and tested data model they know and familiar than experimenting with the alternative data models, which they claimed may not yield the desired result. For instance:

Developer	*We have senior people in the team that has implemented this type of project before, they pretty much know the ins and out of it so it was not difficult to convince a lot of people that dimensional model was what was needed for the reporting warehouse. You stay with what you know, that is very important especially in such a high stake highly visible project like reporting warehouse*
Program Manager	*My preference is always dimensional warehouses but that's just because that is what I know. It was what people in the team knew and have experienced and background*
Data Architect	*My background is all in dimensional modelling and very rarely if ever, do I use any third normal form modelling in data warehousing*

Project Manager — *It's my experience of only ever worked with multidimensional data warehouse. The people I've worked with have always been kind of pushed for that kind of model*

The research participants at GWealth overwhelmingly agreed that their staff experience played a key role in adopting the multidimensional model for their reporting data warehouse. The research participants at GWealth stated there were colleagues in the team with considerable experience in multidimensional data warehousing. The respondents at GWealth were mostly familiar with a multidimensional data model. For instance:

Data Architect — *Yes, I will agree the experience of the team influenced the choice of reporting warehouse. We used dimensional model because that's what we know, all my experiences in data warehousing has been around dimensional modelling so that's what I know and do best. So yes, staff experience played a huge part*

Program Manager — *I think I will agree with that statement, I think staff experience does played a huge role because dimensional model is what we all know, certainly from my perspective and I think from the perspective of others as well*

Lead Data Architect — *I agree the experience of the staff impacted the choice of our data warehouse. I'm sure you will see that is the case if you speak to other members of the team*

Project Manager — *I think it will be difficult to argue the experience of the team did not impact the choice of reporting warehouse data model, the*

	teams have strong views about dimensional model, they've been using it for a long time. I somehow belong to that category as well, as I said earlier; I've only ever worked on dimensional data warehouse. You will find that is the same for everyone
Developer	*I will definitely agree our staff experience swayed the decision to use dimensional model for the reporting warehouse. There are many people here that that's what they've used for many years and are very effective at it*

The prior experience of the development team that built the reporting data warehouse on a multidimensional data model was a major factor in using a multidimensional data model for the reporting data warehouse. The development team at GWealth had an overwhelming preference for a multidimensional data model because they have considerable experience in using multidimensional data model in their careers. The data from GWealth indicated the research participants leveraged their experience of situation awareness, drawing on it to address the business requirement to build the client reporting data warehouse. For this study, the data from GWealth provided the justification to support the prescriptions of naturalistic decision making. The empirical data provides the justification to support the proposition that the implementation orientation of available resources impacts the choice of logical data model for a data warehouse.

4.4.5 Assessing the impact of enterprise data consolidation

In order to assess the impact of enterprise data consolidation to address different reporting and analytics requirements on the choice of a logical model for a data warehouse, the questions below were addressed to research participants at GWealth.

- How diverse is the user base of your data warehouse?
- How often are different areas of the business on-board the reporting data warehouse?

Table 4-6 below outlines the percentage code frequency of the analytic codes related to: *Enterprise Data Consolidation, Limited Consumer Base, Inflexibility of Model, Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration* and *Diversification* analytic codes extracted from Table 4-1.

Analytic Codes	**% Code Frequency**
Enterprise Consumer Base	4.87804878
Enterprise Data Consolidation	53.65853659
Inflexibility of Model	12.19512195
Limited Complexity & Diversity	12.19512195
Generic Requirements	9.756097561
Conceptualization Consideration	4.87804878
Diversification	2.43902439
Total of % Code Frequency	100

Table 4-8: Percentage Code Frequency of Enterprise Data Consolidation vs. Other Analytic Codes

Figure 4-15 below illustrates the frequency analysis of the analytic codes (*Table 4-6*) in relation to questions to assess the impact of enterprise data consolidation on the choice of logical data model for a data warehouse.

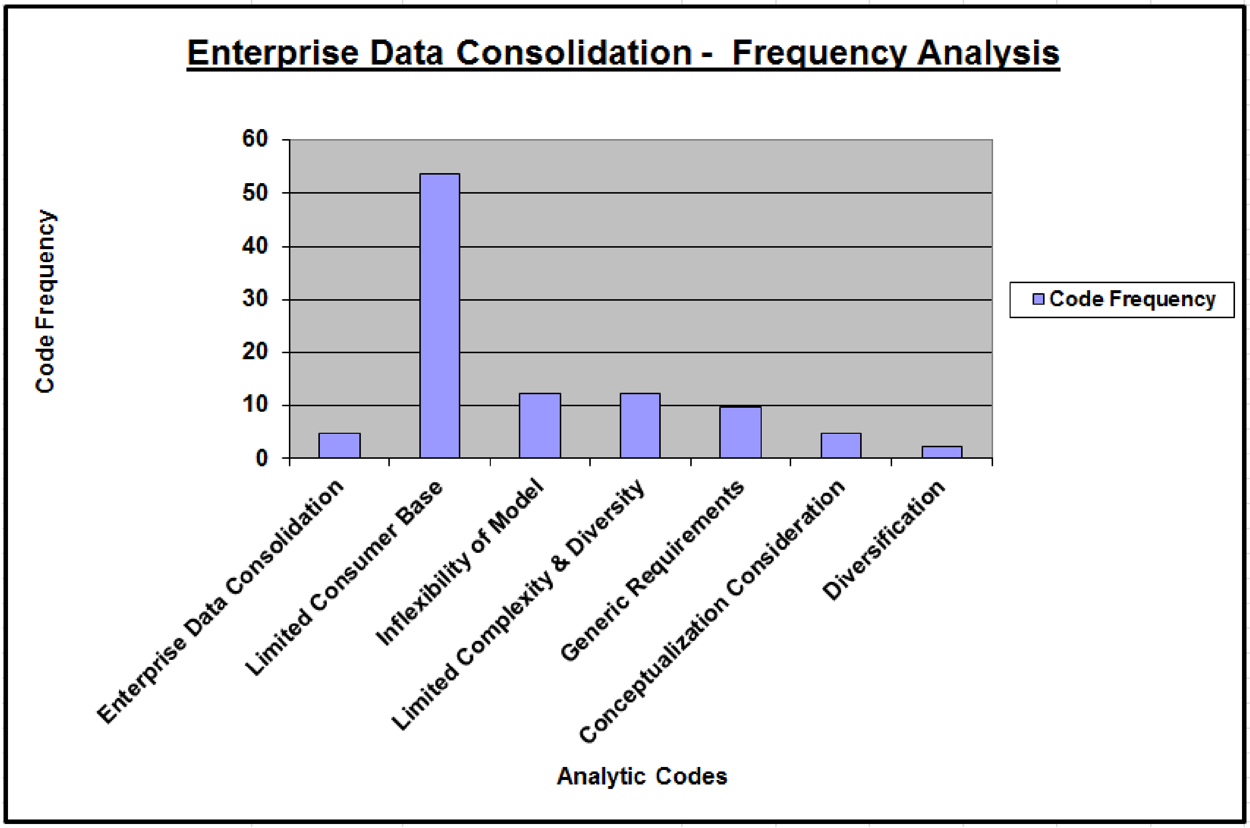

Figure 4-15: Impact of Enterprise Data Consolidation on Logical Data Model Decision

In Figure 4-15, the responses from the research participants were coded along the x-axis, the analytic codes; the percentage code frequency of each of the analytic codes is on the y-axis. The research participants were asked about the diversity of consumer base of the reporting warehouse as a measure to assess the enterprise orientation of the reporting data warehouse in meeting different consumer groups reporting and analytics needs.

The research participants stated 53.6 per cent of the time that the reporting data warehouse has limited enterprise data consolidation. This number is high, indicating that more than half of the respondents at GWealth believed the reporting data warehouse has limited number of users across their organisation. The reporting warehouse also scored low on other analytic codes to assess the enterprise orientation of GWealth reporting warehouse. For instance, *Enterprise Consumer Base* and *diversification of the data model* accounted for just 4.8 and 2.4 per cent of the total percentage code frequency, indicating that the respondents did not consider them as decision factors for adopting the multidimensional data model for client reporting data warehouse. The frequency analysis of the enterprise analytic codes above indicated the consumer base of the reporting warehouse is not diverse and in alignment with the research participants' statement that the reporting warehouse was not created to consolidate or distribute common enterprise data to other consumers groups across GWealth. Consequently, the user base of the reporting warehouse is limited to the team tasked with producing reports for the clients of the firm, the primary goal set by the business for the reporting data warehouse. The research participants at GWealth indicated they tightly executed the business requirements for the reporting data warehouse and did not give any consideration to other consumer groups "*on-boarding*" the reporting data warehouse in line with the business requirements and expectation. The decision to adopt a multidimensional data model for the reporting warehouse was influenced by the requirement from the business for client reporting as opposed to catering for the generic needs of different consumer groups within GWealth. The business requirement for the reporting warehouse prompted the respondents to engage a course of action based on their judgements and assessment of GWealth requirements in alignment with decision making in naturalistic decision theory.

The data presented above provided the evidence to support the proposition that consolidating common enterprise data to address different reporting and analytics requirements of heterogeneous consumer groups is not a factor in adopting a multidimensional data model for a data warehouse.

4.5 Summary of the Findings

This chapter presented the findings of the case study at GWealth; the chapter presented the result of the factors that influenced the decision to adopt a multidimensional data model for GWealth reporting data warehouse. The findings of the case study are summarised below.

4.5.1 Finding 1: *High Query Performance is a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse*

One of the non-functional requirements of the reporting warehouse that featured prominently in the participant responses was the need to quickly process client reports from the reporting data warehouse. An important element of ensuring the client reports were quickly processed is the need to ensure the reporting warehouse data model is portable and optimized for high performance query execution. The data from the case study indicated the importance of performance to the business and the expectation of getting high performance out of the reporting data warehouse. The research participants believed their multidimensional data model enabled them to achieve the performance requirement of their data warehouse.

The observation from the empirical data showed that high proportion of the research participants indicated that performance of query execution was a decision factor in adopting the multidimensional data model for the reporting data warehouse. The empirical data showed that GWealth research participants stated 79 per cent of the time that performance is a key decision factor in using the multidimensional data model for their reporting data warehouse; this is consistent with the literature. In the literature, a multidimensional data model is seen as the technique that presents data in a standard intuitive framework that allows for high performance (Kimball, 1995; 1977). The research participants at GWealth believed the multidimensional data model played a key role in achieving the requirement of performance, having considered the infrastructure supporting their reporting data warehouse. Additionally, the empirical data also indicated the percentage code frequency of performance as a decision factor for implementing the GWealth reporting data warehouse on a multidimensional data model is more than double the combined percentage code frequency of the analytic codes related to: *Inflexibility of Model*, *Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration*, *Enterprise Data Consolidation*, and *Diversification*. In contrast, the percentage code frequency for these decision factors is quite low, collectively; they represented 21 percent of the total decision factors of the observed data for performance.

4.5.2 Finding 2: *Specific Requirement is a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse*

One of the important factors that influenced the decision to implement the GWealth data warehouse on a multidimensional data warehouse was the requirement to build a reporting solution for organisation. The study finds the requirement from the business to create a reporting data warehouse rather an enterprise data warehouse was major factor in the decision to adopt a multidimensional data model for GWealth reporting data warehouse. The empirical data indicated the business requirement had a big impact in influencing the choice of reporting data warehouse logical data model. The research participants stated 81.5 per cent of the time that the reporting data warehouse data model was adopted to address the specific requirement of the business for client reporting, this is consistent with the literature. In the literature, a multidimensional data model is used to implement a data warehouse that supports the specific functionality of the business, as such, it is considered excellent for data processing and ensures good performance for complex slice and dice operations (Imhoff, Galemmo and Geiger, 2003; Kimball and Ross, 2002). Further examination of the empirical data indicated that the percentage code frequency of the analytic codes related to: *Inflexibility of Model*, *Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration*, *Enterprise Data Consolidation*, and *Diversification* is low as a decision factor; collectively; they represented 19 percent of the total decision factors for using multidimensional data model for the reporting data warehouse. Further examination of the empirical data showed that high proportion of the research participants indicated the requirement from the business to create a reporting data warehouse rather an enterprise data warehouse influenced the decision to adopt a multidimensional data model for the reporting data warehouse. The research participants indicated that they found multidimensional data model to be perfect for the type of requirement to create a reporting data warehouse for their organisation.

4.5.3 Finding 3: *Goal and Scope is a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse*

The goal and scope of a data warehouse is an important factor that impacts the choice of logical data model for a data warehouse. The research participants at GWealth recognised the risk involved in choosing a data model for a data warehouse without due consideration for the goal of the data warehouse. The empirical data indicated the business defined goal and limited scope of the reporting warehouse was in alignment with the logical data model adopted for the GWealth data warehouse. The respondents strongly believed the data model for the reporting data warehouse was selected in alignment with the goal and scope defined by their organisation. In the literature, the goal of a data warehouse ultimately defines the scope of a data warehouse. The scope of a data warehouse where the goal of the data warehouse is oriented toward a department within an organisation assumes the requirements for a multidimensional data warehouse emerges from a business area requesting the data warehouse (Kimball and Ross, 2004). The empirical data from GWealth indicated that the research participants stated 77 per cent of the time that the goal and scope of their reporting data warehouse was a factor in adopting a multidimensional data model for the reporting data warehouse. In contrast, the percentage code frequency for the analytic codes related to: *Inflexibility of Model*, *Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration*, *Enterprise Data Consolidation*, and *Diversification* accounted for 23% of the total decision factors for using multidimensional data model for the GWealth reporting data warehouse.

4.5.4 Finding 4: *Staff Experience is a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse*

The implementation orientation of the staff tasked with building a data warehouse is an important factor that impacts the choice of logical data model for a data warehouse. The empirical data from the study indicated that GWealth staff leveraged their prior experiences and drawn on it to address the requirements to build the client reporting data warehouse for their organisation. The participants attributed that a major benefit of their multidimensional data model was that it enabled the participants to focus what they set out to accomplish which was to build the reporting warehouse. The empirical data indicated that the research participants at GWealth stated 81 per cent of the time that their staff experience is a decision factor in adopting a multidimensional data model for the reporting data warehouse. In contrast, the percentage code frequency for other decision factors (*Inflexibility of Model*, *Limited Complexity & Diversity*, *Generic Requirements*, *Conceptualization Consideration*, *Enterprise Data Consolidation*, and *Diversification*) is quite low, collectively; they accounted for just 19 percent of the total decision factors. The research finds that the research participants at GWealth were keen advocates of a multidimensional data model, they expressed their preference for engaging a multidimensional data model for their data warehouse and rarely would they use any other type of data model to build a data warehouse. Vassiliadis (2004) notes that data warehousing landscape is defined by "*do-it-yourself*" advice from experts and proprietary vendor solutions, the empirical data indicated that the research participants at GWealth have considerable experience in multidimensional data model and have mostly used the data model in their careers.

4.5.5 Finding 5: *Enterprise Data Consolidation is Not a Decision Factor in Adopting Multidimensional Data Model for GWealth Data Warehouse*

Enterprise data consolidation to address different reporting and analytics requirements of consumer groups within an organisation is an important consideration impacting the choice of logical data model for a data warehouse. The research finds the consumer base of the reporting warehouse is not diverse because the reporting warehouse was not created to consolidate nor distribute common enterprise data to other consumers groups across GWealth. As a result, the user base of the reporting data warehouse is limited to the team creating reports of statements from the reporting data warehouse. The empirical data indicated that the GWealth data warehouse is not an enterprise data warehouse nor is it a distributor of enterprise data to other consumer groups across GWealth. The observation from the empirical data indicated that lower proportion of the research participants indicated that the decision to adopt a multidimensional data model for reporting data warehouse was influenced by the requirement from the business to cater for the generic needs of different consumer groups within GWealth. In the literature, one of the arguments often recognised as the basis for implementing a data warehouse is the need to capture diverse set of business data in a data warehouse (Inmon, 2005; Kimball, 2002; Tryfona, Busborg and Christiansen, 1999; Weininger, 2002, Husemann, Lechtenbörger, Vossen, 2002). However, integrating enterprise data for various reporting and analytics requirements requires flexible data model that is oriented to the enterprise view of an organisation and focuses on complex data structures and interactions (Imhoff, Galemmo and Geiger, 2003; Tryfona, Busborg and Christiansen, 1999). The empirical data indicated that the research participants at GWealth stated only 4.8 per cent of the time that enterprise data consolidation to address different reporting and analytic requirements is a decision factor in adopting a multidimensional data model for the reporting data warehouse. In contrast, the research participants stated 53.6 per cent of the time that their reporting warehouse has limited enterprise data consolidation indicating that the majority of the research participants believed the use of reporting data warehouse is limited to the reporting team using the solution to produce statements of reports for the clients of their organisation.

CHAPTER 5 - Case Study: ICapital

5.1 Introduction

This chapter presents the outcome of the case study at ICapital, the investment banking division of global financial institution TBank. The chapter focuses on presentation of interview results and alignment of the findings with the research propositions. The data collection method for this research is through interview of research participants. In presenting the result of this case study, the interview data was transcribed and transformed through grounded theory open coding methods involving line-by-line coding, translation of the descriptive codes into analytic codes, repeated comparison of the descriptive and analytic codes for saturation, and identification of core themes from the analytic categories.

This chapter is divided into three Sections; Section 5.2 provides an overview of ICapital including its key business areas. The Section looks at the background of ICapital data warehouse project, the project definition and objectives including the architecture of ICapital Risk and Finance data warehouse. Section 5.3 presents the finding of the case study and the alignment of the interview results with the research propositions. Lastly, Section 5.4 concludes the chapter with a summary.

5.2 Overview of ICapital

ICapital is the investment-banking arm of a major European financial institution, TBank. ICapital has offices around the world and employs over 25,000 people ensuring the company has the global reach, advisory services and distribution power to meet the needs of issuers and investors world-wide. ICapital generated £5.231bn for 2011 financial year and the company more than doubled its income to £11.625bn in 2012. ICapital is organised around a number of businesses including Fund Solutions, Infrastructure Funds, Natural Resource Investments, Commodities, Emerging Markets, Equities, Fixed Income, Foreign Exchange, Capital Market, Liquidity Management, Mergers and Acquisitions, Re-Structuring, Portfolio Management, Prime Services, Quantitative Analytics and Research.

ICapital Fund Solutions was formed in 2005 and it's the asset management business of ICapital. The fund has asset under management in excess of £4bn as of October 2012, it runs its own trading, structuring, quantitative portfolio modelling and discretionary management functions. ICapital Fund Solutions employs over 120 people, and has offices in London, New York, Singapore, Tokyo, Sidney and Hong Kong. The business distributes funds to clients in over 20 countries across 5 continents. ICapital Infrastructure Funds is a fund management division of ICapital; it has been in operation since 1996 and focuses on investment in social infrastructure projects in the UK and continental Europe.

Social infrastructure projects include construction, long term maintenance and management of core public infrastructure such as primary and secondary schools, community healthcare facilities, large acute hospitals, emergency services facilities, local and central government offices. These projects are mainly financed under the Private Finance Initiative (PFI) and Public Private Partnership (PPP) frameworks in the UK and continental Europe. ICapital Natural Resource Investment is a global private equity business focusing on natural resource investment opportunities that provides clients of its parent company (or its affiliates) and strategic investors the opportunity to gain exposure to the natural resources sector by co-investing alongside the ICapital. ICapital Natural Resource Investment invests in global resources by finding and partnering with "*best-in-class*" operational management teams with specific expertise and significant experience and track-record within the niche sectors of the global natural resources sector, primarily in upstream oil and gas, upstream mining, and power, including renewables.

ICapital global sales force provides institutional investors with highly focused, round the clock, team-based service, ranging from specialist coverage on fixed income, foreign exchange, commodities (precious metals and energy), financial futures and derivatives to generalist, multi-product sales coverage across the ICapital product base. ICapital is also acknowledged as one of the leaders in the major commodity asset classes, and remains at the forefront of the industry. As one of the leading providers of commodity solutions, ICapital fields global commodity teams comprising 350 trading, sales and research professionals.

ICapital offers one of the most extensive product suites in the industry, including specialist products and bespoke risk management solutions, with breadth and depth of service unparalleled by many competitors. ICapital is top 3 emerging markets fixed income house and continues to build out its emerging market franchise by scaling up its local footprint in selected locations such as Brazil, Korea, India, Israel, Russia and the Middle East. This strategy enables ICapital to capture opportunities with local client base as well as enhancing the product offering and servicing of offshore clients investing in emerging markets. ICapital emerging-markets product offering spans Sovereign, Corporate Debt and Credit-Default Swaps. It also offers local debt, Interest Rate Swaps, FX, Structured Products, Equities, Indices, as well as Financing and Liability management solutions. ICapital currently has offices in the following emerging market locations: Mexico, Brazil, Argentina, Russia, UAE, Qatar, Saudi Arabia, Israel, China, Hong Kong, Taiwan, Korea, Thailand, Malaysia, Singapore, India and South Africa.

5.3 Background of the Data Warehouse Project

The business rationale for Risk and Finance data warehouse (RFDW) is to address data quality problem across Risk, Finance and Collateral Management functions of ICapital. The rationale called for streamlining data collection processes across Risk, Finance and Collateral Management. Before RFDW was implemented, there were multiple data transformations across the organisation; this makes the original data unrecognisable to its source thus, complicating discussions with the business teams. Additionally, the lack of well-defined data processes impacts the quality and timeliness of information available to the business.

One of the areas where there is a major concern for the business teams is data quality. The lack of proper data quality program within the organisation resulted in inconsistent adjustments and valuation of risk within Risk, Finance and Collateral Management functions of the organisation. Additionally, it was difficult to engage any root cause analysis and little opportunity to get problematic data fixed at the source. This resulted in many areas of the business working from inconsistent and often incomplete datasets. It was also recognised by the firm that bespoke information models limit the bank's ability to provide ad-hoc analysis, including holistic stress testing and one-off requests for external information by the regulatory authorities.

Furthermore, Risk and Finance teams working off their own generated datasets limits the bank's ability to show the relationship between risk, profit and loss (P&L), capital information and client valuations; thus, limiting the bank's ability to provide transparency into these measures. Although multiple data warehouses were considered as part of the legacy acquisition front-to-back (F2B) project and legacy ICapital enterprise warehouse strategic projects, the value proposition is to have one firm-wide data warehouse enabling:

- Increased accuracy and consistency of end of day transactions, positions, risk and P&L across the *Front Office*, *Middle Office*, *Risk*, *Finance and Collateral Management* functions.
- Ability to integrate with a central data quality platform, which will measure quality of source data, identifies and correct root cause.
- One stop shop for *Risk*, *Finance and Collateral Management* teams to cross-reference, slice and dice and report on end of day information.

5.3.1 Project Definition and Objectives

The purpose of RFDW project is to provide a single and consistent dataset for *Market Risk*, *Regulatory Reporting*, *Collateral Management*, *Credit Risk and Finance* Systems. Risk and Finance data warehouse was integrated within the Strategic Architecture Program (SAP), and it is the firm-wide source of end-of-day (EOD) data for Risk and Finance teams at ICapital. The primary objectives of RFDW are:

- To provide single consistent end-of-day dataset for *Market Risk*, *Credit Risk*, *Collateral Management and Finance* functions
- To provide rich information model which will support *Market Risk*, *Credit Risk*, *Collateral Management and Finance* functions with data access tools to support efficient and effective data retrieval
- To provide next level definition of the End of Day (EOD) control processes and integration into the strategic architecture
- To integrate with the firm's central Data Quality platform, this provides data score-carding capabilities.

The ICapital Risk and Finance data warehouse project was supported by four work streams; they included the program management, front office integration and data quality, information model and risk finance integration. The project management work stream is responsible for delivering and monitoring the overall program plan, tracking progress and monitoring the actions, risks and issues throughout the development and delivery phases of the project.

The front office and data quality work stream is responsible for implementing timeline for sourcing data related to *Trades*, *Positions*, *Valuations*, *Explains* and *Risks* into the data warehouse. The information model work stream is responsible for designing and implementing a converged information model to support all of the risk and finance processes using a relational data architecture model. Finally, the risk finance integration work stream is responsible for defining strategy, requirements, implementation and integration of a number of data marts for *Finance*, *Credit Risk*, *Market Risk*, *Regulatory Reporting* and *Collateral Management* for RFDW.

5.3.2 The Data Warehouse Architecture

There are four major components that make up the architecture framework for ICapital Risk and Finance data warehouse. The components include:

- RFDW Data Sourcing Layer
- RFDW Information Model
- RFDW Event Manager
- RFDW Client Interface

5.3.3 RFDW Data Sourcing Layer

The RFDW data sourcing layer is responsible for interfacing with the upstream systems to collect data related to *trades* and *positions*, *explains*, *valuations*, *market data* and other *risk sensitivities*. The data source layer is responsible for collecting the specific business concept data i.e. trades data in "*ICapitalML*" format and sends the file to appropriate message queue (MQ) channel for processing. The "*ICapitalML*" is a XML file specification for exchanging and processing data within ICapital.

5.3.4 RFDW Information Model

The RFDW contains a number of business concepts normalized into a relational model to accommodate the complexity of *trades*, *settlements* and *accounting* entities that make up the data warehouse. Table 5-1 outlines some of the business concepts and entities in RFDW data model.

Risk and Finance Data Model Business Concept
Account
Books
Cash flow
Counterparty
Country
Currency
General Ledger
Industry

Instruments
Journal Postings
Journal Types
Legal Entities
Location
Market Data
Netting Agreement
P & L Item
Product
Ratings
Region
SAP
Securities
Sub Ledger Account/GL
Trade
Traded Exchange
Valuation

Table 5-1: Risk and Finance Data Warehouse Common Business Data *(Adapted from ICapital Artifacts)*

5.3.5 RFDW Event Manager

The event manager provides the central communication hub between the data source layer, the data warehouse and the downstream consumer systems. In RFDW, the data source layer receives data via reference message pointer from the data warehouse, the data source layer then posts "*Write Confirmation*" message to the data warehouse. The data source layer receives control messages from RFDW when journal-posting transaction is completed. The data source layer creates INFO_EVENT entry in the data warehouse; the INFO_EVENT contains "*query-able*" data in RFDW linked to a given information event. Event types related to *trades* and *positions*, *valuation* updates and lifecycle events are available in the data warehouse event manager. Figure

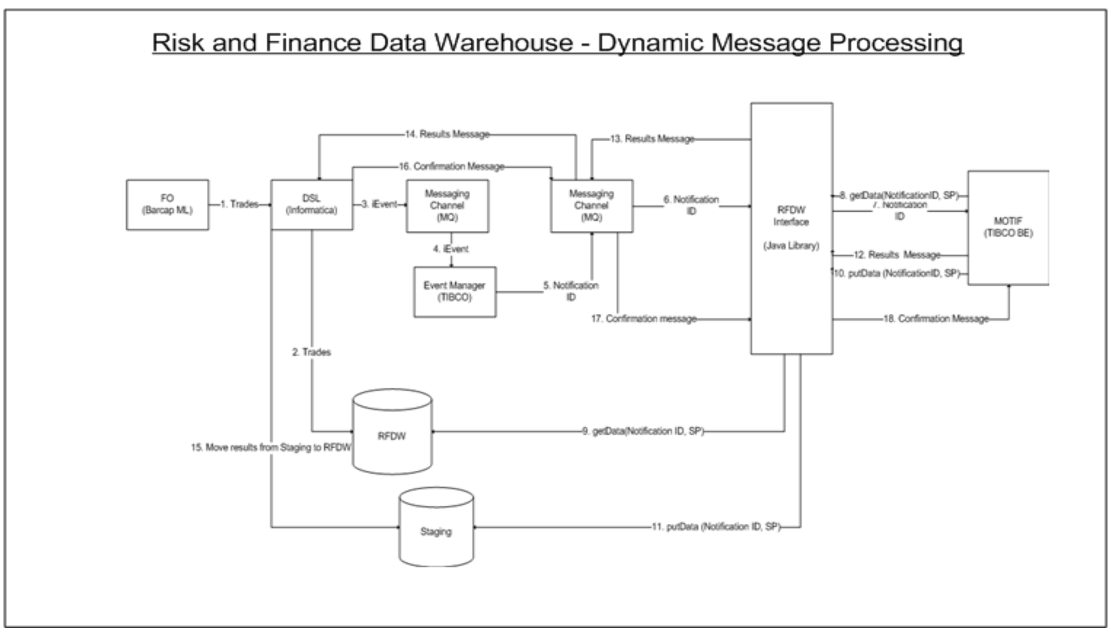

Figure 5-1: Event Manager Dynamic Messaging Processing *(Adapted from ICapital Artifacts)*

5.3.6 RFDW Client Interface

The Risk and Finance data warehouse interface provides a single interface for obtaining data from RFDW. The interface provides notifications to data consumers upon data readiness, and allows the data consumer to access the data in the data warehouse. Additionally, the interface allows write-back of accounting adjustments into Risk and Finance data mart.

The next Section presents the implementation of research design for this study.

5.4 Implementation of Research Design for ICapital

Research design is the blueprint of case study research; it is a plan for assembling, organising and evaluating information according to the problem definition and specific goals for how to use a research finding (McCoy et al, 1993). This Section describes the open coding activities carried out to identify the ICapital analytic category codes by transforming the empirical data collected from the ICapital research participants.

5.4.1 Transcription of ICapital Data and Applying Code to Text

As indicated in Section 3.5.6, a recording device was used to capture the interview sessions with the research participants at ICapital.

Following each interview, empirical data was transcribed into a readable text using Microsoft Word. A sample of transcribed interview data is illustrated below in Figure 5-2.

TRANSCRIBED EMPIRICAL DATA

KO

What is the end-user ability to understand the data relationships in Risk and Finance Data Warehouse?

MRRMBC

Not exactly, the approach to the model, for me, the business have looked at the model, but we show that model to the business it was at extremely high level, we showed the client, trade, party, we showed issuer, things the business is interested in we didn't try to define in a manner that you would think about in physical or logical model. We want our model to be about how fast are we able to enter and read information – as simple as that. Physical implementation is all about the access; correct? It wasn't about whether you can understand what you going to use, it wasn't about any of that stuff. It's all about fast inserts and spontaneous inserts and reads. So we have a high level model which is very high level to the major pieces of information and within each of that high level, you are able to drill into the next detail level of that model – correct? And I think we want to the personal level that is the high level to the second level that is already at the detailed level and we kind of arrive at ER model to allow us to keep information which primarily finance is interested in, risk is interested in. We made the attributes that are associated with that, for example, balance and all of the pieces that hang off of it for example the instrument, counterparty of the balance etc.; so we look at it from the business domain and we said finance is interested in balance, risk is interested in sensitivities and all the other pieces of information related to sensitivities – correct? So, the only place that we sort of really weren't closing the deal is on the finest grain of the model is for FLEX- that's a good model.
And in the case of marts for FLEX, we really sat down with the end users in understanding exactly the product information that they are after for reporting and we created that model which is pure mart model, dimensional model; they get three level of dimensions of balances, at trade level, at the party level and at the group level, right? So we have 3 levels and all the facts and dimensions around it; there are some agreed matrixes; there is a matrix saying we can agree 82000 rows every second, we can write 32000 rows every second so we build our model so that we can achieve that.

Figure 5-2: Transcribed Empirical Data from ICapital Research Participant

As indicated in the research design for this study, the next step carried out by the investigator as part of the open coding process to identify the thematic categories for this research was line-by-line coding of transcribed data. Line-by-line coding for this study involved applying descriptive code to each row of text of transcribed data. A descriptive code is a description of what is happening in a particular row of text that is interpreted by the investigator; therefore, a descriptive code captures the essence of a row of text (Gibbs, 2007) (Gibbs, 2007). For this research, the investigator imported transcribed data from MS Word into MS Excel where line-by-line coding was carried out. A sample of line-by-line coding in MS Excel is illustrated in Figure 5-3 below

EMPIRICAL DATA: LINE-BY-LINE CODING

Transcribed Data	Apply Code to Text: Descriptive Code
but we show that model to the business it was at extremely high level, we showed the client,	The business looked at the model at high level
trade, party, we showed issuer, things the business is interested in we didn't try to define in a	We showed the business what they are interested in
manner that you would think about in physical or logical model. We want our model to be about	Model not defined to business in typical logical or physical manner
So we have a high level model which is very high level to the major pieces of information and	Major pieces of information is model at high level
within each of that high level, you are able to drill into the next detail level of that model –	The data model allows you to drill to next level of detail
correct? And I think we want to the personal level that is the high level to the second level that is	The model has personal and second level of information
already at the detailed level and we kind of arrive at ER model to allow us to keep information	Arrived at ER to address requirements
which primarily finance is interested in, risk is interested in. We made the attributes that are	The model contains information that finance and risk are interested in
associated with that, for example, balance and all of the pieces that hang off of it for example the	The data model contains balance information
instrument, counterparty of the balance etc.; so we look at it from the business domain and we	The data model also contains instrument, counterparties and balances
said finance is interested in balance, risk is interested in sensitivities and all the other pieces of	Finance is interested in balances and risk is interested in sensitivities
information related to sensitivities – correct? So, the only place that we sort of really weren't	Sensitivities and other related information is fact
closing the deal is on the finest grain of the model is for FLEX- that's a good model.	Addressing the grain of FLEX reporting, it's a good model
And in the case of marts for FLEX, we really sat down with the end users in understanding	Sat down with business users to understand their requirements
exactly the product information that they are after for reporting and we created that model which	Sat down with business users to understand their requirements
is pure mart model, dimensional model; they get three level of dimensions of balances, at trade	Created model that addressed information required by business
level, at the party level and at the group level, right? So we have 3 levels and all the facts and	Created model that addressed information required by business
dimensions around it; there are some agreed matrixes; there is a matrix saying we can agree	Created model that addressed information required by business
82000 rows every second, we can write 32000 rows every second so we build our model so that	Created model that addressed information required by business
we can achieve that.	Created model that addressed information required by business
Yeah, I would say absolutely true, correct? So you have to create data architecture, all the access	Model created to address business requirements
on the system right? You have to understand all the loading and the reading access of	You need to understand all your requirements
comparable systems if you don't articulate that, you end up with a system that is un-useable,	You need to understand your requirements to build useable system
that cannot grow overtime, yes I agree with that statement	You need to understand all your requirements
that very well. So I would say in the case of RFDW, we set out the left side and the right side as	Data model contain right side and left side information
you might have known; the left side is the information address driving just-in-time information,	Left side contains just-in-time information

Figure 5-3: Line-By-Line Coding of Transcribed Empirical Data

5.4.2 Convert Descriptive Codes to Analytic Codes and Compare Codes for Saturation

The next step carried out by the investigator as part of the open coding process to identify the thematic categories for this research was transforming the descriptive codes into analytical codes. Gibbs (2007) points out investigators should avoid descriptive codes; instead, investigators should formulate analytic codes, therefore, an analytic code is a conceptualization of the descriptive codes by the investigator. The following descriptive codes in Table 5-2 were transformed into the analytic codes in Table 5-3 as part of the process of identifying the category code *Generic Requirements*.

	Line-By-Line Coding - Descriptive Code of Empirical Data
1	Business looked at the model at high level
2	Showed the business what they are interested in
3	Data model not defined to business in typical logical or physical manner
4	Major pieces of information is model at high level
5	Data model allows you to drill to next level of detail
6	Data model has personal and second level of information
7	Data model has ER and dimensional component
8	Data model contains information that finance and risk are interested in
9	Data model contains balance information
10	Data model contains details related to instrument, counterparties and balances

Table 5-2: Descriptive Codes for Line-By-Line Coding

	Analytic Code for Descriptive Codes in Figure 5-2
1	Data model reviewed with business at high level
2	Data model addressed generic business requirements
3	Data model reviewed with business at high level
4	Data model addressed generic business requirements
5	Data model enables drill down of information
6	Data model enables drill down of information

7	Data model design is generic
8	Data model addressed generic business requirements
9	Data model addressed generic business requirements
10	Data model addressed generic business requirements

Table 5-3: Analytic Codes for Descriptive Codes

As indicated in the research design for this study, the next step engaged by the investigator was ensuring texts with similar themes were assigned the same code; coded passages were further reviewed and the codes compared for saturation. Figure 5.4 illustrates the process of transforming the descriptive codes in Table 5-2 to the analytic codes in Table 5-3; the process ensured texts with similar themes were assigned the same code using MS Excel.

TRANSFORMING DESCRIPTIVE CODES TO ANALYTIC CODES

Apply Code to Text: Descriptive Code	Descriptive Code -> Analytic Code	Compare Codes for Saturation
The business looked at the model at high level	Model reviewed with business at high level	Model addressed generic business requirements
We showed the business what they are interested in	Model addressed generic business requirements	Model addressed generic business requirements
Model not defined to business in typical logical or physical manner	Model reviewed with business at high level	Model addressed generic business requirements
Major pieces of information is model at high level	Model addressed generic business requirements	Model addressed generic business requirements
The data model allows you to drill to next level of detail	Model enables drill down of information	Model addressed generic business requirements
The model has personal and second level of information	Model enables drill down of information	Model addressed generic business requirements
Arrived at ER to address requirements	Model addressed generic business requirements	Model addressed generic business requirements
The model contains information that finance and risk are interested in	Model addressed generic business requirements	Model addressed generic business requirements
The data model contains balance information	Model addressed generic business requirements	Model addressed generic business requirements
The data model also contains instrument, counterparties and balances	Model addressed generic business requirements	Model addressed generic business requirements
Finance is interested in balances and risk is interested in sensitivities	Model addressed generic business requirements	Model addressed generic business requirements
Sensitivities and other related information is fact	Model addressed generic business requirements	Model addressed generic business requirements
Addressing the grain of FLEX reporting, it's a good model	Model addressed generic business requirements	Model addressed generic business requirements
Sat down with business users to understand their requirements	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Sat down with business users to understand their requirements	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Model created to address business requirements	Model addressed generic business requirements	Model addressed generic business requirements
You need to understand all your requirements	Business requirements must be understood	Model addressed generic business requirements
You need to understand your requirements to build useable system	Business requirements must be understood	Model addressed generic business requirements
You need to understand all your requirements	Business requirements must be understood	Model addressed generic business requirements
Data model contain right side and left side information	Data model contains generic information	Model addressed generic business requirements
Left side contains just-in-time information	Data model contains generic information	Model addressed generic business requirements

Figure 5-4: Transformation of Descriptive Codes to Analytic Codes

5.4.3 Identify Sub Category Codes and Main Category Codes

Category codes are logical groupings of the thematic analytic codes; a category code facilitates the retrieval of the underlying codes (Gibbs, 2007). As indicated in Section 3.5.5, the next step carried out by the investigator was to identify a category which a set of analytic codes aligned. The following categories were identified as the initial set of category codes for this case study (*High Query Performance*, *Specific Requirement*, *Generic Requirements*, *Defined Goal & Scope*, *Staff Experience*, *Data Consolidation* and *Diverse Consumers*). Analytic codes which did not aligned with the category codes outlined above were regarded as standalone category codes. The initial category codes were further distilled to identify top level category codes or main category codes which the initial set of category codes aligned; this ensured that category codes rolled up to a top level category code. As a result, the initial category codes (*High Query Performance*, *Specific Requirement*, *Generic Requirements*, *Defined Goal & Scope*, *Staff Experience*, *Data Consolidation* and *Diverse Consumers*) became sub-category codes. As part of the open coding process for this research, the following main analytic category codes (*High Performance of Query Execution, Business Requirement, Firm Objective, Employee Experience* and *Enterprise Data Hub*) were identified for this case study. In addition, the following analytic codes (*Model Inflexibility*, *Limited Complexity & Diversity*, *Conceptualization Consideration*, *Diversification*) were standalone category codes, as indicated above; these analytic codes could not be aligned with any of the sub category codes or main analytic category codes. Figure 5.5 and Figure 5.6 respectively illustrates sample of sub category codes (***Generic Requirements, Specific Requirement***) and main analytic category code (***Business Requirement***) for ICapital case study.

IDENTIFICATION OF SUB-CATEGORY CODES

Descriptive Code -> Analytic Code	Compare Codes for Saturation	Sub Category Code
Model reviewed with business at high level	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model reviewed with business at high level	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model enables drill down of information	Model addressed generic business requirements	Generic Requirements
Model enables drill down of information	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Business requirements must be understood	Model addressed generic business requirements	Generic Requirements
Business requirements must be understood	Model addressed generic business requirements	Generic Requirements
Business requirements must be understood	Model addressed generic business requirements	Generic Requirements
Data model contains generic information	Model addressed generic business requirements	Generic Requirements
Data model contains generic information	Model addressed generic business requirements	Generic Requirements

Figure 5-5: Sub-Category Codes (*Generic Requirements, Specific Requirement*) for the Analytic Codes

MAIN CATEGORY CODES FOR SUB-CATEGORY CODES

Sub Category Code	Main Category Code
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement

Figure 5-6: Main Category Codes for Generic Requirements Sub-Category Codes

5.4.4 Identify Property of ICapital Category Codes

As indicated in the research design framework presented in Section 3.5.5, categories have properties, properties enable analytic category codes to be viewed from multiple perspectives (Gibbs, 2007). As part of the open coding process for this study, sub category codes developed in Section 4.3.3 rolled up to the main category codes, thus, the sub category codes became the properties of top level category codes for this case study. This enables the main category codes to be viewed from multiple perspectives. For instance, the category code *Business Requirement* has two properties *Generic Requirements* and *Specific Requirement* as illustrated in Figure 5.7.

MAIN CATEGORY CODE: PROPERTIES

Main Category Code	Property of Main Category Code
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements

Figure 5-7: Property of ICapital Category Codes

5.4.5 Analytic Codes and Frequency Count of Analytic Codes

As part of the open coding process for the study, a count of frequency occurrence of the analytic category codes was developed, the count of frequency of an analytic code is the number of times an analytic category code is noted in the empirical data. This was done by counting the number of times each instance of an analytic category code appeared in the empirical data, the counts were then aggregated in MS Excel to provide total frequency occurrence for an analytic category code; this process is repeated for all the analytic codes. As indicated in Section 3.5.8, analysis of frequency count of analytic codes is an important step in linking the empirical data with the research propositions. Figure 5.8 illustrates the build of frequency occurrence of an analytic code (*Generic Requirements*) for this study.

FREQUENCY OCCURRENCE OF ANALYTIC CODE

Analytic Code	Frequency Occurrence
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
...	...
Generic Requirements	1
Generic Requirements	1
Total	381

Figure 5-8: Frequency Occurrence of Analytic Code

5.5 Presentation of the Results - ICapital Case Study

The interviews at ICapital were conducted with eleven participants, two of whom were senior data architects at director level, three senior developers, one program manager, one project manager, two development managers and two business analysts. The directors and the program manager were part of the funding group at ICapital, the data architects, developers, project managers and development manager were part of the execution team responsible for implementation and development of Risk and Finance data warehouse at ICapital. This chapter is a result chapter; it presents the result of the interviews at ICapital and aligns the empirical data with the research propositions. Chapter 6 of the thesis is devoted to data analysis of this comparative case study. Table 5-2 below presents the frequency occurrence of the analytical codes for ICapital.

Analytical Codes	**No of Frequency**	**Cumulative Number**	**Cumulative Percentage**
Enterprise Data Consolidation	436	436	33.8
Generic Requirements	381	817	63.3
Defined Goal & Scope	197	1014	78.6
Staff Experience	166	1180	91.5
High Query Performance	87	1267	98.2
User Group Experience	10	1277	99.0
Specific Requirement	5	1282	99.4
Data Conceptualization	4	1286	99.7
Inflexibility of Model	2	1288	99.8
Simplicity of Model	2	1290	100
Total	1290		

Table 5-4: Frequency Analysis of the ICapital Analytic Codes *(See Appendix 4 for ICapital Open Codes)*

The frequency occurrence of each of the analytic codes for ICapital is presented in Table 5-2 above. The column "*No of Frequency*" contains the number of times an analytic code is noted as a decision factor in adopting the relational data model for ICapital Risk and Finance data warehouse. The column "*Cumulative Number*" is the aggregation of each of the frequency number or frequency count of each of the analytic code for ICapital. The column "*Cumulative Percentage*" provides the percentage value for the "*Cumulative Number*". Figure 5-9 below presents the data in Table 5-2 in a graphical format using the *Pareto* chart.

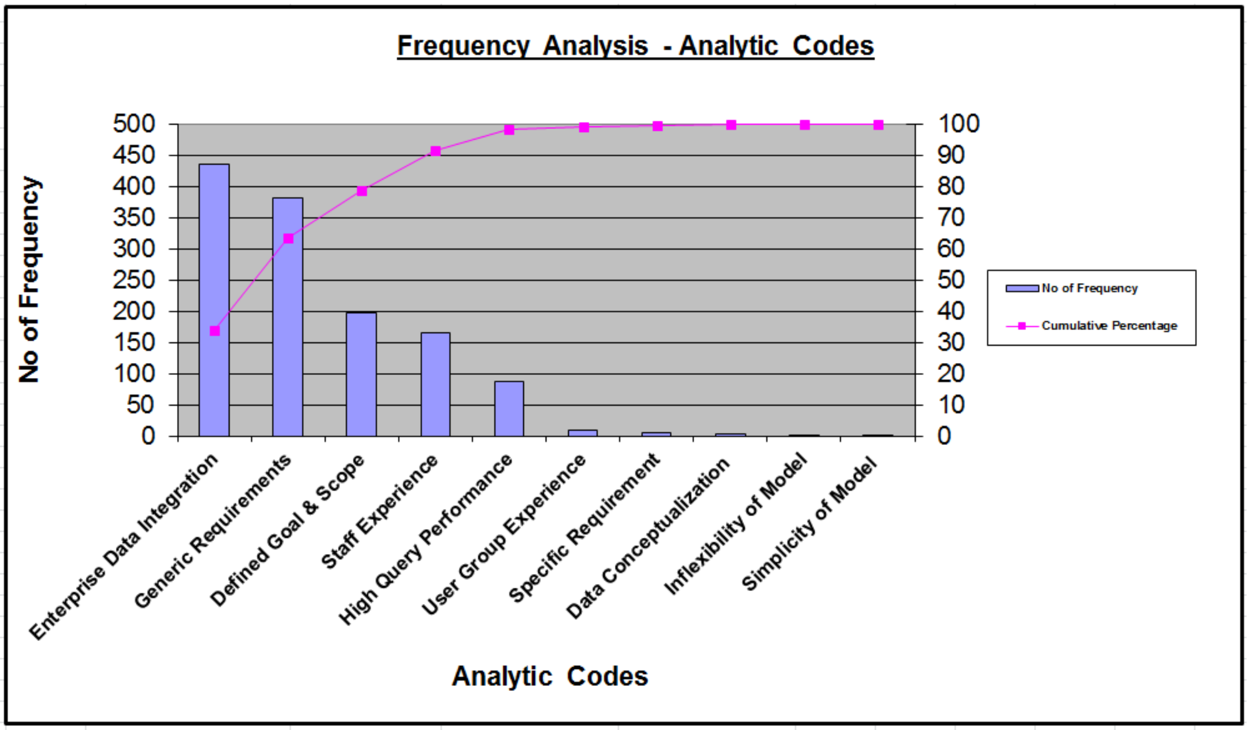

Figure 5-9: Frequency Analysis of the Analytical Codes Using the Pareto Chart

The above Pareto chart presents the frequency analysis of the analytic codes for ICapital case study. As illustrated in the graph, the analytic codes related to: *Enterprise Data Consolidation*, *Generic Requirements*, *Defined Goal and Scope*, *Staff Experience* and *Performance* accounted for 98 per cent of the factors that influenced the decision to adopt a relational data model for

ICapital data warehouse. The rest of this chapter is devoted to presenting the results for these analytic codes.

5.5.1 Assessing the impact of heterogeneous reporting and analytics requirements

In order to assess the impact of enterprise data consolidation to address different reporting and analytics requirements on the choice of logical data model for a data warehouse, the questions below were addressed to the research participants at ICapital.

- How diverse is the user base of your data warehouse?
- How often are different areas of the business on-boarding Risk and Finance data warehouse?

Table 5-3 below outlines the percentage code frequency of the analytic codes related to: *Enterprise Data Consolidation, User Group Experience, Specific Requirement, Data Conceptualization, Inflexibility of Model,* and *Simplicity of Model* analytic codes extracted from Table 5-2.

Analytic Codes	**% Code Frequency**
Enterprise Data Consolidation	94.98910675
User Group Experience	2.178649237
Specific Requirement	1.089324619
Data Conceptualization	0.871459695
Inflexibility of Model	0.435729847
Simplicity of Model	0.435729847
Total of % Code Frequency	100

Table 5-5: Percentage Code Frequency of Enterprise Data Integration vs. Other Analytic Codes

Figure 5-10 below illustrates the frequency analysis of the analytic codes (*Table 5-3*) in relation to questions to assess the impact of enterprise data consolidation on the choice of logical data model for a data warehouse.

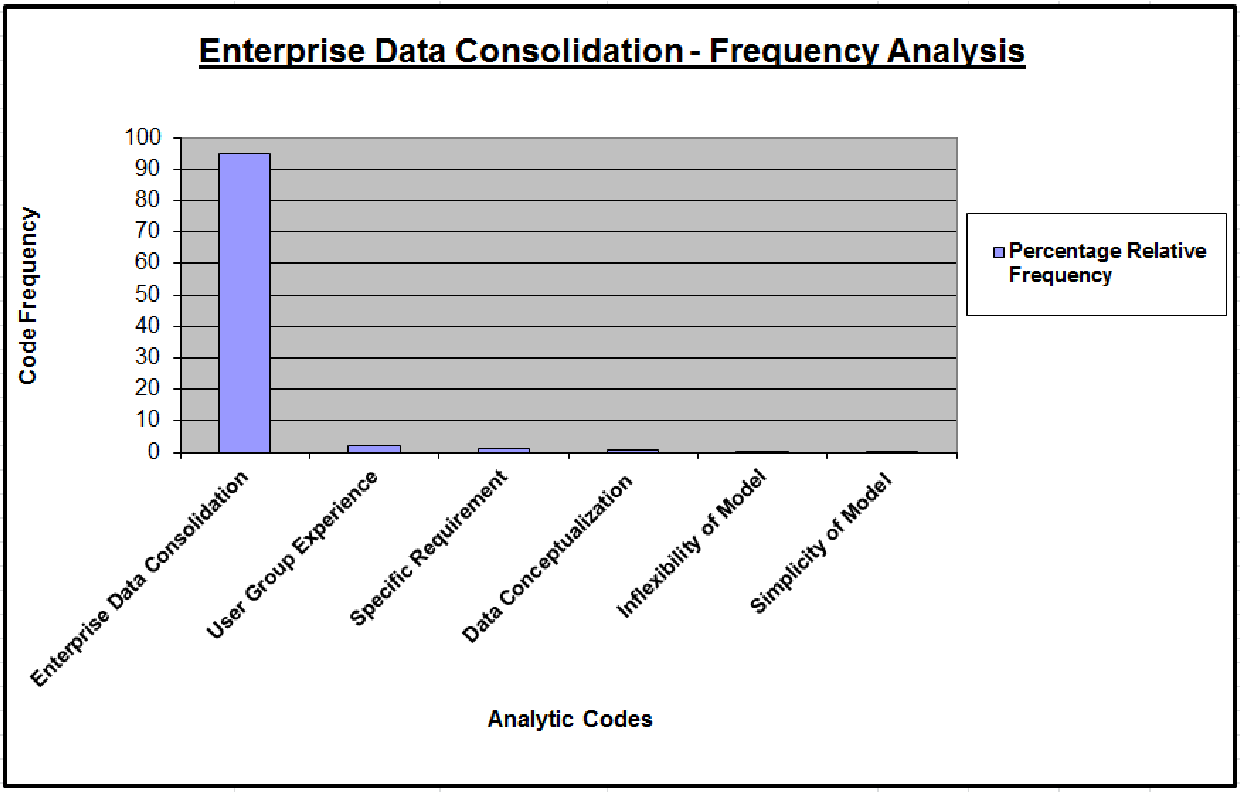

Figure 5-10: Impact of Enterprise Data Consolidation for Multiple Reporting and Analytic Needs on Logical Data Model Decision

In Figure 5-10, the responses from the research participants were coded along the x-axis, the analytical code; the percentage code frequency of each of the analytic code is on the y-axis. The research participants were asked about the diversity of consumer base of Risk and Finance data warehouse as a measure to assess the enterprise orientation of RFDW in addressing different consumer groups reporting and analytics needs. The research participants stated 94 per cent of the time that RFDW has diverse consumer base across the firm; this percentage number is considered high as a decision factor. The percentage code frequency of enterprise data consolidation as a key decision factor for implementing RFDW on a relational data model is 92 per cent higher than the percentage code frequency of all the other analytical codes combined. The percentage code frequency of enterprise data consolidation as a key decision factor is also 93 per cent more than the combined percentage code frequency of the analytical codes related to *data conceptualization* and *specific requirement* decision factors. In contrast, the combined percentage code frequency for other decision factors is quite low, collectively; they represented just 5% percent of the total decision factors for using relational data model for Risk and Finance data warehouse. The research participants at ICapital indicated the consumer base of RFDW is diverse and its data is available to anyone in the firm that requires controlled end of day dataset. The respondent stated the primary consumer of RFDW are the key functions of the firm including *credit risk, market risk, finance, collateral management, regulatory reporting* and *compliance*. For instance:

Development manager	*The data is available to anybody who wants to take the same controlled end of day data. The data is available to everyone as long as they are happy with standardised view of the data*
Lead Data Architect	*I would say it is geared towards anyone that requires signed off and approved information, the middle office and back office functions*
Project Manager	*The user base of RFDW is very diverse, we have user from other areas such as compliance, market risk and counterparty risk*
Data Architect	*As more books are on boarded more, more users will be migrated into RFDW*

The frequency analysis of the analytical code related to *enterprise data consolidation* indicated the consumer base of RFDW is diverse. This is in alignment with the research participants' statement that RFDW was created to consolidate and distribute standardised end of day dataset to the consumers groups across ICapital. The user base of the data warehouse is growing and current consumers include the groups from the middle office and back office functions using RFDW to address varieties of their operational needs. The research participants at ICapital indicated they implemented the business requirements to standardize *trades* and *positions* in the data warehouse enabling all the consumer groups to work from the same dataset across the firm. The empirical data appear to support the decision to adopt a relational data model for RFDW was influenced by the requirement from the business to create a dedicated data hub for capturing the end of day trades and positions for the firm. Consequently, the business requirement for RFDW prompted the respondents to engage a course of action based on their judgements and assessment of ICapital requirements in alignment with the prescription of the naturalistic decision making theory. The data provides evidence to support the proposition that consolidating common enterprise data to address different reporting and analytics requirements of heterogeneous consumer groups is a decision factor in adopting a relational data model for a data warehouse.

5.5.2 Assessing the impact of generic business requirements

In order to assess the impact of generic business requirements on the choice of logical data model for a data warehouse, the questions below were addressed to the research participants

- What is the impact of business requirements on the choice of Risk and Finance data warehouse logical data model?

- How many areas of the business have been on-boarded onto Risk and Finance data warehouse?

Table 5-4 below outlines the percentage code frequency of the analytic codes related to: *Generic Requirements*, *User Group Experience*, *Specific Requirement*, *Data Conceptualization*, *Inflexibility of Model,* and *Simplicity of Model* analytic codes extracted from Table 5-2.

Analytic Codes	**% Code Frequency**
Generic Requirements	94.30693069
User Group Experience	2.475247525
Specific Requirement	1.237623762
Data Conceptualization	0.99009901
Inflexibility of Model	0.495049505
Simplicity of Model	0.495049505
Total of % Code Frequency	100

Table 5-6: Percentage Code Frequency of Generic Requirements vs. Other Analytic Codes

Figure 5-11 below illustrates the frequency analysis of the analytic codes (*Table 5-4*) in relation to questions to assess the impact of generic requirements on the choice of logical data model for a data warehouse. The responses from the research participants provided evidence to support the proposition that the degree of focus on generic business requirement influences the choice of logical data model for a data warehouse. All the research participants at ICapital agreed the business commissioned RFDW to address a major requirement for the firm, to enable all the consumers across the firm to report from the same end of day *trades* and *positions* dataset from the data warehouse.

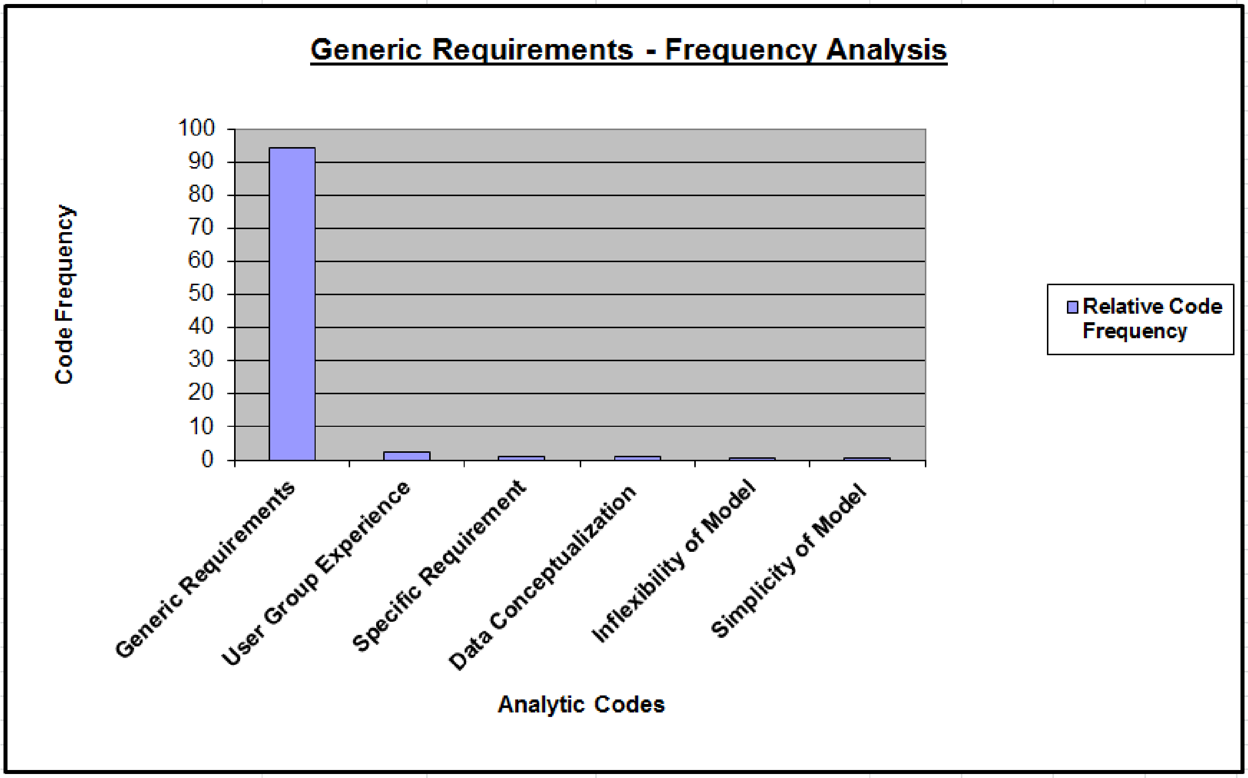

Figure 5-11: Impact of Generic Requirements on Logical Data Model Decision

In Figure 5-11, the responses from the research participants were coded along the x-axis, the analytic code; the percentage code frequency for each of the analytic codes is on the y-axis. The research participants were asked about the impact of addressing the generic business requirements on the choice of logical data model for RFDW. The research participants stated 94 per cent of the time that the data model of RFDW was adopted to address the generic business requirements. The percentage of code frequency of the analytic code to address the generic business requirements is more than 31 times the percentage code frequency of other analytic codes combined and 92 per cent more than the percentage code frequency of the analytic codes related to *data conceptualization* and *specific requirement* decision factors. The data further showed the percentage code frequency of the analytic code to address generic business requirements is more than triple the combined percentage code frequency of the analytic codes related to *model inflexibility*, *simplicity of model* and *specific requirement* decision factors. The percentage of code frequency of the analytic code to address the generic business requirements of RFDW is considered high for a decision factor; in contrast, the rest of the decision factors accounted for just 6% of the total code frequency for using a relational data model for RFDW. The research participants agreed the business requirement had a big impact influencing the choice of logical data model for RFDW. The research participants at ICapital believed the requirement from the business to create an enterprise data warehouse influenced the decision to adopt a relational data model for their warehouse. It was recognised by the respondents that data warehouses are based on business requirement, as such; RFDW was not created because the respondents thought it would be useful for the firm to create it. Instead, the decision to create RFDW was based on explicit requirements from the business and appropriately funded. The respondents believed RFDW relational data model was in alignment with the requirement from the business to create a central data hub for distributing the end of day *trades* and *positions* across the firm. For instance:

Data Architect	*The requirement of RFDW is to present data in a unified dataset to consumers, having standardised architecture for the transactions coming in from varieties of sources in varieties of structures. The objective of FRDW is to normalise source data so those downstream consumers can use data in RFDW to carry out their business*

Lead Data Architect *The business looked at the model at high level to see how requirements are captured, we showed issuer, client, trade, party, position, things the business is interested in*

Program Manager *The primary reason for RFDW is to have one controlled end of data so that everyone work off the same dataset so you have consistency across all the consumers whether it is credit risk, market risk, finance, product control, regulatory reporting or compliance*

In addressing the question related to the number of other areas of the business "*on-boarded*" onto Risk and Finance data warehouse, the research participants responded that Risk and Finance data warehouse has firm wide consumer user base. The RFDW is the data dub for publishing end of day *trades* and *positions* to the downstream consumers at ICapital. The requirement to create a data warehouse that is a data hub for distributing the end of day *trades* and *position* within ICapital was a key decision factor in adopting a relational data model for RFDW. The data from the respondents provided evidence to support the naturalistic decision making, the research participants were able to recognise the requirements from the business is best executed on a relational data model. For instance, the respondents indicated a relational data model is a good fit for the type of requirement requested by the business. This not only suggests situation awareness on the part of the respondents in alignment with the prescription of the naturalistic decision making, it also indicated mastery of the subject matter, exercising judgement on when and when not to engage the relational data model. For instance:

Program Manager *If you are taking data from lots of different sources and primarily you are a data provider, it makes sense to stick to normalise model*

The data from ICapital appear to support the proposition that the degree of focus on generic business requirement influences the choice of logical data model for a data warehouse. The frequency analysis of the analytical code to address the generic business requirements in the narrative provides justification to support the proposition that the need to focus on requirements to address generic business problems is a decision factor influencing the choice of relational data model for a data warehouse.

5.5.3 Assessing impact of goal and scope of data warehouse

In order to understand the impact of the goal and scope of a data warehouse on the choice of logical data model, the research participants were asked the questions below:

- What is the role of the goal of Risk and Finance data warehouse on the choice of its data model?
- What is the impact of the scope of Risk and Finance data warehouse on the choice of its data model?
- What is end-users' ability to understand data relationship in your data warehouse in relation to the goal of the data warehouse?

Table 5-5 below outlines the percentage code frequency of the analytic codes related to: *Defined Goal & Scope, User Group Experience, Specific Requirement, Data Conceptualization, Inflexibility of Model,* and *Simplicity of Model* analytic codes extracted from Table 5-2.

Analytic Codes	% Code Frequency
Defined Goal & Scope	89.54545455
User Group Experience	4.545454545
Specific Requirement	2.272727273
Data Conceptualization	1.818181818
Inflexibility of Model	0.909090909
Simplicity of Model	0.909090909
Total of % Code Frequency	100

Table 5-7: Percentage Code Frequency of Defined Goal & Scope vs. Other Analytic Codes

Figure 5-12 below illustrates the frequency analysis of the analytic codes (*Table 5-5*) in relation to questions to assess the impact of defined goal and scope on the choice of logical data model for a data warehouse.

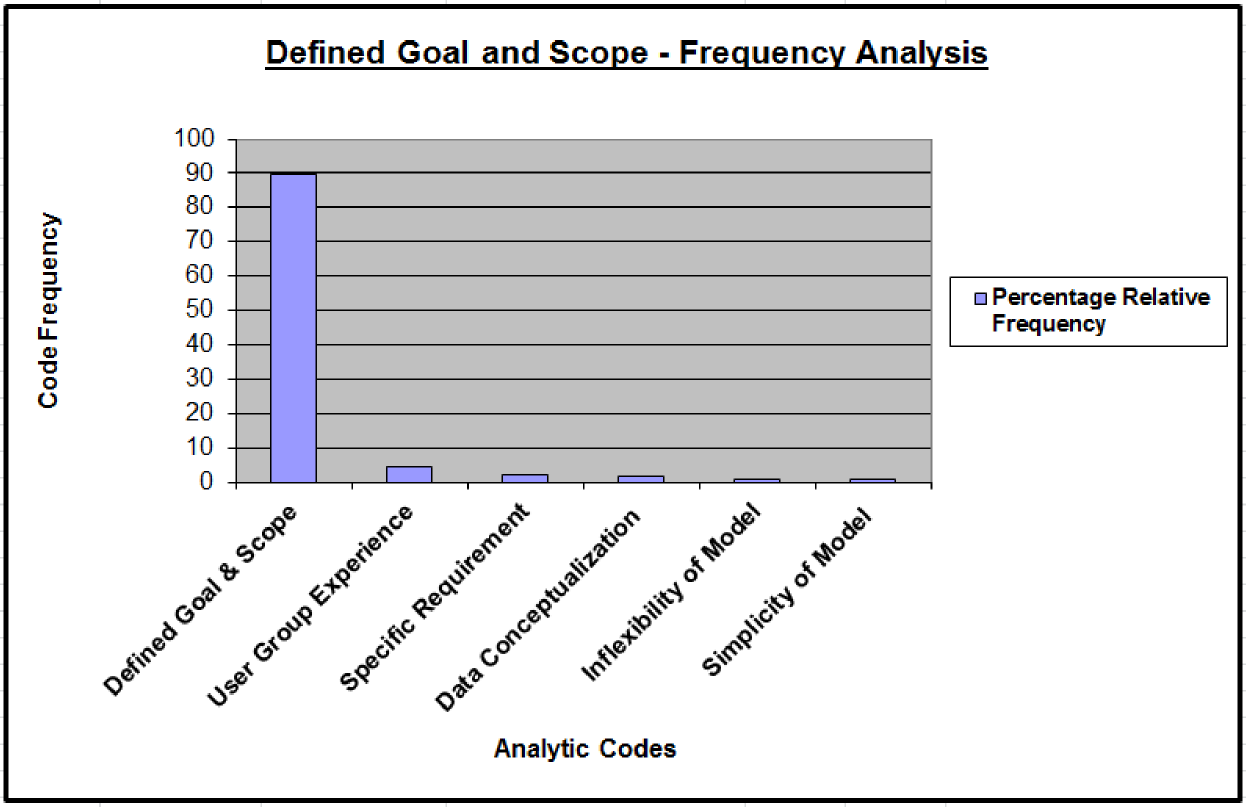

Figure 5-12: Impact of Defined Goal and Scope on Logical Model Decision

In Figure 5-12, the responses from the research participants were coded along the x-axis, the analytic code; the percentage code frequency of each of the analytic code is on the y-axis. The research participants were asked about the goal of Risk and Finance data warehouse as a measure to assess the impact of defined goal and scope of a data warehouse in a decision to adopt a logical data model for RFDW. The research participants stated 90 per cent of the time that the goal and scope of Risk and Finance data warehouse was a factor in adopting a relational data model for RFDW, this number is considered high for a decision factor. This indicated that almost all the research respondents believed the goal and scope of the business is a decision factor in adopting a relational data model for Risk and Finance data warehouse.

As illustrated in Figure 5-12, the percentage rate of code frequency related to defined goal and scope is more than double the percentage code frequency of *user group experience* and *specific requirement* decision factors. The data indicated the percentage code frequency of the analytic code related to business defined goal and scope is 88 per cent higher than the percentage code frequency of the analytic code related to *data conceptualization* decision factor. In contrast, the total code frequency for the rest of the decision factors is quite low, the combined code frequency for the rest of the decision factors accounted for 10% of the decision to use a relational data model for Risk and Finance data warehouse. The data collected at ICapital indicated the respondents considered the business stated goal and scope of Risk and Finance data warehouse in selecting its data model. The data also indicated the respondents were aware of selecting an appropriate data model for RFDW. This suggested the respondents were cognizant of the risks associated with selecting a data model that is not in alignment with the business-defined goal for RFDW. For example:

Data Architect	*The goal of the business has a lot to do with it, one of the success criteria for RFDW is to what extent it satisfies the business goal, the project will be deemed failure if business goal is not met*
Lead Data Architect	*The business looked at the model, we define the model in a manner that showed all the information the business is after, we showed the clients, trades, party and issuer*
Project Manager	*The goal set by the business has impact on the direction of RFDW in general including its choice of data model. The goal is really the building blocks of RFDW and represents the items that RFDW must meet to be declared a success*

Development Manager *I don't think you can separate the goal of the business from the choice of RFDW data model, they all linked together*

The data from ICapital indicated RFDW was intended as the data hub for capturing and distributing the reconciled end of day business data to the heterogeneous business groups within the firm including credit risk, finance, market risk and compliance functions. To this extent, adopting a relational data model is judged to be in alignment with the business goal for RFDW. For instance:

Program Manager *The goal of the business is to have all the business areas use and report from the same set of data. If you are finance and you want to calculate your P&L, the basic transaction data into your calculation comes from RFDW, the same goes for credit risk and market risk*

The data from ICapital indicated the respondents considered the scope of RFDW as a decision factor in adopting a relational data model for the data warehouse. In particular, the data showed the respondents believed a data warehouse that caters for the need of large different consumers within an enterprise tend to have a complex data model:

Project Manager *As you will expect, the data architecture for this type of firm wide data warehouse is complex. I think people will understand why I don't think it will be possible to say you are the golden source of end of day data to hundreds of consumers across the firm and people expects your data model to be simple, I don't think that will be possible*

Lead Data Architect	*I'm actually a believer in the fact that a data warehouse is the wholesale, it is not the detail. I will also say RFDW is the wholesale system. If you are creating a wholesale system, which supports all the information of the firm in the case of RFDW, you probably want a model that is not too dimensional*
Developer	*The aim of RFDW is to support many areas of the business so having a data model that enables us to capture any form of trade data is important to us*

The respondents at ICapital acknowledged the business defined goal and scope of Risk and Finance data warehouse was a decision factor in adopting a relational data model for the data warehouse. The data indicated the respondents implicitly engaged in situation awareness and consequently modelled categorisation based on recognising the goal and scope that relational data model is suited. The frequency analysis of the analytic code to address defined goal and scope in the narrative provides the justification to support the proposition that the business defined goal and scope is a decision factor influencing the choice of relational data model for a data warehouse.

5.5.4 Assessing the impact of implementation orientation of available resources

In order to assess the impact of the implementation orientation of available resources on the choice of logical data model for a data warehouse, the questions below were addressed to ICapital research participants.

- What is the impact of staff experience in the decision to adopt a relational data model for Risk and Finance data warehouse?
- Would you agree or disagree that your staff experience influenced the choice of the data model for your data warehouse?

Table 5-6 below outlines the percentage code frequency of the analytic codes related to: *Staff Experience*, *User Group Experience, Specific Requirement*, *Data Conceptualization*, *Inflexibility of Model*, and *Simplicity of Model* analytic codes extracted from Table 5-2.

Analytic Codes	**% Code Frequency**
Staff Experience	87.83068783
User Group Experience	5.291005291
Specific Requirement	2.645502646
Data Conceptualization	2.116402116
Inflexibility of Model	1.058201058
Simplicity of Model	1.058201058
Total of % Code Frequency	100

Table 5-8: Percentage Code Frequency of Staff Experience vs. Other Analytic Codes

Figure 5-13 below illustrates the frequency analysis of the analytic codes (*Table 5-6*) in relation to questions to assess the impact of staff experience on the choice of logical data model for a data warehouse. The implementation orientation of available resources is defined by the experience of the employees tasked with building a data warehouse.

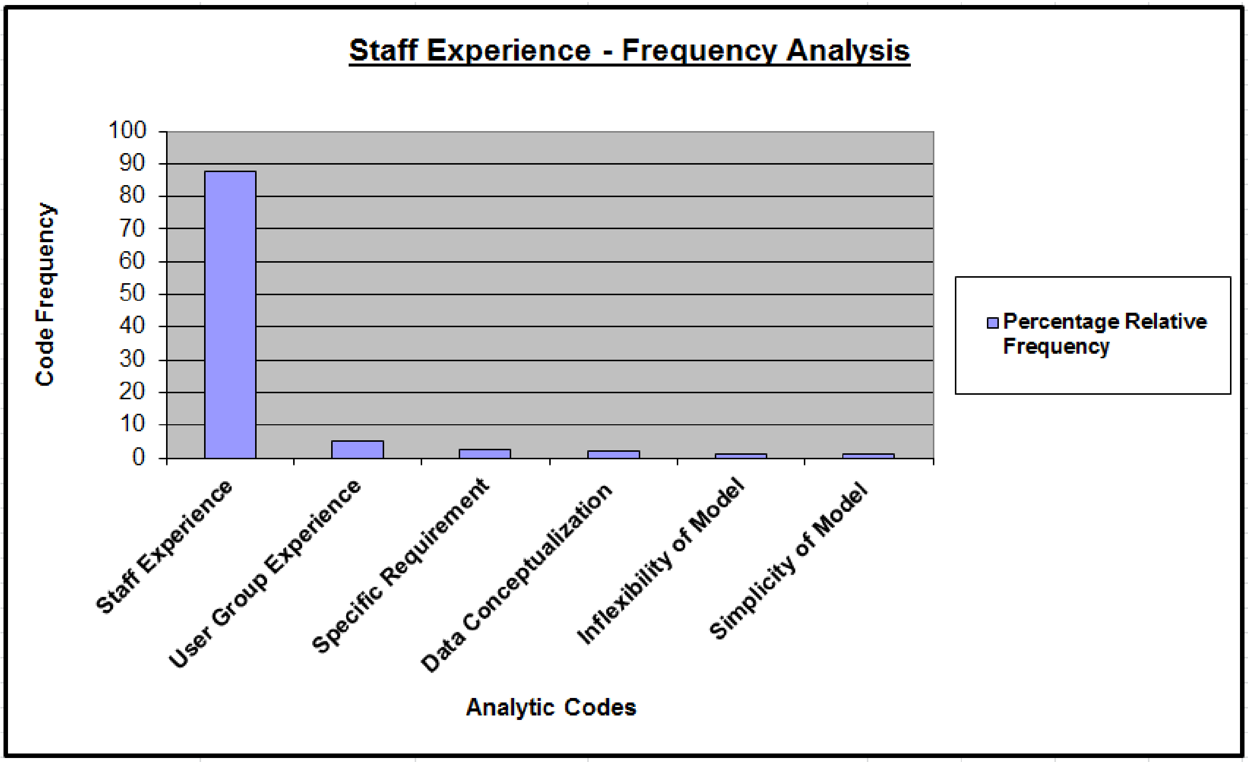

Figure 5-13: Impact of Staff Experience on Logical Data Model Decision

In Figure 5-13, the responses from the research participants were coded along the x-axis, the analytic code; the percentage code frequency of each of the analytical code is on the y-axis. The research participants were asked about the impact of staff experience as a measure to assess the impact of implementation orientation of available resources in adopting a logical data model for a data warehouse. The research participants at ICapital stated 87 per cent of the time that their staff experience was a decision factor in adopting a relational data model for RFDW.

The percentage code frequency of staff experience as a decision factor for adopting a relational data model for RFDW is more than 15 times the percentage code frequency of all the other analytic codes combined, this number is considered high for a decision factor. In contrast, the code frequency for rest of the decision factors accounted for 12% which is considered low for the decision factors. The research participants at ICapital expressed their preference for a relational data model as the most sensible option to address the business requirements and goal of RFDW. For instance:

Program Manager	*If you are taking data from lots of different sources and primarily you are a data provider, it makes sense to stick to normalise model*
Lead Data Architect	*I would say the dimensional model is perfect for just pure mart, a dimensional model for RFDW wouldn't have worked*
Project Manager	*You want to be able to able to support as much as possible different requirements so that those consumers in turn can support operations in their respective areas*
Data Architect	*If you look at RFDW you will see it is evident from the way the entities are laid out, you can bring in more data later on with considerable less difficulty*

The research participants at ICapital agreed that their staff experience played a key role in adopting a relational model for their data warehouse. The development team at ICapital had an overwhelming preference for a relational data model and acknowledged the model was the appropriate choice to address the business requirement for RFDW. The data from ICapital provides justification to support the prescriptions of naturalistic decision making.

The business requirement for RFDW prompted the respondents to engage a course of action based on their judgements and assessment of ICapital requirements in alignment with the naturalistic decision making theory. The empirical data provides the justification to support the proposition that implementation orientation of available resources impacts the choice of logical data model for a data warehouse.

5.5.5 Assessing the impact of high performance of query execution

In order to understand the impact of high performance of query execution on the choice of logical data model for a data warehouse, the questions below were addressed to ICapital research participants.

- What non-functional requirements impacted the choice of Risk and Finance data warehouse logical data model?
- What other non-functional factors influenced the decision to adopt a relational data model for Risk and Finance data warehouse?
- How would you characterise the advantages of Risk and Finance data warehouse data model?

Table 5-7 below outlines the percentage code frequency of the analytic codes related to: *High Query Performance, User Group Experience, Specific Requirement, Data Conceptualization, Inflexibility of Model,* and *Simplicity of Model* analytic codes extracted from Table 5-2.

Analytic Codes	**% Code Frequency**
High Query Performance	79.09090909
User Group Experience	9.090909091
Specific Requirement	4.545454545
Data Conceptualization	3.636363636
Inflexibility of Model	1.818181818
Simplicity of Model	1.818181818
Total of % Code Frequency	100

Table 5-9: Percentage Code Frequency of Query Performance vs. Other Analytic Codes

Figure 5-14 below illustrates the frequency analysis of the analytic codes (*Table 5-7*) in relation to questions to assess the impact of high performance of query execution on the choice of logical data model for a data warehouse.

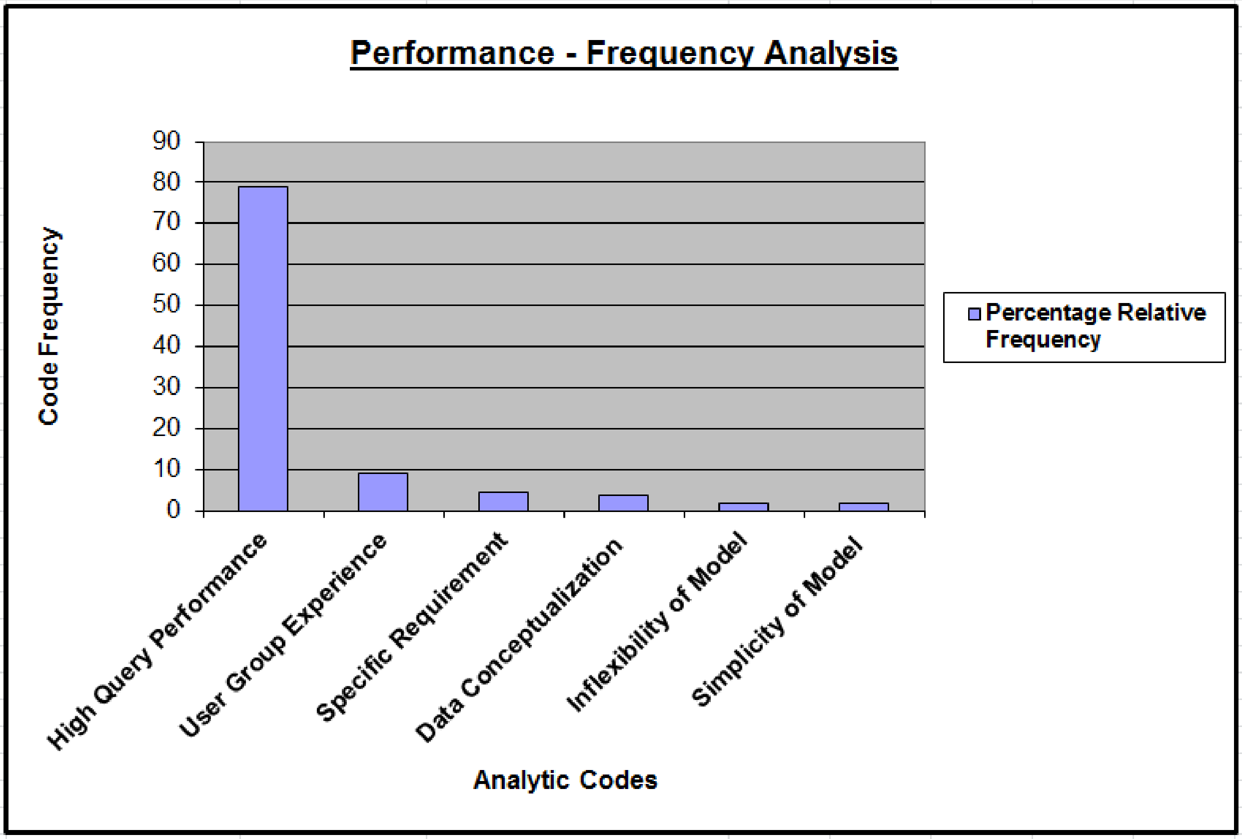

Figure 5-14: Impact of Performance in Data Warehouse Logical Data Model Decision

In Figure 5-14, the responses from the research participants were coded along the x-axis, the analytic code; the frequency analysis of the analytic code is on the y-axis. The research participants were asked about the impact of performance as a measure to assess the performance of query execution of RFDW. The research participants stated 79 per cent of the time that performance was a key consideration in using a relational data model for ICapital Risk and Finance data warehouse, this is considered high for a decision factor. The percentage code frequency of performance as a key decision factor for implementing the RFDW on a relational data model is more than 7 times the percentage code frequency of all the other analytic codes combined. In contrast, the percentage code frequency for the rest of the decision factors is 21%, this is considered lower for the decision factors. The responses from the research participants indicated performance was the most important consideration of any non-functional requirement for RFDW. One of the non-functional requirements of Risk and Finance data warehouse that featured prominently in the participant responses was the need to quickly process large volumes of data that comes into the Risk and Finance data warehouse on a daily basis. All the research participants stated the importance of performance to the business, downstream consumers and the expectation of getting high performance out of Risk and Finance data warehouse. One key element of ensuring RFDW performance is in line with the business expectation is the use of database appliance for the data warehouse. The participants indicated RFDW database appliance is designed and optimized for processing large volumes of data that the data warehouse receives daily. One noticeable observation of the research participants at ICapital is how they strongly believed in their choice of relational data model for Risk and Finance data warehouse. In order to shed light on other non-functional factors, which influenced the choice of RFDW logical data model, the participants reaffirmed that the business requirements was a factor in adopting a relational data model for Risk and Finance data warehouse. The research participants outlined a number of advantages of RFDW data model. One of the advantages of RFDW data model outlined by the research participants is flexibility. For instance:

Project Manager	*The data model has enabled RFDW to capture vast array of data in one place what you will normally find in different systems*
Lead Data Architect	*The attraction of the model is flexibility in the sense that the model resolves the hidden mismatch between the way information arrives and the way downstream users is interested in using that information*

Development Manager	*The model is incredibly flexible, we can integrate new types into the model as required*

The empirical data from ICapital suggests an alignment with the prescription of naturalistic decision theory. The data indicates the research participants at ICapital leveraged their experience of situation awareness, drawing on it to address the business requirement to build Risk and Finance data warehouse. The respondents at ICapital were able to articulate the benefits of RFDW data model; they stated the data model allowed them to address the business requirement of making RFDW the data provider of end of the day *trades* and *positions* across the firm. The participants stated their team had to implement a data model that allowed them to capture as many diverse dataset as possible and process the data as quickly as possible with minimal data transformations during data loading into RFDW. The research participants acknowledged that a relational model is not necessarily considered the optimal data model for performance, however, the respondents stated they improved performance by minimising data transformation in RFDW and use a database appliance for data processing for their data warehouse. In alignment with naturalistic decision making, this is an indication the respondents were using their experience of situation awareness to make judgement and engage an appropriate course of action to address the issues related to the performance of RFDW. The frequency occurrence of the analytic code for performance in the narrative provided added justification to support the proposition that high performance of query execution is a decision factor that influenced the choice of relational data model for the ICapital data warehouse.

5.6 Summary of the Findings

This chapter presented the findings of the case study conducted at ICapital; the chapter presented the result of the factors that influenced the decision to adopt a relational data model for ICapital Risk and Finance data warehouse. The findings of the study are summarised below.

5.6.1 Finding 1: *Enterprise Data Consolidation is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse*

The study finds ICapital Risk and Finance data warehouse is an enterprise data warehouse and distributor of standardised dataset to the other consumer groups within ICapital. The consumer base of Risk and Finance data warehouse is diverse because RFDW is a data provider of end of the day *trades* and *positions* dataset to varieties of consumers within ICapital. In the literature, integrating enterprise data for various reporting and analytics requirements requires flexible data model that is oriented to the enterprise view of an organisation and focuses on complex data structures and interactions (Imhoff, Galemmo and Geiger, 2003; Tryfona, Busborg and Christiansen, 1999). The data from the research participants showed that the decision to adopt a relational data model for Risk and Finance data warehouse was influenced by the requirement for a data warehouse that cater for the generic needs of different consumer groups within ICapital. The observation from the empirical data showed high proportion of the research participants indicated the decision to adopt a relational data model for Risk and Finance data warehouse was influenced by the requirement to create a dedicated data hub for capturing the end of the day *trades* and *positions* for the organisation. The research participants stated 95 per cent of the time that RFDW has diverse consumer base across the firm.

The percentage code frequency of enterprise data consolidation as a key decision factor for implementing RFDW on a relational data model is 89.9 per cent higher than the percentage code frequency of the analytic codes for *user group experience*, *specific requirement, data conceptualization inflexibility of the model* and *simplicity of model*. The observation from the empirical data indicated that lower proportion of the research participants indicated the decision factors relating to *user group experience, specific requirement, data conceptualization inflexibility of the model* and *simplicity of model* were responsible for adopting a relational data model for Risk and Finance data warehouse. The combined percentage code frequency for these decision factors represented 5% percent of the total decision factors for using relational data model for Risk and Finance data warehouse. The empirical data provided evidence to support the proposition that consolidating common enterprise data to address different reporting and analytics requirements is a decision factor in adopting a relational data model for a data warehouse.

5.6.2 Finding 2: *Generic Requirements is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse*

One of the important factors that influenced the decision to implement the ICapital data warehouse on a relational data warehouse was the requirement to build an enterprise data warehouse that is a data provider of end of the day *trades* and *positions* for varieties of consumers within ICapital. The empirical data indicated the business requirement had a big impact in influencing the choice of logical data model for RFDW.

The respondents at ICapital recognised that data warehouses are requirements based; consequently, the requirement to create RFDW was based on the explicit requirements from the business and appropriately funded. The respondents at ICapital agreed the business commissioned Risk and Finance data warehouse to address a major requirement for the firm, to enable all consumers across the firm to report from the same end of the day business data from RFDW. In the literature a relational data model is used for a data warehouse designed to support an enterprise enabling the possibility of different types of reporting and analytics across within an organisation (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003). The observation from the empirical data showed high proportion of the research participants indicated the decision to adopt a relational data model for Risk and Finance data warehouse was influenced by the requirement to address generic business requirements for a host of business areas and regions for the organisation. The research participants stated 94 per cent of the time that the data model of RFDW was adopted to address the generic business requirements. The percentage of code frequency of the analytic code to address the generic business requirements is more than 31 times the percentage code frequency of the analytic codes for *user group experience*, *specific requirement*, *data conceptualization*, *inflexibility of model,* and *simplicity of model*. Also, the percentage of code frequency of the analytic code to address generic business requirements is 92 per cent more than the percentage code frequency of the analytic codes related to *data conceptualization* and *specific requirement* decision factors. Additionally, the data showed that percentage code frequency of the analytic code to address generic business requirements is more than triple the combined percentage code frequency of the analytic codes related to *model inflexibility*, *simplicity of model* and *specific requirement* decision factors.

5.6.3 Finding 3: *Goal and Scope is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse*

The goal and scope of a data warehouse is an important factor that impacts the choice of logical data model for a data warehouse. The research participants at ICapital recognised the risk involved in choosing a data model for a data warehouse without due consideration for the goal of the data warehouse. The empirical data indicated the business defined goal and scope of RFDW was in alignment with the logical data model adopted for the ICapital data warehouse. The respondents strongly believed the data model for RFDW was selected in alignment with the goal and scope defined by their organisation. The study finds the goal of ICapital Risk and Finance data warehouse was clearly defined and the scope of the project was large. The empirical data indicated the business defined goal and the scope of Risk and Finance data warehouse were in alignment with the logical data model adopted for the ICapital data warehouse, this is consistent with the literature. In the literature, the goal of a data warehouse ultimately defines the scope of a data warehouse. The scope of a data warehouse where the goal of the data warehouse is oriented toward an enterprise is larger than the scope of a data warehouse where its goal is oriented to the need of a particular business area within an enterprise (Imhoff, Galemmo and Geiger, 2003). The observation from the empirical data showed that high proportion of the research participants indicated the decision to adopt a relational data model for RFDW was influenced by the consideration of the goal and scope of Risk and Finance data. In particular, the stated goal and the scope of Risk and Finance data warehouse was one of the drivers of the decision that led to adopting a relational data model for Risk and Finance data warehouse. The research participants stated 90 per cent of the time that the goal and scope of Risk and Finance data warehouse was a factor in adopting a relational data model for RFDW.

The percentage code frequency related to defined goal and scope is more than double the percentage code frequency of *user group experience* and *specific requirement* decision factors. Additionally, the data indicated the percentage code frequency of the analytic code related to business defined goal and scope is 88 per cent higher than the percentage code frequency of the analytic code related to *data conceptualization* decision factor. In contrast, the observation from the empirical data indicated lower proportion of the research participants indicated the decision factors relating to *user group experience, specific requirement, data conceptualization inflexibility of the model* and *simplicity of the model* were responsible for adopting a relational data model for Risk and Finance data warehouse, collectively, these decision factors accounted for 10% of the decision to use a relational data model for Risk and Finance data warehouse.

5.6.4 Finding 4: *Staff Experience is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse*

The implementation orientation of the staff tasked with building a data warehouse is an important factor that impacts the choice of logical data model for a data warehouse. The empirical data from the study indicated that ICapital staff leveraged their prior experiences and drawn on it to address the requirements to build a data warehouse for their organisation. Vassiliadis (2004) notes that data warehousing landscape is defined by "*do-it-yourself*" advice from experts and proprietary vendor solutions, the empirical data indicated that the research participants at ICapital have considerable experience in relational data model and recognised situations where using any other type of data model apart from relational data model would not have worked to address the requirements from their organisation.

The observation from the empirical data showed that high proportion of the research participants indicated the decision to adopt a relational data model for Risk and Finance data warehouse was influenced by the experience of their implementation teams. The research participants at ICapital stated 87 per cent of the time that their staff experience was a decision factor in adopting a relational data model for RFDW. The percentage code frequency of staff experience as a decision factor for adopting a relational data model for RFDW is more than 15 times the percentage code frequency of the analytic codes for *user group experience*, *specific requirement, data conceptualization inflexibility of the model* and *simplicity of the model*. In contrast, these decision factors accounted for 12% of the decision to use a relational data model for Risk and Finance data warehouse. The empirical data provided the justification to support the proposition that the implementation orientation of available resources impacts the choice of logical data model for a data warehouse.

5.6.5 Finding 5: *High Query Performance is a Decision Factor in Adopting Relational Data Model for ICapital Data Warehouse*

The study finds the research participants indicated performance is the most important non-functional requirement that influenced the choice of Risk and Finance data warehouse data model. One of the non-functional requirements of Risk and Finance data warehouse that featured prominently in the participant responses was the need to quickly process large volumes of data that comes into Risk and Finance data warehouse on a daily basis. The research participants stated the importance of performance to the business, downstream consumers and the expectation of getting high performance out of Risk and Finance data warehouse.

The study finds the research participants engaged an appliance database to ensure the performance of RFDW is in line with the business expectation. The participants indicated RFDW database appliance is designed and optimized for processing large volumes of data that the data warehouse receives daily. This finding is in contrast with the literature, in the literature, the most important performance consideration in data warehousing concerns the optimal selection of database objects such as indexes, which is based on the logical data model adopted for a data warehouse (Golfarelli and Rizzi, 1998). The empirical data showed that high proportion of the research participants indicated that performance was a key consideration in using a relational data model for ICapital Risk and Finance data warehouse. The empirical data indicated the importance of performance to the business, downstream consumers and the expectation of getting high performance out of Risk and Finance data warehouse. The research participants stated 79 per cent of the time that performance was a key consideration in using a relational data model for ICapital Risk and Finance data warehouse. The percentage code frequency of performance as a key decision factor for implementing the RFDW on a relational data model is more than 7 times the percentage code frequency of the analytic codes for *user group experience*, *specific requirement*, *data conceptualization inflexibility of the model* and *simplicity of the model*. In contrast, these decision factors accounted for 21% of the decision to use a relational data model for Risk and Finance data warehouse.

CHAPTER 6 - Discussion of the Case Studies

6.1 Introduction

This chapter presents a detailed discussion of the results presented in Chapters 4 and 5 for the case studies conducted at GWealth and ICapital, the wealth management and investment banking division of a global financial institution, TBank. In presenting the discussion of the research findings, the chapter analyses the observations from the empirical data in the light of the research propositions. In addition, the chapter offers a critique of the findings from the case studies and presents a revised conceptual model for logical data model adoption.

The chapter is divided into six Sections; Section 6.2 discusses the research findings in detail and provides an in-depth discussion of each of the factors that influenced the decision to adopt multidimensional and relational data models in the case organisations. Based on the findings of the research, Section 6.3 examines the relationships between the research analytic code variables and the logical data models. Section 6.4 presents a revised conceptual decision model for logical data model adoption and Section 6.5 discusses the implications of the study for practice and theory respectively. Lastly, Section 6.6 concludes with a summary.

6.2 Discussion of the Research Findings

In this Section, the research propositions are discussed based on the empirical findings from the GWealth and ICapital case studies. Table 6-1 presents a comparison of the findings at GWealth and ICapital. In Table 6-1, the percentage of code frequency that is 60% and above is rated *High*, those between 50% - 59% are rated *Medium*, those between 0% - 49% are rated *Low*. The data points between the ratings represent areas where there are gaps in the frequency analysis; this presents a cut off for the rating of a category. In the following Sections, each of the findings is discussed in more detail.

Analytic Category	Property	GWealth			ICapital		
		% Code Frequency		Decision Factor Rating	% Code Frequency		Decision Factor Rating
		High	Low		High	Low	
High Performance of Query Execution	High Performance	80	20	High	79	21	High
Business Requirement	Specific Requirement	75	-	High	-	6	Low
	Generic Requirement	-	25	Low	94	-	High
Firm Objective	Defined Goal & Scope	68	32	High	90	10	High
Employee Experience	Staff Experience	60	40	High	88	12	High
Enterprise Data Hub	Data Consolidation	(98)	2	Low	93	7	High
	Diverse Consumers	(53)	47	Low	95	5	High

Table 6-1: Summary of the Findings at GWealth and ICapital Case Studies

6.2.1 High Performance of Query Execution

Improving faster data access and query performance of data warehouses are important considerations in data warehousing (Kimball, 1995, 1997; Weininger, 2002). In the literature, multidimensional data modelling is seen as a technique that presents data in a standard intuitive framework that allows for high performance as queries in decision support systems requires significant data joining and aggregations (Kimball, 1995; 1977). High performance of query execution was rated high as a decision factor for the GWealth and ICapital cases, the research participants at GWealth stated 80% of the time that performance was a key factor in using a multidimensional for their reporting data warehouse. Equally, the research participants at ICapital stated 79% of the time that performance was a factor in their consideration for using a relational data model for their data warehouse. In the case of GWealth, the research participants used a multidimensional data model as the primary tool kit that enabled them to achieve high performance. This is consistent with the literature; a multidimensional data model is seen as better for performance. In the literature, the attractiveness of using a multidimensional schema for a data warehouse is that it improves performance (Kimball and Ross, 2002; Kimball, 2005; Kimball 1997). The key decision-makers at GWealth believed a multidimensional data model enabled them to achieve high performance of query execution for their data warehouse. One of the reasons why the research participants at GWealth adopted a multidimensional data model for their data warehouse was because of the need to quickly process reports of statements for their client:

Development Manager	*Getting the reports out to clients is very important to us so we used the data model that enabled us to get data out very fast so that we can report on it*

In alignment with the literature, one of the advantages of using a multidimensional data model is that multidimensional data warehouses are easier to understand (Kimball and Ross, 2002; Herden, 2000). This view is supported by the empirical data; the research participants at GWealth outlined some advantages of the data model for reporting data warehouse. One of the advantages outlined by the research participants was that the reporting warehouse data model is easier to understand:

Project Manager	*I would say for me, it is easier to understand and easier to isolate issues and of course, it allows you to focus on what you initially set out to do which is to implement a reporting warehouse*
Data Architect	*the main one for me is around the query performance and the end-user being able to understand exactly what you build*

High performance of query execution as a decision factor is also rated high for the research participants at ICapital. This outcome is consistent with the research proposition RP1 – high performance of query execution is a factor influencing the choice of data model for a data warehouse. Although the observed outcome is consistent with the research proposition RP1, it is necessary to question the extent to which a multidimensional data model is solely responsible for the performance gain claimed by the research participants at GWealth.

First, it is not clear from the empirical data whether GWealth operates a "*Shared Services*" operating model. Shared services enables organisation to leverage and consolidate business operations that are used by multiple functions within an organisation, the aim of which to eliminate redundancies and reduce the operating cost (Rouse, 2013). As GWealth move towards a shared services operating model, it is likely the performance of reporting data warehouse will degrade overtime as the reporting team implements shared services agreement with other technology functions within GWealth. As more services and applications are placed on the reporting data warehouse server as part of the shared services operating model, the faster the performance degradation the reporting team will experience overtime. In this situation, a question can be asked of the reporting data warehouse team whether they still believe their multidimensional data model is responsible for the performance gain of the reporting data warehouse or it is the "*un-shared, resource rich*" infrastructure environment of the reporting data warehouse that provided the performance gain that was wrongly attributed to a multidimensional data model. The answer considered likely would be the latter.

Second, even if the management of GWealth has no intention of operating shared services model in the future, it is likely that as the management attempt to justify Return on Investment (ROI) on their existing infrastructure expenditure, they will request that future applications and services leverage the existing infrastructure environment, over a period of time, this will impact the performance of all the applications and services on the infrastructure. This brings us back to the question that was raised earlier whether it is truly a multidimensional data model that improves performance or it is the infrastructure environment that ultimately improves the performance of a data warehouse.

As indicated previously, the ICapital team adopted a relational data model for their data warehouse; the research participants at ICapital stated 79% of the times that performance was a key consideration in their consideration of using relational data model for Risk and Finance data warehouse. In addition, the ICapital team used a database appliance to improve the performance of their Risk and Finance data warehouse. The use of an appliance database by ICapital team raises some observations that should be explored. First, it can be argued this is an indication that a relational data model is not an optimal choice for improving the performance of a data warehouse unless a data warehouse project is well funded to the extent that there is a budget to purchase an appliance database that can improve the performance of a data warehouse. Second, knowing that a relational data model is not considered an optimal for performance and given that a bus architecture could be used if high performance is imperative for RFDW, the question can be asked as to whether the ICapital team made the right decision in adopting a relational data model for the Risk and Finance data warehouse (*considering the research participants indicated 79% of the times that performance was a key consideration for their data warehouse*). However, it may be very difficult or almost impossible to use a bus architecture framework for RFDW if the business objectives of RFDW are taken into consideration. In this situation, the empirical data appear to suggest that the research participants at ICapital knew precisely why they adopted a relational data model for their data warehouse. For instance:

Director	*I would say the dimensional model is perfect for just pure mart, a dimensional model for RFDW wouldn't have worked*

First, it can be interpreted from the above statement that the implementation team at ICapital is aware of the of logical data model that is suitable for their data warehouse, it can also be inferred from the statement that the implementation team is aware of the potential dangers of using an inappropriate logical model for their data warehouse. On the surface, it appears that the director's comment does not go beyond the assertion that a multidimensional data model is "*perfect for just pure mart*", however, the observation from the empirical data suggests that there are pressing requirements that the management at ICapital required their data warehouse to address. For instance, the overriding requirement for ICapital data warehouse is the need for the RFDW to be a data hub for their organisation. In this respect, it can be argued that the ICapital team viewed a database appliance as an integral part of the strategy to address the performance limitations of the data model they adopted for their data warehouse.

Second, from an operations perspective, the RFDW was commissioned as a global data warehouse and funded accordingly. It can be interpreted that funding was not an impediment for the ICapital team in developing the type of data warehouse their management required. As RFDW is processing data feeds from key regional locations where ICapital operates its businesses, the empirical data suggests that, longer term, a relational data model provides the capabilities that ICapital requires as it integrates additional line of businesses into its data warehouse. In this respect, the cost of purchasing a database appliance could be interpreted as being "*passed on*" as part of the cost of doing business. As such, funding for this type of expensive hardware presents no immediate obstacle for the ICapital team.

6.2.2 Focus on Specific and Generic Business Requirements

In the literature, the data architecture of a data warehouse is driven by user requirements; requirements' engineering is a reality in demand-driven and supply-driven design frameworks (Giorgini, Rizzi and Garzetti, 2008). In data warehousing, for demand and supply driven design frameworks, the degree of focus on specific or generic business requirements is a decision factor that impacts the choice of logical data model for a data warehouse. In the literature, the user requirements outlined the vision and the tasks of a data warehouse (Inmon, 2005; Inmon, Imhoff and Sousa, 2001; Kimball and Ross, 2002; Sperley, 1999; Reeves, Ross, Thornthwaite and Kimball, 1998). The empirical data from the study is consistent with the literature on the requirement driver for implementing multidimensional and relational data warehouses. A multidimensional data model is used to implement a data warehouse that supports the functionality of a specific part of the business (Imhoff, Galemmo and Geiger, 2003; Kimball and Ross, 2002). A data warehouse built on a relational data model is engineered to support the business rules of an entire enterprise; therefore, a relational data model should not be used for a data warehouse that is oriented to the need of a specific functional area in an organisation. Instead, a relational data model should be used for a data warehouse that supports an enterprise enabling the possibility of different types of reporting and analytics across within an organisation (Inmon, 2005; Imhoff, Galemmo and Geiger, 2003). The outcomes from GWealth and ICapital cases are consistent with the literature and the research proposition RP2 – degree of focus on specific requirement or generic requirements is a factor influencing the choice of data model for a data warehouse

Focusing on the requirements to address the business needs is rated high as a decision factor by the research participants at GWealth and ICapital cases. For the research participants at GWealth, focusing on the specific business requirement to create a client reporting data warehouse was rated high as a decision factor. The research participants at GWealth stated 81 per cent of the time that the data model for their reporting data warehouse was adopted to address a specific requirement of their business for client reporting, consequently, this category property was rated high by the research participants. The research participants at GWealth equally rated the simplicity of their data model high. In contrast, research participants at GWealth rated low enterprise properties such as *generic requirements* for their data warehouse. As a result, the implementation team at GWealth did not find a relational data model suitable for their reporting data warehouse. For instance:

Data Architect	*The requirement was to develop a data warehouse for creating reporting statements for clients of the firm. I found multidimensional model to be perfect for that type of requirement*

As described in Section 4.3.2, the research participants at GWealth considered a multidimensional data model to be particularly suitable for implementing a business requirement that is not perceived as complex or a data warehouse that is not intended to support heterogeneous consumers with diverse needs and requirements. For instance:

Data Architect	*If the scope is very small as in the case of the reporting warehouse, a multidimensional data architecture lend itself to a much better data warehouse*

Although the research participants at GWealth stated 84 per cent of the time that a multidimensional data model was adopted for the reporting data warehouse to address a specific requirement of client reporting for their organisation, it is necessary to point out that the GWealth implementation team did not specifically considered the future requirements of their business users while building the reporting data warehouse. While it is true the implementation team at GWealth developed the reporting data warehouse to address current business requirement to address the challenges faced by their organisation today, there is no guarantee the reporting data warehouse will meet the reporting needs of the organisation in future. As GWealth operates in a dynamic and changing business environment; it is unlikely the organisation will be able to fully leverage today’s solution to address tomorrow’s needs without significant modifications to the existing data warehouse. In this situation, although a bus architecture framework is an option available to the GWealth team to address potential future requirements; this option is only useful if the granularity of data for future enhancements is the same as the granularity of the existing data in the reporting data warehouse. Additionally, it is likely that such a retrofitting exercise may prove too costly to the extent that the management decide to abandon the retrofitting effort in favour of building a new data warehouse. The implementation team at GWealth could avoid this type of problematic situation by been proactive and thinking strategically about the potential future needs of their clients in addition to addressing the existing requirements of their business for client reporting data warehouse.

Focusing on the requirements to address the generic requirements of their organisation was equally rated high by the research participants at ICapital. For the participants at ICapital, the overriding requirements driving the need for a data warehouse was the requirement to capture, consolidate and distributes the reconciled end of day (EOD) trades data in their organisation. The research participants at ICapital stated 94 per cent of the time that the RFDW data model was adopted to address generic business requirements of their organisation. The ICapital requirements tasked its data warehouse to ‘*on board*’ consumers across the enterprise and ensure that all the functional areas of the organisation use the same dataset. These requirements influenced the research participants to use a relational data model for their data warehouse. In the light of the requirements from their management, the research participants at ICapital did not find a multidimensional data model suitable for their data warehouse.

Program Manager	*If you are taking data from lots of different sources and primarily you are a data provider, it makes sense to stick to normalised model*

The research participants at ICapital rated high enterprise properties such as *generic* and *diverse requirements* for their data warehouse. Additionally, they rated properties such as *specific requirement*, *simplicity of requirement* low; these properties were considered suitable for a multidimensional data model. As a result, the implementation team at ICapital did not find a multidimensional data model suitable for their data warehouse. For instance:

Director	*I would say the dimensional model is perfect for just pure mart, a dimensional model for RFDW wouldn’t have worked*

The business requirements underpinning the ICapital relational data warehouse is complex and the data warehouse is intended to support the needs of numerous functional departments within the organisation. For instance:

Program Manager	*The primary reason for RFDW is to have one controlled end of day data so that everyone work off the same dataset so you have consistency across all the consumers whether it is credit risk, market risk, finance, product control, regulatory reporting or compliance*

The above outcome is also consistent with the literature; a relational data model is usually oriented to the enterprise view of the organisation and focuses on complex data interactions (Imhoff, Galemmo and Geiger, 2003; Tryfona, Busborg and Christiansen, 1999).

6.2.3 The Goal and Scope of the Data Warehouse

In the literature, the goal of a data warehouse defines the scope of a data warehouse (Giorgini et al, 2005; Imhoff, Galemmo and Geiger, 2003). However, the goal of a data warehouse does not exist in isolation; it reflects the priorities of an organisation as defined in the business requirements from a data warehouse (Giorgini, Rizzi and Garzetti, 2008). Addressing the goal and the scope of a data warehouse was rated high as a decision factor for the research participants at GWealth and ICapital respectively. The research participants at GWealth stated 77 per cent of the time that the goal and scope of their reporting data warehouse was a factor in adopting a multidimensional data model for the reporting data warehouse. Equally, the research participants at ICapital stated 90 per cent of the time that the goal and scope of Risk and Finance data warehouse was a factor in adopting a relational data model for RFDW. The outcomes of goal and scope decision factors for GWealth and ICapital cases are consistent with the literature and the research proposition RP3 – the goal and scope of data warehouse is a factor influencing the choice of a logical data model for a data warehouse. For the research participants at GWealth, the requirement from their management is in alignment with their organisation objective. In the business requirement, their organisation objective was translated into the goal and scope of the reporting data warehouse in support of the firm objective.

Director	*It is important to consider the goal of your data warehouse before you select your data model, this enables you to select the appropriate architecture having considered all other factors that are critical to achieving the goal of the data warehouse*
Data Architect	*You want to use the data model that is best fit for the goal of your data warehouse, otherwise, you may find your initial goals are not fully realised by using an inappropriate data model; that would be unacceptable to the business*

As described in Section 4.3.3, the firm objective was to stay ahead of the competition, of which, the reporting data warehouse was the key enabler. Additionally, the scope of GWealth data warehouse was limited to client reporting to achieve the firm goal. The goal and the scope defined for the reporting data warehouse in the business requirement document (BRD) influenced the decision to use a multidimensional data model for GWealth data warehouse.

Director	*The scope of the data warehouse is important as the goal of the data warehouse. In most cases, the goal of the data warehouse will determine the scope and size of the data*

warehouse, both are important factors in determining the choice of your data warehouse data model

In the literature, the requirements underpinning a multidimensional data warehouse is derived from a business area requesting the data warehouse (Imhoff, Galemmo and Geiger, 2003). The goal and the scope defined in the business requirement for the reporting data warehouse impacted the choice of the data model adopted for the data warehouse. Specifically, the GWealth data warehouse was used to produce defined type of client reports on a periodic basis. According to the research participants at GWealth, their data warehouse was not implemented to cater for the needs of different departments or functional areas at GWealth, nor was their data warehouse implemented to '*on-board*' diverse set of consumers within GWealth.

Program Manager	*The sole objective of the reporting warehouse is to use it for reporting purposes, our data model has allowed us to meet that objective*

While the empirical data indicated that the GWealth team has adopted a data model that is considered appropriate for their reporting data warehouse, it is questionable whether the implementation team at GWealth will use a relational data model for their data warehouse even if the scope of the reporting data warehouse is updated by their management to include addressing the reporting and analytics need of other departments at GWealth. In this situation, the scope of the reporting data warehouse will increase significantly and span the entire GWealth organisation.

However, the implementation teams at GWealth are inclined to use a multidimensional data model irrespective of the goal and scope requirements for a data warehouse. In this situation, it is difficult to see how the research participants at GWealth can be persuaded to consider a relational data model for their data warehouse even though there is a change in the scope and goal of the data warehouse.

The research participants at ICapital also rated the goal and the scope of their data warehouse high as decision factor and high priority item to address by the implementation team. The goal of ensuring the functional areas uses the same trade and finance dataset and the scope of their data warehouse spanning all the functional areas of their organisation combined to influence the research participants to use a relational data model for ICapital data warehouse. In the context of this research, for ICapital case study, the goal and the scope set out in their business requirements document impacted the choice of the data model adopted for their data warehouse. For instance:

Data Architect	*The goal of the business has a lot to do with it, one of the success criteria for RFDW is to what extent it satisfies the business goal, the project will be deemed failure if business goal is not met*

In alignment with the literature, data warehouses that are oriented to the needs of the enterprise are implemented using a relational data model in the CIF framework (Inmon, Imhoff and Sousa, 2001). As described in Section 2.3, a corporate information factory is a landscape of organisation systems.

Director	*If you are creating a wholesale system, which supports all the information of the firm in the case of RFDW, you probably want a model that is not too dimensional*

6.2.4 Implementation Orientation of Available Resources

In the literature, people make decisions using their previous experience and engage in different course of action for each situation based on that experience (Baron, 2004; Klein, 2008). The finding from this study is consistent with the literature on naturalistic decision making. As presented in Table 6-1, the influence of employee past experience as a decision factor was rated high by the research participants at GWealth and ICapital. This is also consistent with the research proposition RP4 – the implementation orientation of available resources is a factor influencing the choice of data model for a data warehouse. The research participants at GWealth stated 60 per cent of the time that their staff experience is a decision factor in adopting a multidimensional data model for the reporting data warehouse. Equally, the research participants at ICapital stated 81 per cent of the time that their staff experience was a decision factor in adopting a relational data model for their data warehouse. For the research participants at GWealth, the past experience of the key decision-makers in multidimensional data modelling is pronounced and in most cases, a multidimensional data model is the only data model they have used in their previous data warehousing projects. For the research participants at GWealth, a multidimensional data model was the model of choice to address the business requirement for client reporting.

As described in Section 4.3.4, the empirical data from the case study at GWealth provided the evidence that research participants' prior experience was one of the main drivers for adopting a multidimensional data model for the GWealth data warehouse. For instance:

Program Manager	*My preference is always dimensional warehouses but that's just because that is what I know. It was what people in the team know and have experienced and background*

The observation from the empirical data is that although other factors such as *performance*, *business requirements*, *goal and scope* of client reporting influenced the decision to adopt a multidimensional data model for the reporting data warehouse, it is also evident that the research participants previous experience played a major role in their decision to adopt a multidimensional data model for their data warehouse. For instance:

Director	*We have a number of senior architects in the team that are very familiar with dimensional architecture; that had big influence in choosing it*
Project Manager	*It's my experience of only ever worked with multidimensional data warehouse. The people I've worked with have always been kind of push for that kind of model*

The research participants at GWealth have mostly used a multidimensional data model in their previous careers. Their preference for a multidimensional data model was so strong to the extent that it is difficult to see how they could be persuaded to use any data model that is not multidimensional to address the reporting or non-reporting requirements of their data warehouse. While no one questions that the research participants at GWealth are ardent supporter of multidimensional data modelling, it is important to point out some shortcomings of their reluctance to consider an alternative data model such as a relational data model for their data warehouse. First, it appears the research participants at GWealth is used to using the same methodology for so long that they believed it provides solution to all data warehousing problems. This is problematic in the sense that their organisation may end up with silos of data warehouses which, in aggregate, may be costly to maintain and integrate for their organisation.

Second, it is evident the research participants at GWealth had considerable experience in multidimensional data modelling and the benefit that such experienced workforce brings to an organisation is considerable. Conversely, it can be argued that, over time, unless there is change in their implementation approach, the research participants experience may also hinder their long-term contribution to their organisation. The research participants may consider their approach appropriate for as long as they are presented with similar type of requirement as client reporting, however, this cannot be guaranteed. If the management at GWealth requires a data warehouse that captures its entire businesses and disseminates operational data to several departments for reporting and analytics, it is difficult to see how the GWealth research participants can apply their previous experience (in multidimensional data modelling) to requirements that demands new approach in addressing their organisation's requirements.

The research participants at ICapital equally rated their employee past experience high as a decision factor influencing the choice of data model for their data warehouse. The research participants at ICapital had overwhelming preference for using a relational data model for their Risk and Finance data warehouse.

The implementation team at ICapital knew from past experience which data warehouse project a relation data model is suited and recognised RFDW as one of such projects. For instance:

Director	*I would say the dimensional model is perfect for just pure mart, a dimensional model for RFDW wouldn't have worked*

In the literature, the impact and influence of previous experience in decision making is widely discussed. In naturalistic decision making, as people make decisions, they access and categorise a situation using their previous experience and make judgements about the appropriate course of action to take (Baron, 2004; Klein, 2008). As described in Section 2.5.4, naturalistic decision making shifts our understanding of decision making from a model that is based on rational choice to an approach that is knowledge based drawing on considerable experience on the part of a decision maker (Klein, 2008; Azuma et al, 2006). Similarly, it is evident from the empirical data that previous experience of the ICapital team was a factor in their decision to adopt a relational data model for their data warehouse. As presented in Section 5.3.4, the research participants at ICapital outlined a number of business requirements that a relational data model is suitably placed to address. One of the requirements identified by the ICapital research participants is building a data hub for their organisation. This type of acknowledgement on the part of the ICapital research participants is an indication of the level of awareness drawing on the experience of the implementation team at ICapital (Klein, 2008). For instance:

Project Manager	*You want to be able to able to support as much as possible different requirements so that those consumers in turn can support operations in their respective areas*

6.2.5 Enterprise data consolidation for heterogeneous reporting and analytics requirements

Enterprise capabilities such as firm-wide *data consolidation*, *integration* and *diversified consumers* were rated low by the research participants at GWealth for their data warehouse. However, these capabilities were rated high as a decision factor by the research participants at ICapital for their data warehouse. This outcome is consistent with the research proposition – RP 5 consolidating enterprise data to meet heterogeneous reporting and analytics requirements is a factor influencing the choice of data model for a data warehouse. The research participants at GWealth stated 53 per cent of the time that the reporting data warehouse has limited consumer base. This number indicated that more than half of the research participants at GWealth believed the reporting data warehouse has limited number of diverse users. On the other hand, the research participants at ICapital stated 95 per cent of the time that RFDW has diverse consumer base across their organisation. In the literature, multidimensional data warehouses are driven from the perspective of the business users (Bruckner, List and Schiefer, 2001; Prakash and Gosain, 2003; Sperley, 1999; Golfarelli and Rizzi, 1998; Husemann, Lechtenbörger and Vossen, 2002). A multidimensional data model is used to implement a data warehouse that supports a specific functionality of the business (Imhoff, Galemmo and Geiger, 2003; Kimball and Ross, 2002). The outcome of GWealth case study is in alignment with the literature. Their data warehouse is not used to address any enterprise requirement for GWealth; rather, the reporting data warehouse was used to support the creation of monthly statement reports for the clients of GWealth. As a result, the research participants rated their client reporting data warehouse low against enterprise capabilities such as firm wide *data consolidation*, *integration* and *diversified enterprise consumers*. For instance:

Data Architect	*The requirement was to develop a data warehouse for creating reporting statements for clients of the firm*
Program Manager	*The sole objective of the reporting warehouse is to use it for reporting purposes, our data model has allowed us to meet that objective*

In the literature, a data warehouse based on a relational data model enables a data warehouse to support multiple, diverse reporting and analytics requirements within an enterprise (Inmon, 2005; Drewek, 2005; Imhoff, Galemmo and Geiger, 2003). In alignment with the literature, the flexibility provided by a relational data model enabled the research participants at ICapital to integrate additional front office businesses into their data warehouse with less difficulty. This enabled the implementation team at ICapital to address varieties of stakeholders' requirements supporting multiple perspectives of their organisation. For instance:

Data Architect	*If you look at RFDW you will see it is evident from the way the entities are laid out, you can bring in more data later on with less difficulty*
Project Manager	*You want to be able to able to support as much as possible different requirements so that those consumers in turn can support operations in their respective areas*

The aim of this research is not to take a position on which logical data model is suitable for addressing the business requirements of the case organisations, rather, the aim of this research is to examine the decision factors that influenced the research participants in the case organisations to adopt the logical data model that was chosen for their data warehouses.

In the light of the literature, the findings presented in Section 4.4 and Section 5.5 and the discussion above, the next Section examines the degree of relationship between the research analytic code variables and the logical models adopted at GWealth and ICapital respectively.

6.3 Test of Relationship Between Analytic Codes and Logical Data Models

Table 6-4 (below) is a multivariate contingency table of the research logical data models and the frequency of the analytic codes presented in Section 4.4 and Section 5.5. A contingency table is used to analyse and record the relationship between two or more variables in hypothesis testing in order to decide whether or not there is a significant relationship between the variables in a contingency table. Significance in hypothesis testing implies that interpretation of the cell frequencies in a contingency table is warranted, on the other hand, an in-significant relationship between the variables in a contingency table implies that any differences in the cell frequencies may be explained by chance (Stockburger, 1996).

In alignment with the data analysis framework presented in Section 3.5.9, Gibbs (2007) notes the aim of selective coding is to identify the core category variables that are central to the phenomenon under study, as a result, the code variables in Table 6-2 from GWealth case study and the code variables in Table 6-3 from ICapital case study are the core category variables that are central to this study; these variables are the subset of the analytical code variables presented in Section 4.3 and Section 5.4 respectively. The code variables in the contingency table (Table 6-4) were derived from the analytic code variables in Table 6-2 and 6-3 respectively.

ACV	GWealth Analytical Codes Variables (ACV)	No of Frequency
1	High Query Performance	72
2	Specific Business Requirement	84
3	Defined Goal & Scope	65
4	Staff Experience	81
5	Enterprise Data Consolidation	22
	Total	324

Table 6-2: Number of Frequency of GWealth Analytic Codes

ACV	ICapital Analytical Codes Variables (ACV)	No of Frequency
1	High Query Performance	87
2	Generic Business Requirements	381
3	Defined Goal & Scope	197
4	Staff Experience	166
5	Enterprise Data Consolidation	436
	Total	1267

Table 6-3: Number of Frequency of ICapital Analytic Codes

In table 6-4 (below), the research **A**nalytic **C**ode **V**ariables (ACV) related to High Performance = ACV1, Business Requirements = ACV2, Goal and Scope = ACV3, Staff Experience = ACV4, Enterprise Data Consolidation = ACV5.

Logical Data Model (LDM)	OUTCOME					**Total**
	ACV1	ACV2	ACV3	ACV4	ACV5	
Multidimensional Data Model (**MDM**)	72	84	65	81	22	324
Relational Data Model (**RDM**)	87	381	197	166	436	1267
Total	159	465	262	247	458	1591

Table 6-4: Contingency table of Logical Data Models and Observed Frequency of Analytic Codes

In alignment with the aim of this research in Section 1.2, Table 6.5 presents the observed and expected code frequencies of the analytic codes to test the degree of significant relationship between the logical data models (*MDM, RDM*) and the research analytic code variables (*High Performance***,** *Business Requirements (Specific & Generic), Goal and Scope, Staff Experience, Enterprise Data Consolidation*), this is necessary to disprove the null hypothesis (*that there is no relationship between the analytic codes and the choice of logical data model*) inherent between the analytic code variables and the research logical models in support of the research propositions.

Logical Data Model (LDM)	OUTCOME (Observed & Expected)					**Total**
	ACV1	ACV2	ACV3	ACV4	ACV5	
Multidimensional Data Model (**MDM**)	72 (32.38)	84 (94.70)	65 (53.36)	81 (50.30)	22 (93.27)	324
Relational Data Model (**RDM**)	87 (126.62)	381 (370.30)	197 (208.64)	166 (196.70)	436 (364.73)	1267
Total	159	465	262	247	458	1591

Table 6-5: Observed and Expected Frequency of Analytic Codes and Logical Data Models

Table 6.5 computes the expected code frequency for the analytic codes using the formula below:

$$Ei,j = \frac{TiTj}{T}$$

Where

Ei,j = The expected frequency for cell i, j

Ti= Total column for the ith row

Tj= Total column for the jth column

T = Total number of observation

The expected cell frequency for MDM/ACV1 is 32.38 ((159*324)/1591). This value represents the impact of *High Performance* on choice of LDM indicating that if there is no relationship between *High Performance* and MDM the expected outcome of ACV1 is 32.38. Equally, expected cell frequency for RDM/ACV1 is 126.62 ((159*1267)/1591). This value represents the impact of *High Performance* on choice of LDM indicating that if there is no relationship between *High Performance* and RDM the expected outcome of ACV1 is 126.62. Table 6.5 shows the expected frequencies in parenthesis for the rest of the analytic code variables. In determining the significant relationship between the data models (MDM and RDM) and the research analytic code variables (ACV1, ACV2, ACV3, ACV4 and ACV5), the Chi Square formula below is used.

$$X^2 = \sum \frac{(O - E)2}{E} = \mathbf{157.5}$$

Where:

O = Observed code frequency

E= Expected code frequency

***157.5** = Chi Square value for ACV1 –ACV5 (See Appendix 5 for calculations)*

The degree of freedom (d.f.) is (r-1) (c-1) or (2-1) (5-1) = 4, *where r is the number of rows for the logical data models in the case studies and C is the number of columns for the research analytic code variables or outcomes in Table 6.5*. Table 6.6 (*below*) presents the distribution table to determine the *p* value. In Table 6.6 (below), the Critical Value for this test is **14.860** with degree of freedom (DF4); the critical value represents the value for accepting or rejecting the null hypothesis.

Chi Square Distribution Table										
DF	**0.995**	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	**14.860**
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997

Chi Square Distribution Table										
DF	**0.995**	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Table 6-6: Chi square distribution table, Source: *https://people.richland.edu/james/lecture/m170/tbl-chi.html, [Date Accessed: 26/05/2014; 30/09/2014]*

As indicated above, the critical value for the significant test is **14.860**, however, the value of Chi Square to test for significant relationship between the research analytic code variables (***ACV1***, ***ACV2***, ***ACV3***, ***ACV4***, ***ACV5***) the logical data models (***MDM***, ***RDM***) = **157.5**, this number is greater than critical value of **14.860**, thus; ***the null hypothesis is rejected indicating there is a significant relationship between the analytic code variables and the decision to adopt multidimensional or relational logical data model for a data warehouse***. In the light of the result presented above, the empirical data presented in this research provided the evidence to conclude that the research propositions (*RP1, RP2, RP3, RP4 and RP5*) are the important decision factors influencing the choice of logical data model for a data warehouse in the industry. Although, the value of Chi Square for all the analytic variables is higher than the *p* value for the significant test leading to a rejection of the null hypothesis, it is necessary to further examine the contribution of each of the analytic code variables to the value of Chi Square presented above for the logical models (*MDM, RDM*). Table 6.7 presents the percentage contribution of the ACV's for Table 6.4.

Logical Data Model (LDM)	Percentage (%) Contribution of ACVs' For MDM & RDM					**% Total**
	ACV1	ACV2	ACV3	ACV4	ACV5	
Multidimensional Data Model (**MDM**)	22.22%	25.93%	20.06%	25%	6.79%	20.4
Relational Data Model (**RDM**)	6.87%	30.07%	15.55%	13.10%	34.41%	79.6%
% Total	9.99%	29.23%	16.47%	15.52%	28.79%	100%

Table 6-7: Percentage Contribution of the Analytic Code Variables

The Sections are devoted to examining each of the analytic code variables and their significance to both multidimensional and relational data models.

6.3.1 ACV1 and the Significant Relationship Test

As presented in Table 6-5 above, the values of observed and expected outcomes of ACV1 for MDM are 72 and 32.38 respectively; these values are used for ACV1 together with the values of observed and expected outcomes of other ACVs' to determine the level significant relationship between the analytical variables ACVs' and the logical data models (MDM and RDM). As presented in Section 6.3, the Chi Square value equals 157.5; the contribution of ACV1 to the Chi Square value for ACV1 and MDM equals 48.48 indicating that the percentage contribution of ACV1 to the total value of adopting MDM for the GWealth reporting data warehouse is 22.22%. The Chi Square value indicated that there is a significant relationship between ACV1 (*high* *performance*) and the choice of multidimensional data model (MDM) for a data warehouse. Further examination of the observed and expected outcomes for ACV1 indicated that there is also a significant relationship between ACV1 and RDM (*i.e. the decision to use a relational data model for a data warehouse*). The contribution of ACV1 to the Chi Square value for ACV1 and RDM equals 12.40. Although the Chi Square for all the combined variables indicated that there is a significant relationship between the variables (*ACV1, ACV2, ACV3, ACV4, ACV5*) combined, further examination of the variable for ACV1 and RDM indicated that ACV1 contributed just 6.87% of the decision to adopt RDM for ICapital data warehouse. In a way, this is probably explain why a relational data model is not considered an optimal choice if high performance is a major requirement from a data warehouse; this may also explain why the RFDW team at ICapital adopted a database appliance as part of the strategy to achieve high performance for their data warehouse. The implication of this outcome is that, overtime; there is likely to be an increase in the use of appliance database (*as a means of improving performance)* in relational data warehousing.

6.3.2 ACV2 and the Significant Relationship Test

The values of observed and expected outcomes of ACV2 for MDM are 84 and 94.70 respectively; these values are used for ACV1 together with the values of observed and expected outcomes of other ACVs' to determine the level significant relationship between the analytical variables ACVs' and the logical data models (MDM and RDM). As presented in Section 6.3, the Chi Square value equals 157.5; the contribution of ACV2 to the Chi Square value for ACV2 and MDM equals 1.21. Although the Chi Square for all the combined variables indicated that there is a significant relationship between the combined variables (*ACV1, ACV2, ACV3, ACV4, ACV5*), the Chi Square value for ACV2/MDM approach indicated that there is a significant relationship between the variable ACV2 and MDM, for instance the percentage contribution of ACV2 to total value of adopting MDM is 25.93%, this value is 3.71% more than the contribution of ACV1 in the decision to adopt a MDM for a data warehouse. This indicates that the need to focus on addressing specific requirement for an organisation is an important factor in adopting a MDM for a data warehouse. The same is also true for ACV2 and RDM, for instance, the values of observed and expected outcomes for ACV2 and RDM are 381, 370.30 respectively; the contribution of ACV2 to the Chi Square value equals 0.31, however, the percentage contribution of ACV2 to the total value of adopting RDM for a data warehouse is 30.07%, this is an increase of 23.2% on the percentage contribution of ACV1 for adopting RDM, indicating that there is a significant relationship in adopting a relational data model to address generic requirements for an organisation.

6.3.3 ACV3 and the Significant Relationship Test

The values of observed and expected outcomes of variable ACV3 for MDM are 65 and 53.36 respectively; these values are used for ACV3 together with the values of observed and expected outcomes of other ACVs' to determine the level significant relationship between the analytical variables ACVs' and the logical data models (MDM and RDM). As presented in Section 6.3, the Chi Square value equals 157.5; the contribution of ACV3 to the Chi Square value for ACV3 and MDM equals 2.54 indicating that the percentage contribution of ACV3 to the total value of adopting MDM for the GWealth reporting data warehouse equal 20.06%, this value is less than the percentage contribution of ACV1 and ACV2 (22.22%, 25.93%) respectively. Equally, the contribution of ACV3 to the Chi Square value for ACV3 and RDM equals 0.65. As presented in Table 6-5, the percentage contribution of ACV3 to the total value of adopting a RDM for a data warehouse equals 15.55%, this value is 14.52% less than the percentage contribution of ACV2 (30.07%) to the total value of adopting a RDM for a data warehouse. This suggests that ACV2 is very significant in the decision to adopt a relational data model for a data warehouse.

6.3.4 ACV4 and the Significant Relationship Test

The values of observed and expected outcomes of variable ACV4 for MDM are 81 and 50.30 respectively; these values are used for ACV4 together with the values of observed and expected outcomes of other ACVs' to determine the level significant relationship between the analytical variables ACVs' and the logical data models (MDM and RDM). As presented in Section 6.3, the Chi Square value equals 157.5; the contribution of ACV4 to the Chi Square value for ACV4 and MDM equals 18.74 indicating that the percentage contribution of ACV4 to the total value of adopting MDM for the GWealth reporting data warehouse equals 25%. This value is 4.94% greater than the contribution of ACV3 for adopting MDM and 2.78% greater than the contribution of ACV1 for adopting a MDM for a data warehouse. This suggest that employee past experience is considered a significant decision factor (*than performance and goal and scope decision factors*) in adopting a multidimensional data model for a data warehouse. The values of observed and expected outcomes of variable ACV4 for RDM are 166 and 196.70 respectively. The percentage contribution of ACV4 to the total value of adopting RDM equals 4.79, in percentage term, the value of contribution of ACV4 to the total value of adopting a RDM is

13.10%, this value represents an increase of 6.23% compared to the contribution of ACV1 for adopting a RDM for a data warehouse. This also suggest that employee past experience is considered a significant decision factor (*than performance decision factor*) in adopting a relational data model for a data warehouse.

.

6.3.5 ACV5 and the Significant Relationship Test

The values of observed and expected outcomes for ACV5 and MDM are 22 and 93.27 respectively; these values are used for ACV5 together with the values of observed and expected outcomes of other ACVs' to determine the level significant relationship between the analytical variables ACVs' and the logical data models (MDM and RDM). As presented in Section 6.3, the Chi Square value equals 157.5; the contribution of ACV5 to the Chi Square value for ACV5 and MDM equals 54.46, however, the percentage contribution of ACV5 to the total value of adopting MDM for the GWealth reporting data warehouse equals 6.79%. As presented in Section 5.4.1, the reporting data warehouse also scored low on other analytic codes to assess the enterprise properties of GWealth reporting data warehouse. For instance, *enterprise consumer base* and *diversification of the data model* accounted for just 4 and 2 per cent of the total percentage of the code frequency for ACV5, indicating that the research participants did not consider those enterprise factors (*i.e. enterprise consumer base, enterprise consumer base and diversification of the data model*) as decision factors for adopting a multidimensional data model for client reporting data warehouse. On the other hand, the values of observed and expected outcomes for ACV5/RDM for ICapital data warehouse are 436 and 364.73 respectively; the percentage contribution of ACV5 to the total value of adopting RDM equals 13.92. This outcome is consistent with the empirical data, for instance, the research participants at ICapital stated 95 per cent of the time that enterprise factors such as *enterprise data consolidation* and *enterprise consumer base* were important factors in adopting a relational data model for RFDW. In percentage term, the percentage contribution of ACV5 to the total value of adopting a RDM for data warehouse is 34.41%, this value is 4.34% greater than the percentage contribution of ACV2 (30.07%) and 5.76% greater than the combined percentage contributions of ACV3, ACV4. The percentage contribution of ACV5 for ACV5/RDM is also 27%.54% greater than the contribution of ACV1 for adopting a RDM for a data warehouse, indicating that ACV5 is considered a significant decision factor compared to other analytic code variables in adopting a relational data model for a data warehouse. In the light of the analysis presented above, the next Section presents the revised conceptual decision model for adopting a logical data model for a data warehouse.

6.4 Decision Matrix for Logical Data Model Adoption

Table 6-9 presents a decision matrix for logical data model adoption for this study. A decision matrix enables a decision maker to evaluate and prioritise his/her options; it is useful in situations where a decision maker has a number of alternatives to choose from and where there are many different factors to take into account (Brooks, 2014). The following processes were engaged in constructing the decision matrix presented in Table 6-9 and are in alignment with the American Society for Quality (ASQ, 2014):

- Identify evaluation criteria, the evaluation criteria for the decision matrix in Table 6.9 are the analytic code variables (ACV) presented in Section 6.3; these are the decision factors that are considered to be impacting the choice of logical data model (LDM) in data warehousing.
- Assign relative weight to each ACV based on importance of each ACV to the choice of LDM, Note 2 in Table 6-8 presents the derivation of relative weight for the decision matrix.
- Establish and assign rating scale to each ACV in order to evaluate each ACV for logical data model adoption. Note 1 in Table 6-8 presents the derivation of rating scale for the decision matrix.

In order not to introduce bias into the decision matrix, the relative weight for each of the analytic variables is kept constant at 3 indicating that the relative weight of each analytic code variable is high as indicated in Note 2 in Table 6-8. Equally, the rating scale of ACV is derived from the percentage contribution of each ACV to the value of Chi-Square presented in Sections 6.3.1 – 6.3.5; the values in column "*% Contribution to Chi-Square*" were extracted from Table 6-7 in Section 6.3. Additionally, Note 1 in Table 6-8 provides the derivation and translation of rating scale for decision matrix in Table 6-9.

ACV	Analytical Codes Variables (ACV)	LDM	Relative Weight (RL)	% Contribution to Chi-Square	Rating Scale of %Contribution to Chi-Square (RS)
1	High Query Performance	MDM	3	22.22%	3
		RDM	3	6.87%	1
2	Specific Business Requirements	MDM	3	25.93%	3
	Generic Business Requirements	RDM	3	30.07%	3
3	Defined Goal & Scope	MDM	3	20.06%	3
		RDM	3	15.55%	2
4	Staff Experience	MDM	3	25%	3
		RDM	3	13.10%	2
5	Enterprise Data Consolidation	MDM	3	6.79%	1
		RDM	3	34.41%	3

Note 1:	**Derivation of Ratings is Based on % of Contribution ACV to Value of Chi Square:** • Low = (<=10%) • Medium = (>10%, <=20%) • High = >20% **Rating Scale of %Contribution to Chi-Square** • Low = 1 • Medium = 2 • High = 3
Note 2:	**Relative Weight Values: 1 = Low; 2 = Medium; High = 3 (***These values represent the relative importance of ACV to LDM adoption. For this study, all ACVs are considered important and assigned relative value of 3 (High) in Table 6-9***)**
Note 3:	**RL** = Relative Weight **RS** = Rating Scale of %Contribution to Chi-Square
Note 4:	The values of % contribution of all ACVs to Chi-Square is presented in Table 6-7, Section 6.3

Table 6-8: Derivation and Translation of Relative Weight and Rating Scale for Decision Matrix

ACV	Analytical Codes Variables (ACV)	Weight (Constant)	MDM (RL * RS) Points	RDM (RL * RS) Points
1	High Query Performance	3	3 * 3 = 9	3 * 1 = 3
2	Business Requirement	3	3 * 3 = 9	3 * 3 = 9
3	Defined Goal & Scope	3	3 * 3 = 9	3 * 2 = 6
4	Staff Experience	3	3 * 3 = 9	3 * 2 = 6
5	Enterprise Data Consolidation	3	3 * 1 = 3	3 * 3 = 9
	Total		**39**	**33**

Table 6-9: Decision Matrix for Logical Data Model Adoption in Data Warehousing

In decision matrix, option with the highest score is the decision that should be made (Brooks, 2014), in this context, based on the criteria defined by the analytic code variables (ACVs), multidimensional data model (MDM) scored the highest number of points (39) against 33 for the relational data model (RDM). Based on the total score of relative weight and rating scale of percentage contribution of ACVs, MDM is suggested by the decision matrix as the logical data model that should be adopted for a data warehouse.

Although, the decision matrix suggested that MDM should be adopted for a data warehouse based on the total score of the decision matrix, there are interesting observations in Table 6-9 that should be addressed. For instance, there are variations in total score of individual ACVs when considered on individual basis. As can be observed in Table 6-9, the total score of relative weight and rating scale of percentage contribution of ACV5 for RDM outweighed that for

MDM, in this case, if the only criterion (*discounting other criteria i.e. ACV1 – ACV4*) for implementing a data warehouse is *enterprise data consolidation* to address different reporting and analytics requirements within an organisation, based on the score of this decision factor (*9 points for RDM against 3 points for MDM for ACV5*), RDM should be adopted as the logical data model for a data warehouse. Furthermore, if decision factors relating to *business requirements* and *enterprise data consolidation* are the primary criteria for implementing a data warehouse (*i.e. discounting ACV1, ACV3 and ACV4*), RDM should be adopted as the logical data model for a data warehouse. As can be observed in Table 6-9, the aggregation of total score of relative weight and rating scale of percentage contribution of ACV2 and ACV5 for RDM equals 18 points against 12 points for MDM.

6.5 The Revised Conceptual Model for Data Model Decision

Causal relationship is a process in which data is used to infer causality by an investigator. Causality is defined in terms of observable and unobservable events (Hidalgo and Sekhon, 2011). In the light of the empirical findings presented in Sections 4.4 and Sections 5.5, the discussions presented in Section 6.2, the results presented in Section 6.3 and Sub-Sections 6.3.1-6.3.5 (*testing the degree of significant relationship between the research analytic code variables and the research logical modes (LDM)*) and the decision matrix presented in Section 6-4, Figure 6-1 presents the revised conceptual model introduced in Section 2.7.6 for adopting a logical data model for a data warehouse.

6.5.1 Difference Between Proposed and Revised Conceptual Models

The revised conceptual model in Figure 6-1 builds on the proposed conceptual model presented in Section 2.6.6; however, both models are different in the following ways. First, the revised conceptual model is grounded in the empirical data from the case studies; also, the decision matrix presented in Section 6.4 is introduced into the revised conceptual model. In contrast, the proposed conceptual model presented in Section 2.6.6 was developed from the review of literature.

Second, inherent null hypothesis exists in this study, for instance, it can be argued that ***there is no relationship between the analytic codes and the choice of logical data model (null hypothesis)***; however, this inherent null hypothesis is rejected. The value of Chi Square presented in Section 6.3 confirmed that there is a significant relationship between the research analytic code variables (***ACV1***, ***ACV2***, ***ACV3***, ***ACV4***, ***ACV5***) the logical data models (***MDM***, ***RDM***)**.** The conceptual model is revised based on the result of this test.

Third, based on the empirical data, analysis of the decision matrix presented in Section 6.4 provided pathways for logical data model adoption based on *total score of relative weight* and *rating scale of percentage contribution of ACVs* that are the key determinants of implementing a data warehouse. As observed from the analysis of the decision matrix presented in Section 6.4, if decision factors relating to *High Query Performance, Defined Goal & Scope, Staff Experience* (*ACV1, ACV3 and ACV4*) are the primary criteria for implementing a data warehouse, decision matrix suggests that MDM should be adopted as the logical data model for a data warehouse. However, if decision factors relating to *business requirements* and *enterprise data consolidation* (*ACV2 and ACV5*) are the primary criteria for implementing a data warehouse, decision matrix suggests that RDM should be adopted as the logical data model for a data warehouse. The conceptual model is also revised based on the prescription of the decision matrix.

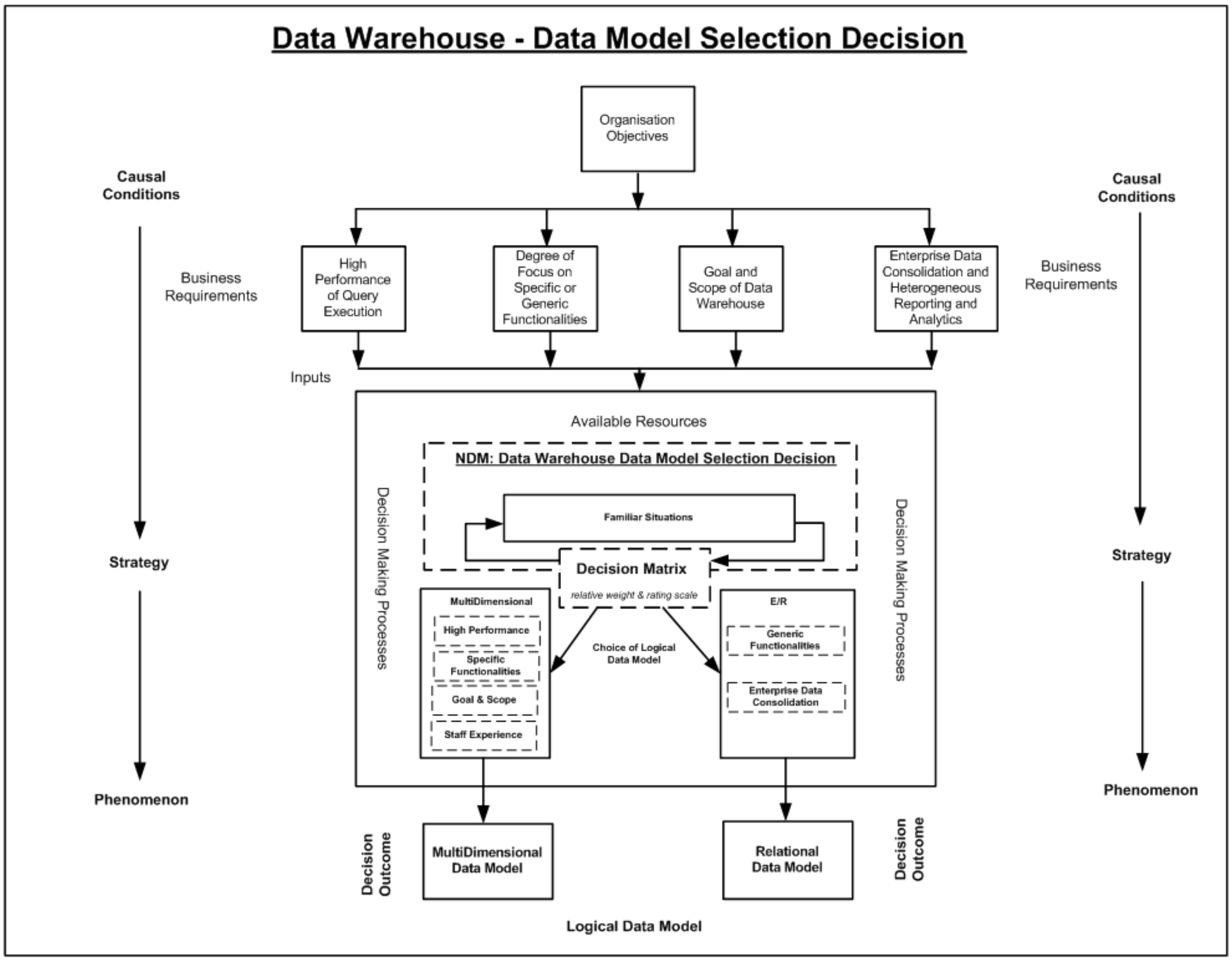

Figure 6-1: Revised Conceptual Decision Model for Selecting Logical Data Model

6.5.2 The Revised Conceptual Model

As illustrated in Figure 6-1, the factors influencing the choice of logical data model presented in this thesis are inputs into the proposed decision model; these factors are the decision criteria in the decision matrix presented in Section 6.4. As discussed in Sections 4.3.3 and 5.3.3, an organisation objective represents the priority of a firm for its data warehouse. In Figure 6-1, the decision factors representing *high performance of query execution*, *degree of focus on specific or generic functionalities*, *goal and scope*, *enterprise data consolidation and reporting* are collectively documented as the business requirements from a data warehouse. In the case organisations, the business requirements encapsulated the firm objectives and provided the requirements to be addressed by both data warehouses (Inmon, 2005; Inmon, Imhoff and Sousa, 2001; Kimball and Ross, 2002; Sperley, 1999; Reeves, Ross, Thornthwaite and Kimball, 1998).

In naturalistic decision making, people make decisions using their previous experience and engage a course of action based on that experience (Baron, 2004; Klein, 2008). As presented in Section 2.6.4, the implementation orientation of the available resources is defined by the past experience of the employees tasked with building a data warehouse. In the literature, in naturalistic decision making, familiar situation is characterised by pattern recognition, this provides reference, context and highlights potential cues between current decision situation and past situations, if there is a match; a course of action similar to previous decision situation is taken (Klein, 2008).

Naturalistic decision making is depicted in the revised conceptual model as a process titled "***Familiar Situation***"; this is also considered the OODA loop of the conceptual model as presented in Section 2.5.4. As presented in Section 4.4.4 and Section 5.5.4; in the case organisations, the past experience of the research participants in multidimensional and relational data modelling is pronounced and in most cases the research participants were able to identify the requirements which multidimensional and relational data models are suited. As presented in Section 4.3.4, the research participants at GWealth have mostly used a multidimensional data model in their previous careers. Their preference for the data model was strong to the extent that it is difficult to see how they could be persuaded to use any data model that is not multidimensional to address any requirement for a data warehouse. As presented in Section 6.2.4, the research participants at ICapital knew from past experience which data warehouse project a relation data model is suited and recognised their data warehouse project as one of such projects; this is an evidence of pattern recognition, a key component in naturalistic decision making. As presented in Section 5.5.4 and 6.2.4, the research participants not only articulated that they considered multidimensional data model to be perfect for implementing pure data mart, they also recognised that in light of the business requirements for their data warehouse, they considered it appropriate to engage a "*normalised model*" for ICapital data warehouse. In the literature, in naturalistic decision making, as people make decisions, they access and categorise a situation using their previous experience and make judgements about the appropriate course of action to take (Baron, 2004; Klein, 2008).

Based on the outcome of the test of significant relationship between the research analytic code variables (***ACV1***, ***ACV2***, ***ACV3***, ***ACV4***, ***ACV5***) the logical data models (***MDM***, ***RDM*)**, the decision matrix presented in Section 6.4 is introduced into the revised conceptual model. As discussed in Sections 4.3 and 5.3, the decision factors relating to *high performance of query execution, goal and scope* are non-functional requirements. They both accounted for 80% and 68% of the total percentages of the decision factors that influenced the choice of multidimensional data model for GWealth data warehouse. As a percentage of the research propositions, these factors accounted for 27% and 24% respectively. Additionally, the Chi Square presented in Section 6.3.1 indicated that there is a significant relationship between ACV1 and MDM and ACV1 and RDM, the Chi Square for both methods were 48.48 and 12.40 respectively. Furthermore, as presented in Section 6.3.1, the percentage contribution of ACV1 to the total value of adopting MDM or RDM for a data warehouse is 22.22% and 6.87% respectively. Equally, Section 6.3.3 indicated that there is a causal relationship between ACV3 and MDM and causal relationship between ACV3 and RDM. The value of Chi Square for a significant relationship between ACV3/MDM and ACV3/RDM are 2.54 and 0.65 respectively, in percentage term, the contribution of ACV3 to the total value of adopting MDM or RDM for a data warehouse is 20.06% and 15.55% respectively. As discussed in Section 4.3 for GWealth case study, the decision factor relating to *the degree of focus on specific business requirement* accounted for 75% of the overall total decision for adopting a multidimensional data model for GWealth data warehouse. This factor also accounted for 26% of the research proposition.

Similarly as discussed in Section 5.3, for ICapital, *the degree of focus on generic business requirement*, *enterprise data consolidation, heterogeneous reporting and analytics requirements* accounted for 94% and 95% of the total decision percentages for adopting a relational data model for ICapital data. As presented in Section 6.3.2, there is a causal relationship between ACV2 (*degree of focus on specific business requirement*) and MDM and a causal relationship between ACV2 (*degree of focus on generic business requirements*) and RDM. The value of Chi Square for a significant relationship between ACV2/MDM and ACV2/RDM are 1.21 and 0.31 respectively, in percentage term, the contribution of ACV2 to the total value of adopting MDM or RDM for a data warehouse is 25.93% and 30.07% respectively.

As discussed in Sections 4.3 and 5.3, decision factor relating to the *orientation of available resources* represented 60% and 88% of the total decision factors for adopting the multidimensional data model for GWealth data warehouse and relational data model for ICapital data warehouse respectively. As shown in Section 6.3.4, there is a causal relationship between ACV4 (*staff experience*) and MDM and causal relationship between ACV4 and RDM. The value of Chi Square for a significant relationship between ACV4/MDM and ACV4/RDM are 18.74 and 4.79 respectively, in percentage term, the contribution of ACV4 to the total value of adopting MDM or RDM for a data warehouse is 25% and 13.10% respectively.

The decision matrix presented in Figure 6-1 provided the pathways for adopting multidimensional and relational data models in the conceptual decision model, the paths are presented as M - PATH and R - PATH respectively. As presented in Section 6.4, Brooks (2014) suggests that option with the highest score in a decision matrix is the decision that should be adopted. In Table 6-9, the total score of relative weight and rating scale of percentage contribution of *ACV1, ACV3 and ACV4* for MDM outweighed that for RDM MDM (*27 points for MDM, 15 points for RDM*). In Figure 6-1, business requirements for data warehouses with ***high degree of focus on performance***, ***high degree of focus on specific functionality***, ***high degree of focus on goal and scope*** *and* ***high degree of staff experience*** are categorized via the M - PATH, this path represents the multidimensional route. In accordance with the analysis of decision matrix presented in Table 6-9, the M - PATH suggests such business requirements are best implemented using a multidimensional data model. As presented in the decision matrix (*Table 6-9 in Section 6.4*), RDM accumulated 18 points against 12 points for MDM in total score of relative weight and rating scale for ACV2 (*generic business requirements*) and ACV5 (*enterprise data consolidation for heterogeneous reporting and analytics requirements*) respectively, as a result, in Figure 6-1, business requirements for a data warehouse with ***high degree of focus on generic functionalities***, ***high degree of enterprise reporting and analytics*** are categorized via R - PATH, this path represents the relational route. The R - PATH suggests such business requirements are best implemented using a relational data model provided that the primary purpose of implementing a data warehouse is to address generic business requirements to cater for various reporting and analytics requirements within an organisation.

6.6 Implications for Practice

Cavaye (1996) pointed out that case study is applicable where knowledge is limited; the success of a research in information system is judge by whether knowledge is improved by a research and to what extent the knowledge gained can be applied in practice (Galliers and Land, 1987). The studies presented in this thesis are important because the research findings provide the understanding necessary to explain why pluralistic data architectures exist in data warehousing. The research adds to knowledge in data warehousing by providing a framework that guides decision making in adopting a logical data model for a data warehouse. Unlike the research in data warehousing that investigated which data architecture of a data warehouse is most successful (Ariyachandra and Watson, 2006); this research makes no assertion about which logical data model that is considered successful or otherwise. Rather, this research investigated and documented the nature of the business requirements for a data warehouse that is empirically suitable for multidimensional and relational data models. In so doing, the conceptual decision model presented in Figure 6.3 provided the template for solution architects and practitioners in the industry when considering the type of logical data model on which to implement a data warehouse. The implications of this research for practice focuses on two areas namely: (a) Requirements and funding considerations (b) Data integration considerations.

6.6.1 Requirements and Funding Considerations

In reflecting on the interviews conducted with the research investigator, the key research participants at GWealth indicated that they did not go through any rigorous decision making process when they adopted a multidimensional data model for their reporting data warehouse. The research participants also commented that they did not consider the possibility of exploring or using any other data model for the reporting data warehouse in the light of the business requirements that was presented to them at the time. However, they commented that given the interviews and the discussions that took place in this study, they now realised that, much more could have been achieved for the business with the reporting data warehouse. For Instance:

Lead Data Architect	*If I were to do it all again, I wouldn't do it in a Kimball style schema, because it is too difficult to change*

The research participants also commented that the funding model for the reporting data warehouse did not enable them to expand the scope of the reporting data warehouse:

Development Manager	*We could have got more out of it had the funding model being different and then, had the business witnessed our ability to get more out of it*

The research participants commented that this study got them to think about the other reporting systems they have in the organisation. They commented that the reporting systems: "*they all do pretty much the same thing*", with appropriate funding they would have the capability to cover the broadest set of requirements for the firm:

Lead Data Architect	*Having the right funding model, we would have captured the broadest sets of requirements than we had… from then, we start to synthesize a solution that meets our requirements. I would be very surprised if we didn't end up with relational schema and non-relational marts implemented either in views or hard marts*

The conceptual decision model for selecting a logical data model will help the research participants at GWealth in addressing the problems described above. The conceptual model exposes the industry to the varieties of business requirements that multidimensional and relational data models are suited thereby allowing data warehouse implementation teams to evaluate the pros and cons of each the data model in relation to their requirements. By engaging in the study, the GWealth team realised that rather than focusing on the narrow requirement of client reporting, their organisation could have benefited long term by developing a data warehouse that allows for various forms of reporting and analytics capabilities for their organisation. The GWealth team could deploy the conceptual decision model to articulate to the business the costs and benefits associated with a data warehouse that focuses only on client reporting versus a data warehouse that addresses the current reporting requirements and also lay a foundation for future reporting and analytics solutions for the organisation. By deploying the conceptual decision model to demonstrate to the business the benefits that would be realised by the organisation in the immediate future, the implementation team at GWealth (*or any team in any organisation in similar situation*) can secure necessary funding to implement an appropriate data warehouse for their organisation. Such a data warehouse will allow the business to address current and potential future requirements; this would provide better value and long term benefits to their organisation.

6.6.2 Data Integration Considerations

The research participants commented that data integration and operational processing are the other areas where they experienced some difficulties. Because the GWealth team is sourcing data from multiple sources, their extraction, transformation and data loading (ETL) processes become so complex to the extent that it impacted the overall performance of their data warehouse.

Worse still, the complexity of the ETL processes and the performance degradation they experienced (*as a result of the ETL*) affected the morale of some research participants with ownership responsibilities for the data warehouse:

Program Manager	*We try to bring in multiple sources so the loading of the fact became quite complex….we ended up putting in these defects fixes in our end of day process to sort of go back and fix it…that got us over certain hurdles and obviously the intention was to go back and fix it properly…that never happened…I never liked that*

The research participants did not see the connection between the problem described above and their data model:

Project Manager	*They are particularly not related to data architecture*

This report is beneficial to the research participants (*and teams in other organisations*) enabling them to establish the connection between their data model and the complexity of their ETL processes. Although this report will not immediately address the particular data integration and operational issues described above for the GWealth team, however, this report enables other organisations to be aware of this type of "*foundation issues*" when making decisions about the type of logical data model that will be used for their data warehouse.

6.7 Implications for Theory

Theory is an abstract entity whose aim is to explain, describe and enhance our understanding of the world, in some cases, it makes predictions of what will happen in the future and provide a basis for intervention and action (Gregor, 2006). The contributions of this research are explanatory and predictive in nature ensuring an investigation of the phenomenon under study; the thoughts put forward were developed and integrated to support the explanation of the phenomenon leading to a development of the conceptual model for data model decision. The implication of the research to theory is that the study provides an explanation why the logical data models based on multidimensional and relational data models defined data architecture in data warehousing. It is impractical for a study to claim it has provided all the explanations associated with adopting logical data model for a data warehouse, however, as organisations implements data warehouses as the foundation for their business intelligence solutions, the study provides important considerations that should be taken into account when deciding the type of logical data model to be adopted for a data warehouse. The conceptual model presented in Figure 6-1 addressed the functional requirements and the humanistic experience of executing a business requirement for a data warehouse. In so doing, the conceptual model (*for data model decision*) provided the association between the research propositions; synthesises of business requirements and recognition-primed decision model (NDM) leading to decision pathways for logical data model adoption in data warehousing. In the conceptual model, the M - PATH suggests a requirement with *high degree of focus on performance, high degree of focus on specific functionality, high degree of focus on goal and scope and high degree of staff experience* is executed via M - PATH, indicating adopting a multidimensional data model for a data warehouse. The R - PATH indicated a requirement with *high degree of focus on generic functionalities, high degree of enterprise reporting and analytics* is executed via R - PATH, indicating adopting a relational data model for such a data warehouse.

6.8 Summary

This chapter discussed the findings of the case studies at GWealth and ICapital respectively; in so doing, the chapter discussed the decision factors impacting the choice of logical data model in alignment with the research propositions. The chapter comparatively examined each of the analytic codes, their properties and how they aligned with the research propositions. The chapter engaged the literature and the empirical data to drive the discussion on the research propositions. Additionally, the chapter addressed the *null hypothesis* inherent in the research (*i.e. that there is no relationship between the analytic codes and the choice of logical data model*), using a Chi Square method on the contingency table of observed and expected code frequencies for the case organisations; the chapter examined the relationship between the research analytic code variables (*ACV1, ACV 2, ACV 3, ACV 4, ACV 5*) and the logical data models (*MDM, RDM*). The Chi Square test yielded a positive result leading to a rejection of the *null hypothesis*, the chapter further examined the contributions of each of the analytic code variables to the Chi Square value and their significance for each of the logical models (*MDM, RDM*). While the Chi Square test for the analytical code variables indicated that there is a significant relationship between the analytic code variables (*ACV1, ACV2, ACV3, ACV4, ACV5*) and the logical models (MDM and RDM), further analysis and examination of each of the analytical code variables also indicated that there is a significant relationship between ACV1/MDM, ACV1/RDM, ACV2/MDM, ACV2/RDM, ACV3/MDM, ACV3/RDM, ACV4/MDM, ACV4/RDM and ACV5/MDM, ACV5/RDM.

Furthermore, based on the analysis of the contributions and the significance of each of the analytic code variables to the logical data models, the chapter presented a revised conceptual decision model for logical data model adoption in data warehousing. The decision model examined the causal relationships between the decision factors, provided synthesises of the business requirements and recognition-primed decision model (NDM) leading to the decision pathways for adopting a logical data model in data warehousing. Finally, the chapter examined the implications of the research for practice and theory; the chapter discussed the contributions of the work to industry and theory by identifying the nature of the business requirements applicable to the implementation pathways in the revised conceptual model for logical data model adoption.

CHAPTER 7 - Conclusion

7.1 Research Summary

This research was motivated by polarization of data architecture in data warehousing arising from the adoption of multidimensional and relational data models in the industry. The aim of this research is to examine the decision factors influencing the choice of logical data model for data warehouses in the industry. In pursuit of that aim, the research examined naturalistic decision making as the theoretical foundation for the research. The empirical data from the study provided the means of data analysis necessary to achieve the objectives of the research. The objectives set out in Chapter 1 are summarised below:

Objective 1: Review the state-of-the-art in data warehouse development in order to distill the factors affecting the choice of data model in data warehousing.

Objective 2: Synthesise the literature in order to try and develop a framework impacting the choice of logical data model in data warehousing.

Objective 3: Undertake a comparative case study in the organisations using the multidimensional and relational data models to test the factors impacting the selection of logical data model for a data warehouse.

Objective 4: Evaluate the outcome of the study to determine the alignment of the decision factors in practice and contributions to theory

In achieving the aim and objectives of the study, **Chapter 2** presented an overview and key components of a data warehouse system, a review of the literature in data warehousing was explored and different phases of implementing a data warehouse was discussed. In the chapter, a review of the decision theories was examined, in particular, the normative, descriptive, prescriptive and naturalistic decision theories. The decision to adopt a logical data model for a data warehouse was explored and multidimensional and relational data models were compared and contrasted. The chapter presented and discussed the research propositions, the key factors considered to be influencing the choice of logical data model in data warehousing as indicated by the literature.

Chapter 3 described the means of achieving the aim and objectives of this study through the research design. The design outlined the practical steps and processes engaged to address the research problem. This study used a comparative case study for the case organisations, GWealth and ICapital, where multidimensional and relational data models respectively were implemented for their data warehouses. The components of the research design framework were presented and discussed. For the study, the interview of research participants provided the source of data collection for the research. The method of linking the empirical data to the research propositions were presented and discussed. Finally, grounded theory provided the framework for data analysis for the study using qualitative and quantitative methods.

Chapter 4 presented the empirical data from the interviews conducted at GWealth. The main focus of the case at GWealth was to examine the factors that influenced the research participants to adopt a multidimensional data model for their data warehouse and the extent to which those factors supported or disproved the research propositions presented in Chapter 2. The empirical findings of the case study were presented and aligned with the propositions they addressed. For each of the research findings, the frequency analysis of the analytic codes were presented and discussed in alignment with the research propositions. The summary of the findings were presented and discussed.

Chapter 5 presented the empirical data from the interviews conducted at ICapital. The case study at ICapital examined the factors that influenced the research participants to adopt a relational data model for their data warehouse and the extent to which the empirical findings supported or disproved the research propositions presented in Chapter 2. In presenting the results, the empirical findings from the case study were presented and aligned with the propositions they addressed. The frequency analysis of the analytic codes were presented and discussed in accordance with the research design framework presented in Chapter 3

Chapter 6 provided an in depth discussion of the factors that influenced the choice of logical data model in the case organisations using the empirical data presented in Chapter 4 and Chapter 5. The chapter critically examined the research findings in alignment with the literature and naturalistic decision making theory. The analysis of each of the empirical code categories were presented and discussed in alignment with the research propositions.

The chapter further presented a contingency table of observed and expected code frequencies for the research analytic code variables (ACVs') and tested the degree of significant relationship between the logical data models (***MDM***, ***RDM***) and the research ACVs' (***ACV1, ACV2, ACV3, ACV4, ACV5***). The chapter further examined the contributions of each of the analytic code variables and their significance to each of the logical models (***MDM***, ***RDM***). Following that, the conceptual decision model introduced in Chapter 2 was revised in the light of the findings presented in Section 6.3 and Sub Sections 6.3.1 – 6.3.5. Finally, the implications of the study for practice and theory were discussed.

7.2 Research Contributions

The findings of this research provided the understanding to explain why multidimensional and relational data models defined data architecture landscape in data warehousing. Apart from the findings that a choice of logical data model for a data warehouse is impacted by decision factors such as: *high degree of focus on performance, high degree of focus on specific or generic functionalities, high degree of focus on goal and scope* and *high degree of focus on enterprise data consolidation* as presented in Sections 4.5 and 5.6 respectively, the research finds the research participants in the case studies were keen advocates of their respective logical data models (*MDM* and *RDM*), the research participants had strong preference for the logical data models they adopted for their respective data warehouses as presented in Sub Sections 4.4.4 and 5.5.4 respectively . The study finds that prior experience of the research participants in multidimensional and relational data modelling respectively is an important factor in adopting the MDM and RDM for the data warehouses in the case organisations.

Furthermore, this study has researched an area where there is no known research on the subject; although a study was conducted on which data warehouse architecture is most successful (Ariyachandra and Watson, 2006), that study largely compared different architecture methods, based on four measures (*information quality*, *system quality*, *individual impact* and *organisation impact*) it determined which method was successful. The decision factors that influenced the choice of logical data model for a data warehouse which, this study investigated has not been previously researched and documented in the literature. The study presented in this thesis investigated data warehousing from a logical data model perspective; the study focuses on understanding and providing explanation why logical data models based on multidimensional and relation data models defined the data architecture landscape in data warehousing; consequently the study expanded the literature in this area, as related studies is limited.

The revised conceptual decision model provides the foundation to elucidate the decision pathways for adopting a logical data model in data warehousing; this is a contribution of this research. As presented in Section 6-5, the conceptual model is grounded in the empirical data and confirms the existence of significant relationship between the research analytic code variables ACVs (*ACV1, ACV2, ACV3, ACV4 and ACV5*) and LDM (*MDM, RDM*) leading to rejection of null hypothesis inherent in the study in Section 6.3. As presented in Section 6.5.2, in accordance with the analysis of decision matrix presented in Table 6-9, the conceptual decision model suggested that multidimensional data model should be adopted for a data warehouse with ***high degree of focus on performance***, ***high degree of focus on specific functionality***, ***high degree of focus on goal and scope*** *and* ***high degree of staff experience***.

Equally, the conceptual decision model suggested that relational data model should be adopted for a data warehouse with ***high degree of focus on generic functionalities***, ***high degree of enterprise reporting and analytics***. This study will be an invaluable resource for various implementation teams in several organisations in both industry and research sectors. Additionally, the study provides the starting point for investigation into the future role of multidimensional and relational data models in "*big data*" data warehousing.

7.3 Research Limitations

Although this research has made contributions to our understanding of why pluralistic data architectures exists in data warehousing, there are limitations that should be taken into account when interpreting the findings of this study.

First, this study is limited to two case organisations using multidimensional and relational data models for their data warehouses respectively. The research participants were mainly from the case organisations and provided the empirical data for this research. It is possible that, as the business environment and technology changes overtime, the position and views of the research participants regarding choice of their data model may change or reflect the realities of the point in time.

Second, only the technical implementation teams in the case organisations took part in the study. The end-users were not part of the study mainly because the decision regarding the choice of logical data model for a data warehouse is a technology decision. As a result, the criteria by which the user communities judge the success of their data warehouse may be different from that of the implementation teams.

Third, this research is a qualitative study, although some quantitative methods were used in analysing the empirical data; qualitative research by nature relies on interpretation of social reality (Walliman, 2006). The findings of this research were based on interpretation of the evidence from the research participants, as such, the findings of the study should be seen in that context.

Fourth, although not all the IT staff in the case organisations took part in this research, however, key decision makers from the implementation teams in the case organisations took part in the study; it is possible that IT employees that were not part of the study may have different views or opinions regarding the data model that was adopted for their data warehouse.

Fifth, the research propositions were the key factors considered to impact and influenced the choice of logical data model for a data warehouse. It is possible that this study have not accounted for all the factors that influenced the decision to adopt a logical data model in data warehousing; therefore, this provides an opportunity for further research.

7.4 Areas of Future Research

Adopting an appropriate logical data model for a data warehouse is an important decision in data warehousing, this study used the literature to distill the factors considered to be influencing the choice of logical data model for a data warehouse, however, the findings in this study indicated that while there is a significant relationship between a set of research analytic variables and the choice of a logical model (i.e. multidimensional), the study also indicated that the same variables has no significant relationship in the decision to adopt a relational data model. First, further research could be done to explore reason(s) for this outcome. Second, further research could be done to identify other factors that have not been considered by this research which may also influence the choice of logical data model in data warehousing. The outcome of such a research could be used to validate or disprove the findings of this research.

The theoretical framework for this research is decision theory, in particular, the naturalistic decision theory. In naturalistic decision making, decisions are made based on prior experience (Klein, 2008; Azuma et al, 2006). Further research could be done to examine whether decision makers consider alternative choices and engages logic of consequence when faced with decision situation such as adopting a logical data model for a data warehouse (Simon, 1995, March and Heath, 1994). Such a study could also be used as the basis to validate the findings of this research.

BIBLIOGRAPHY

Abelló, A., Samos, J. and Saltor, F., (2002), *YAM2 (Yet Another Multidimensional Model)*: An Extension of UML. In: Intl. Database Engineering & Application Symposium (IDEAS'02), Edmonton: Canada, 172-181

Adamson, C., (2006), *Mastering Data Warehouse Aggregates: Solutions for Star Performance*, Crosspoint Boulevard: Indianapolis, Wiley

Adelman, S., Moss, L. and Abai, M., (2005), *Data Strategy*, Upper Saddle River: New Jersey, Pearsons Education, Inc.

Agrawal, R., Gupta, A. and Sarawagi, A., (1997), *Modeling Multidimensional Databases*, Proceedings of the Thirteenth International Conference on Data Engineering, p.232-243, April 07-11, 1997

Allison, G.T., (1971), *Essence of Decision: Explaining the Cuban Missile Crisis*, Little, Brown

Alter, S., (2002), *A work System View of DSS in its Fourth Decade*, Eight Americas Conference on Information Systems, pp. 150-151

Antón, A., I. and Potts, C., (1998), *The use of goals to surface requirements for evolving systems*, In Proceedings of the 20th international conference on Software engineering, p.157-166, April 19-25, 1998, Kyoto: Japan

Ariyachandra, T. and Watson, H., (2006), *Which Data Warehouse Architecture Is Most Successful*, Business Intelligence Journal, Vol.11, No.1

ASQ, (2014), *Decision Matrix*, Available from: http://asq.org/learn-about-quality/decision-making-tools/overview/decision-matrix.html [Date Accessed: 11 December 2014]

Azuma, R., Daily, M., Furmanski, C., (2006), *A Review of Time Critical Decision Making Models and Human Cognitive Processes*, Proc. 2006 IEEE Aerospace Conference (Big Sky, MT, 4-11 March 2006).

Babbie, E. R., (2010), *The Practise of Social Research*, Wadsworth Cengage Learning, Belmont, CA

Baron, J. (2004), *Normative models of judgement and decision making*. In D. J. Koehler & N. Harvey (Eds.), Blackwell Handbook of Judgement and Decision Making, pp. 19–36. London: Blackwell

Barlow, S. (2013), *6 Reasons Why Healthcare Data Warehouses Fail*, https://www.healthcatalyst.com/why-healthcare-data-warehouses-fail/, Date Accessed: 11/8/14

Beinum, H.V., (1999), *On the Design of the ACRES Program*. In:Greenwood, Davydd, J., (1999), (eds) *Action Research: From practise to writing in an international action research development program*, Amsterdam, The Netherlands: John Benjamins Publishing Company. P.3 – 24

Bell, D.E., Raiffa, H., Tversky, A., (1988), *Decision Making: Descriptive, Normative and Prescriptive Interactions*, Cambridge University Press

Bentayeb, F., Favre, C. and Boussaid, O., (2008), *A user-driven data warehouse evolution approach for concurrent personalized analysis needs*, Integrated Computer-Aided Engineering, v.15 n.1, p.21-36

Bernard, R. H., (2000), *Social Science Methods: Qualitative and Quantitative Approaches*, Thousand Oaks, CA: Sage Publications, Inc

Bernard, R. H, (2006), *Research Methods in Anthropology: Qualitative and Quantitative Approaches*, 4th Edition, Lanham, MD: AltaMira, Press

Boeije, Hennie, (2010), *Analysis in Qualitative Research*, Thousand Oak, CA: Sage Publishing

Böhnlein, M. and Ulbrich-vom Ende, A., (2000), *Business Process Oriented Development of Data Warehouse Structures*, Data Warehousing 2000 – Methoden, Anwendungen, Strategien; Physica: Heidelberg

Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A. and Papaboschi, S., (2001), *Designing data marts for data warehouses*, In ACM Transactions on Software Engineering and Methodology (TOSEM), v.10 n.4, p.452-483, Oct. 2001

Bordens, K. S. and Horowitz, I. A., (2002), *Social Psychology*, Mahwah, NJ: Lawrence Erlbaum Associate, Inc. Publishers

Border, K.C., (1984), *Aspects of Normative Decision Theory, Division of Humanities and Social Sciences*, California Institute of Technology, California

Brennan, P.F., (1995), *Patient Satisfaction and Normative Decision Theory*, Journal of the American Medical Informatics Association, Vol. 2, No 4, July/August

Brooks, C., (2014), *What Is a Decision Matrix*. Available from: http://www.businessnewsdaily.com/6146-decision-matrix.html [Date Accessed: 29 November 2014]

Brown, C. and Lloyd, K., (2001), *Qualitative methods in psychiatric research*, Advances in Psychiatric Treatment, Vol. 7

Bruckner, R.M., List, B. and Schiefer, J., (2001), *Developing Requirements for Data Warehouse Systems with Use Cases*, In Proceedings of the Seventh Americas Conference on Information Systems (AMCIS 2001), pp. 329-335; Boston, Massachusetts, USA, August 2001

Bryman, Alan, (1989), *Research Methods and Organisation Studies*, New York, NY: Routledge

Bryman, A. and Bell, E., (2007), *Research Methods*, Second Edition, New York, Oxford University Press Inc.

Bryne, M.M., (2001), *Evaluating the findings of qualitative research*, AORN Journal In: HighBeam Research Inc., Farmington Hills, Michigan: Gale Group

Bush, V., (1945), *As We May Think*, Atlantic Monthly, July, http://www.theatlantic.com/doc/194507/bush, Date Accessed: 15/10/2009

Cabibbo, L., and Torlone, R., (1998), *A Logical Approach to Multidimensional Databases*, In Proceedings of the 6th International Conference on Extending Database Technology: Advances in Database Technology, p.183-197, March 23-27, 1998, Volume 1377 of LNCS., Valencia: Spain

Caserta, J. and Kimball, R., 2004, *The Data Warehouse ETL Toolkit, Practical Techniques for Extracting, Cleaning, Conforming and Delivery Data*, Indianapolis, IN: Wiley Publishing

Cavaye, A.L.M., (1996), *Case study research: a multi-faceted research approach for IS*, Information Systems Journal, 6(3)

Charmaz, Kathy, (2006), *Constructing Grounded Theory*, London: Sage Publications Ltd.

Chaudhuri, S. and Dayal, U., (1997), *An Overview of data warehousing and OLAP technology*, ACM SIGMOD, Volume 26, Issue 1, March 1997, p.65-74

Clark, A.M., (1998), *The qualitative-quantitative debate: moving from positivism and confrontation to post-positivism and reconciliation*, Blackwell Science Ltd., Journal of Advanced Nursing, 27, 1242-1249

Crossan, F., (2003), *Research Philosophy: towards an understanding*, Nurse Researcher, Vol. 11, No. 1

Culnan, M.J. and Swanson, E.B., (1978), *Document-Based Systems for Management Planning and Control: A Classification, Survey, and Assessment*, MIS Quarterly, December

Czaja, R. and Blair, J., (2005), *Designing Surveys: A Guide to Decisions and Procedures*, Second Edition, Thousand Oaks, CA: Sage Publications, Inc

Devlin, B.A. and Murphy, P.T., (1988), *An architecture for a business and information systems*, IBM Systems Journal. Volume 27, No. 1, p.60

DeWitt, D. J. D., Samuel Madden, and Michael Stonebraker, (2008), *How to build a high-performance data warehouse*. Available from: http://db. lcs. mit. edu/madden/high_perf. pdf [Date Accessed: 07 December 2009]

Dhar, V. and Stein, R., (1997), *Intelligent Decision Support Methods*, Upper Saddle river, NJ: Prentice Hall

Dillon, S.M., (1998), *Descriptive Decision Making: Comparing Theory with Practice*, In Proceedings of 33rd conference of the Operational Research Society of New Zealand

Dilthey, W. (1977). *Die Wissenschaften vom Menschen, der Gesellschaft und der Geschichte* (Gesammelte Schriften XVIII. Band) [Human, social, and historical sciences (Collected works: Volume 18)]. Göttingen: Vandenhoeck & Ruprecht

Drewek, K., (2005), *Data Warehousing: Similarities and Differences of Inmon and Kimball*. B-EYE-NETWORK. Available from: http://www.b-eye-network.com/view/743 [Date Accessed: 07 March 2010]

Druzdzel, M.J., Flynn, R.R., (2000), *Decision Support Systems*,
In: Kent, A. (ed.), Encyclopedia of Library and Information Science,
Vol.67 (30), p.120-133, New York: Marcel Dekker, Inc

Dul, J. and Hak, T., (2008), *Case Study Methodology in Business Research*, First Edition, Oxford, Elsevier Publishing

Eisenhardt, K.M., (1989), *Building Theories From Case Study Research*, The Academy of Management Review, Vol. 14, Issue 4

Eom, S. B., (2001), *Decision Support Systems*, *International Encyclopedia of Business and Management*, 2nd Edition, Edited by Malcom Warner, London: International Thomson Business Publishing Company

Firestone, J.M., (2002), *Data Warehouses, Data Marts, and Data Warehousing: New Definitions and New Conceptions*, Vol. 6, DKMS Brief No. Six, http://www.dkms.com/papers/dwdmed.pdf, Date Accessed: 19/10/2009

Forgionne, G., Moral, M. and Gupta, J.N.D., (2003), *Decision Making Support Systems, Achievements and Challenges for the New Decade*, Hershey PA: Idea Group Publishing

Franconi, E. and Sattler, U., (1999), *A data warehouse conceptual data model for multidimensional aggregation*, In Proceedings of the Workshop on Design and Management of Data Warehouses (DMDW '99), June, Heidelberg: Germany

Galliers R.D. and Land, F.F., (1987), *Viewpoint: Choosing Appropriate Information Systems Research Methodologies*, Communications of the ACM, vol. 30, No. 11, pp. 901-902

Gartner, (2005), *More Than 50 Percent of Data Warehouse Projects Will Have Limited Acceptance or Will Be Failures Through 2007*, Available From: http://www.gartner.com/it/page.jsp?id=492112, [Date Accessed: 07 March 2010]

Gerring, John, (2007), *Case Study Research: Principles and Practices*, New York, NY: Cambridge University Press

Gibbs, R., Graham, (2007), *Analyzing Qualitative Data*, London: Sage Publications

Gillham, B., (2000), *Case study research methods*, London: MPG Books Ltd.

Giorgini, P., Rizzi, S. and Garzetti, M, (2008), *A goal-oriented approach to requirement analysis in data warehouses, Decision Support Systems*, DOLAP'05, November 4-5, v45 n.1, p.4-21, April, 2008, Bremen: Germany

Golfarelli, M., Maio, D. and Rizzi, S., (1998), *The Dimensional Fact Model: a conceptual model for data warehouses*, International Journal of Cooperative Information Systems, V7, p.215-247

Glaser, B.G. and Strauss, A. L., (2009), *The Discovery of Grounded Theory: Strategies for qualitative research*, Piscataway, NJ: Transaction Publishers

Golfarelli, M., Maio, D. and Rizzi, S., (1998), *The Dimensional Fact Model: a conceptual model for data warehouses*, International Journal of Co-operative Information Systems, V7, p.215-247

Golfarelli, M. and Rizzi, S., (1998), *A methodological framework for data warehouse design*, In Proceedings of the 1st ACM international workshop on Data warehousing and OLAP (DOLAP 98), Nov 7, 1998, Washington, D.C.

Gravetter, F. J. and Forzano, Lori-Ann, B., (2009), *Research Methods for the Behavioural Sciences*, 3rd Edition, Bellmont, CA: Wadsworth Cengage Learning

Green, J.L., Patricia, C.C. and Elmore, P.B., (2006), *Handbook of complementary methods in education research: American Education Research Association*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc

Gregor, S. (2002), *A theory of theories in information systems*. In S. Gregor and D. Hart (Eds.), Information Systems Foundation: Building the Theoretical Base. Australian National University, Canberra, pp.1-20

Gregor, S. (2006), *The Nature of Theory in Information Systems*, Management information Systems, Quarterly, vol. 30, No. 3, pp. 611 - 642

Grobelnik, M., and Lavrac, N., (2003), *Data Mining and Decision Support: Integration and Collaboration*, Norwell, Massachusetts, Kluwer Academic Publishers

Guba, E.G., (1990), *The Alternative Paradigm Dialog* In: Guba, E.G., (1990), (eds) *The Paradigm Dialog*, Newbury Park, CA: Sage Publications, Inc. P.17 – 27

Gutierrez, A. and Marotta, A., (2000), *An Overview of Data Warehouse Design Approaches and Techniques*, White paper, Instituto de Computacion, Universidad de la Republica, Montevideo, Uruguay

Hammersley, M., (1993), *Social Research: Philosophy, Politics and Practice,* London, Sage Publishing

Hancock, D. R. and Bob Algozzine, (2006), *Doing Case Study Research: A Practical Guide for Beginning Researchers*, New York, NY: Teachers College Press

Hansson, S.O., (1994), *Decision Theory: A Brief Introduction*, Stockholm

Harvey, N., (2004), *Blackwell handbook of judgment and decision making*, Oxon, Blackwell Publishing

Hans Van Beinum, (1999), *On the Design of the ACRES Program*. In:
Greenwood, Davydd, J., (1999), (eds) *Action Research: From practise to writing in an international action research development program*, Amsterdam, The Netherlands: John Benjamins Publishing Company. P.3 - 24

Hashmi, N., (2000), *Business Information Warehouse for SAP*, Roseville, CA: Prima Publishing

Hassan, Q.U., Spector, M.J., P.I. Davidsen, (2008), *Complex decision making; theory and practice*, New York, Springer

Hayen, R., L., Rutashobya , C., D. and Vetter, D., E., (2007), *An Investigation of Factors affecting Data Warehousing Success*, *Issues in Information Systems*, Vol. VIII, No. 2, pg. 547-553

Herden, O., (2000), *A Design Methodology for Data Warehouses*, 7th Doctoral Consortium on Advanced Information Systems and Engineering (CAiSE'00), Sweden

Hidalgo, F., & Sekhon, J. (2011), *Causality*. In B. Badie, D. Berg-Schlosser, & L. Morlino (Eds.), International encyclopedia of political science. (pp. 204-211). Thousand Oaks, CA: SAGE Publications, Inc

Holsapple, C.W., (2008), *DSS Architecture and Types*, Handbook on Decision Support Systems Vol 1, Chapter 9, Editors (Burstein, F. and Holsapple, C.W.), Berlin: Springer

Holsapple, C.W. and Whinston, A. B., (1996), *Decision Support Systems: A Knowledge-based Approach*, Minneapolis, MN: West Publishing Company

Horkheimer, M., (2002), *Critical Theory, Selected Essays*, Lexington Avenue, NY:, The Continuum Publishing Company

Humphries, M., Hawkins, M., W., and Dy, M., C, (1999), *Data Warehousing: Architecture and Implementation*, Upper Saddle River: New Jersey, Prentice Hall PTR

Hurtado, C.A. and Mendelzon, A., O., (2004), *OLAP Dimension Constraints*. In: Proc. ACM PODS, Pages 169-179

Husemann, B., Lechtenbörger, G. and Vossen, G., (2002), *Conceptual Data Warehouse Design*. In Proceedings of the International Workshop on Design and Management of Data Warehouses, June 5-6, Stockholm, Sweden

Imhoff, C., Galemmo, N. and Geiger, J. G., (2003), *Mastering Data Warehouse Design: Relational and Dimensional Techniques*, Indianapolis: Indiana, Wiley Publishing

Inmon, B., 2009, *The Elusive Data Warehouse*, March 19, Online Article, http://www.b-eye-network.com/view/9956/, Date Accessed: 11/11/09

Inmon, W.H., (2005), *Building the Data Warehouse*, Fourth Edition, Indianapolis: IN, Wiley Publishing, Inc

Inmon, W.H., Imhoff, C. and Sousa, R., (2001), *Corporate Information Factory*, Second Edition, New York: New York, John Wiley & Sons Inc

Inmon, W.H., Strauss, D. and Neushloss, G., (2008), DW 2.0: *The Architecture for the Next Generation of Data Warehousing*, Burlington: MA, Morgan Kaufmann Publishers

Jerome, K., (1992), *Managing with Information*, 4th Edition, Englewood Clifs, New Jersey: Prentice Hall.

Johnson, J.G., Busemeyer, J.R., (2005), *Rule-Based Decision Field Theory: A Dynamic Computational Model of Transitions Among Decision-Making Strategies* In: Tilman, B. and Haberstroch, S. (eds.) The Routines of Decision Making. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. pp. 3-20

Johnson J.G., Busemeyer, J.R., (2010), *Decision making under Risk and Uncertainty*, WIREs Cognitive Science, Vol 1, September/October, John Wiley & Sons, Ltd. pp 736-749

Jones, B.D., (1999), *Bounded Rationality, Annual Review of Political Science*, Vol. 2 pp 297-321

Jones, M. and Alony, I., (2011), *Guiding the Use of Grounded Theory in Doctoral Studies – An Example from the Australian Film Industry*, International Journal of Doctoral Studies, Vol. 6

Jones, C. H. and Ferguson, R.L., (1969), *A Computer Aided Decision System*, Management Science, Vol. 15, No.10, June, pp. B550 – B561

Jorgensen, D.L., (1989), *Participant Observation: A Methodology for Human Studies*, Applied Social Research Methods Series, Volume 15, Thousand Oaks, CA: Sage Publications, Inc

Kaplan, B and Duchon, D., (1998), *Combining qualitative and quantitative methods in information systems research: a case study*, MIS Quarterly, 12(4)

Khan, A., (2005), *SAP and BW Data Warehousing: How to Plan and Implement*, Lincoln, NE: iUniverse Publishing

Kimball, R. and Ross, M., (2002), *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modelling*, New York: Wiley Publishing

Kimball, R. and Ross, M., (2004), *Differences of Opinion, The Kimball Bus Architecture and Corporate Information Factory: What are the fundamental differences?*, Information Week, March 6, http://www.informationweek.com/news/showArticle.jhtml?articleID=17800088, [Date Accessed: 30th March 2011]

Kimball, R., (1995), *Is ER Modeling Hazardous to DSS?*, DBMS, October

Kimball, R., (1997), *A Dimensional Modeling Manifesto: Drawing the Line Between Dimensional Modeling and ER Modeling Techniques*, DBMS, August

Klien, G., (2008), *Naturalistic Decision Making, Human Factors*, Vol. 50, No. 3, June 2008, pp. 456-460

Kotler, P., Armstrong, G., Wong, V. and Sauders, J., (2008), *Principles of Marketing*, Fifth European Edition, Harlow, Essex: Pearsons Education Limited

Kunreuther, H., Meyer, R., Zeckhauser, R., Slovic, P., Schwartz, B., Schade, C., Luce, M.F., Lippman, S., Krantz, D., Kahn, B., and Hogarth, R., (2001), *High Stakes Decision Making: Normative, Descriptive, Prescriptive and Considerations*, http://marketing.wharton.upenn.edu/documents/research/High_Stakes_Decision_Making.pdf, [Date Assessed: 27th November 2011]

Lechtenbörger, J., (2001), *Data Warehouse Schema Design*, http://dbms.uni-muenster.de/dbms/media/publications/2003/diss-jl-btw2003.pdf, [Date Accessed 29th August 2009]

Lechtenbörger, A. Vossen, G., (2003), *Multidimensional Normal Forms for Data Warehouse Design*, Information Systems, 28(5), 415-434

Leigh, W.E. and Doherty, M.E., (1986), *Decision Support and Expert Systems*, Cincinnati: South-Western Publishing Company.

Levene, M. and Loizou, G., (2003), *Why is the snowflake schema a good data warehouse design, Information Systems,* Vol. 28, issue 3, pg225 – 240

Liang, T.P., (1988), *Development of a Knowledge Based Model Management System,* Operations Research, Vol. 36, No. 6

Liang, T.P., Lee, C.C., and Turban, E., (2008), *Model Management and Solvers for Decision Support*, Handbook on Decision Support Systems Vol 1, Chapter 12, Editors (Burstein, F. and Holsapple, C.W.), Berlin: Springer

Lincoln, Y. S., and Guba, E. G., (1985), *Naturalistic Inquiry*, Beverly Hills: Sage Publication

Linstedt, D., (2004), *Data Vault Model*, http://www.danlinstedt.com/AboutDV.php, Date accessed 20/08/09.

Lisl, Zach, (2006), *Using a multiple-case studies design to investigate the information-seeking behavior of arts administrators*, Library Trends, 22(6)

List, B., Bruckner, R.M., Machaczek, K. and Schiefer, J., (2002), *A Comparison of Data Warehouse Development Methodologies Case Study of the Process Warehouse*, In Proceedings of the 13th International Conference on Database and Expert Systems Applications, 02-06 September 2002, Heidenberg: Berlin

Loshin, D., (2009), *Master Data Management*, Burlington, MA: Elsevier

Luján-Mora, S. and Trujillo, J., (2003), *A Comprehensive Method for Data Warehouse Design*, In Proceedings DMDW, Berlin: Germany

Malaxa, V. and Douglas, I., (2005), *A Framework for Metadata Creation Tools*, Interdisciplinary Journal of Knowledge and Learning Objects, Vol. 1

Mallach, E. G., (2000), *Decision Support and Data Warehouse Systems*, International Edition, McGraw-Hill Higher Education, Singapore

March, J.G., Heath, C., (1994), *A primer on decision making: how decisions happens*, New York, New York, The Free Press

Marco, D., (2000), *Building and Managing the Meta Data Repository: A Full Lifecycle Guide*, New York: New York, John Wiley & Sons, Inc

Marco, D. and Jennings, M., 2004, *Universal Meta Data Models*, Indianapolis, IN: Wiley Publishing

Mark, M., (2005), *Bounded Rationality: Analysis of architecture, Integration, BI & IT Strategy*, http://it.toolbox.com/blogs/bounded-rationality/a-50-data-warehouse-failure-rate-is-nothing-new-4669, [Date Accessed, 30th March 2011]

Marschan-Piekkari, R and Welch, C., (2004), *Handbook of Qualitative Research Methods for International Business*, Cheltenham, Gloucester: Edward Elgar Publishing Limited

Mayers, M.D., (1997), *Qualitative Research in Information Systems,* MIS Quarterly, Vol. 21, Issue 2

Mazon, J., Trujillo, J., Serrano, M. and Piattini, M., (2005), *Applying MDA to the Development of Data Warehouses*, In Proceedings of the 8th ACM international workshop on Data warehousing and OLAP, Bremen, Germany, Pages, 57-66

McCoy, E.D., Shrader-Frechette, K. S., (1993), *Method in Ecology: Strategy for Conservation*, Cambridge, England: Cambridge University Press

McNiff, J., Lomax, P. and Whitehead, J., (2003), *You and Your Action Research Project*, Abingdon, Oxon: RoutledgeFalmer

Mertler, C.A., (2006), *Action Research: Teachers as Researchers in the Classroom*, Thousand Oaks, CA: Sage Publications, Inc

Miles, M. B. and Huberman, A. M., (1994), *Qualitative Data Analysis: An Expanded Sourcebook*, Thousand Oak, CA: Sage Publishing

Mills, A.J., Wiebe, E. and Durepos, G., (2010), *Encyclopedia of Case Study Research*, London: Sage Publications Ltd.

Mingers, J., (2001), *Combining IS Research Methods: Towards a Pluralist Methodology*, Information Systems Research, 12 (3)

Moody, D. L. and Kortink, M.A.R., (2000), *From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design*, In, Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW), June 5-6, Stockholm: Sweden

Mrunalini, M. Kumar, T.V.S., Geetha, D.E. and Rajanikanth, K., (2006), *Modelling of Data Extraction in ETL Processes Using UML 2.0*, DESIDOC, Journal of Library & Information Technology, Vol. 26, No 5

Newman, I. and McNeil, K.A., (1998), *Conducting Survey Research in the Social Sciences*, Lanham, Maryland: University Press of America, Inc

NISO, (2004), *Understanding Metadata*, National Information Standards Organisation, Bethesda: Maryland, NISO Press

O' Sullivan, O., (1996), *Data Warehousing-without the Warehouse*, ABA Banking Journal, 88 (12)

Orlikowski, W.J. and Baroudi, J.J., (1991), *Studying Information Technology in Organisations: Research Approaches and Assumptions*, Information Systems Working Papers Series

Paim, F. R. S. and Castro, J.F.B., (2003), *DWARF: AN Approach for Requirements Definition and Management of Data Warehouse Systems*, In: 11th IEEE International Requirements Engineering Conference (RE'03), September 08-September 12, Monterey Bay, California, USA

Ponniah, P., (2001), *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals*, 605 Third Avenue, New York, New York: John Wiley & Sons Inc.

Power, D., (2001), *What is data Mining and How is it related to DSS*, DSS News, Vol. 2, No. 25, December 2, http://dssresources.com/newsletters/42.php, [Date Accessed: 19/10/2009]

Power, D. J., (2007), *A Brief History of Decision Support Systems*, DSSResources.COM, http://DSSResources.COM/history/dsshistory.html, version 4.0, March 10

Power, D. J., (2008), *Decision Support Systems: A Historic Overview*, Handbook on Decision Support Systems Vol. 1, Chapter 7, Editors (Burstein, F. and Holsapple, C.W.), Berlin: Springer

Power, D. J., (2008), *Handbook on Decision Support Systems* Vol. 1, Chapter 7, Decision Support Systems: A Historical Overview, Editors (Burstein, F and Holsapple C.W.), Berlin: Springer

Prakash, N. and Gosain, A., (2003), *Requirements Driven Data Warehouse Development*, In CaiSE Short Paper Proc.

Punch, K., (2006), *Developing Effective Research Proposals*, Second Edition, London, Sage Publishing

Rainardi, V., (2008), *Building a Data Warehouse: With Examples in SQL Server*, New York: NY, Apress Publishers

Reeves, L., Ross, M., Thornthwaite, W. and Kimball, R., (1998), *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses*, 605 Third Avenue, New York, New York: John Wiley & Sons Inc.

Riabacke, A., (2006), *Managerial Decision Making Under Risk and Uncertainty*, IAENG International Journal of Computer Science, Vol.32 No 4, February

Rizzi, S., Abelló, A., Lechtenbörger, J. and Trujillo, J., (2006), *Research in Data Warehouse Modeling and Design: Dead or Alive*, In: Proceedings of the 9th ACM international workshop on Data warehousing and OLAP, Arlington: Virginia, 3-10

Rouse, M., (2012), Shared Services, http://searchcio.techtarget.com/definition/Shared-services, [Date Accessed, 30th May 2014]

Rubin, H.J. and Rubin, I.S., (2005), *Qualitative Interviewing, The Art of Hearing Data*, Second Edition, Thousand Oaks, CA: Sage Publications, Inc

Russom, Phillip, (2009), *Next generation data warehouse platforms*, TDWI Best Practices Report, Fourth Quarter

Ryan, A. B., (2006), *Post-Positivists Approaches to Research*. In: Researching and Writing your thesis: a guide for postgraduate students. MACE: Maynooth Adult and Community Education

Saldana, J., (2009), *The Coding Manual for Qualitative Researchers*, London, Sage Publishing, p. 3

Sapia, C., Blaschka, M., Hoefling, G. and Dinter, B., (1998), *Extending the E/R Model for the Multidimensional Paradigm*, In Proc. International Workshop on Data Warehouse and Data Mining, (DWDM '98), Vol. 1552 of LNCS., Singapore

Sarkar, P., (2002), *Information Quality in the Knowledge Driven Enterprise*, Journal, Information Management Direct, November

Saunders, M., Thornhill, A. and Lewis, P., (2009), *Research Methods for Business Students*, 5th Edition, Harlow, Essex: Pearson Education Limited

Seidel, J. V., (1998), *Qualitative Data Analysis*, Qualis Research, www.qualisresearch.com, Date Assessed: 14/10/2010

Sen, A. and Sinha, A., (2005), *A comparison of Data Warehouse Methodologies*, Communications of the ACM, Vol. 48, No. 3

Shanks G., (2002), *Guidelines for Conducting Positivist Case Study Research in Information Systems*, AJ IS, December, Vol.10 No 1

Siddiqi, N., (2006), *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, Hoboken, N.J.: John Wiley and Sons

Simons Helen, (2009), *Case Study Research in Practice*, City Road, London: Sage Publications Ltd

Simon, H.A., (1955), *A Behavioural Model of Rational Choice, Quarterly Journal of Economics*, Vol. 69, pp.99-118

Sperley, E., (1999), *Enterprise Data Warehouse, Planning, Building, and Implementation*, Upper Saddle River: New Jersey, Prentice Hall PTR

Speziale, Streubert, H.J. and Carpenter, D. R., (2007), *Qualitative Research in Nursing: Advancing the humanistic Imperative*, 4th Edition, Philadelphia, PA: Lippincott, Williams & Wilkins

Stake, R. E., (1995), *The Art of Case Study Research*, Thousand Oaks, CA: Sage Publications, Inc

Stockburger, D. W., (1996), *Introductory Statistics: Concepts, Models, and Applications*, http://www.psychstat.missouristate.edu/introbook/sbk28m.htm, [Date Accessed, 27th November 2014]

Strauss A. L., and Corbin, J.M., (1997), *Grounded Theory in Practise*, Thousand Oaks, CA: Sage Publications, Inc

Sultan, F. and Aziz, A., (2010), *Ideal Strategy to Improve Data Warehouse Performance*, International Journal on Computer science and Engineering, Vol.2, No.2, pg. 409-415

Teory, T.J., Yang, D. and Fry, J.P., (1986), *A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model*, Computing Survey, Vol. 18, No. 2

Theodoratos, D. and K. S., Timos, (1997), *Data Warehouse Configuration*, Proceedings of the 23rd International Conference on Very Large Data Bases, p.126-135, August 25-29

Thrash, W., (2009), *Enterprise Data Warehouse or Star Schema?*, Available from: http://willthrash.wordpress.com/2009/04/29/enterprise-data-warehouse-or-star-schema-aka-inmon-vs-kimball [Accessed 07 March 2010]

Todman, C., (2001), *Designing a Data Warehouse: Supporting Customer Relationship Management*, Upper Saddle River, NJ: Prentice Hall PTR

Torraco, R. J., (2005), *Theory Development Research Methods*, The BK Whitepaper Series, San Francisco, CA: Berrett-Koehler Publishers, Inc.

Trujillo, J., Palomar, M., Gomez, J. and Song, I., (2001), *Designing Data Warehouses with OO Conceptual Models*, IEEE, Special Issue on Data Warehouses, 66-75

Tryfona, N., Busborg, F. and Christiansen, J.G.B., (1999), *starER: a conceptual model for data warehouse design*, In Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, p.3-8, November 02-06, 1999, Kansas City: Missouri

Turban, E., Aronson, J.E., Liang, T., and Sharda, R., (2007), *Decision Support and Business Intelligence Systems*, Eight Edition, Upper Saddle River, New Jersey: Pearson Publishing.

Watson, J. and Ariyachandra, T., (2005), *Factors in the Selection Decision and the Success of the Architectures*, http://www.uga.edu/~hwatson/DW_Architecture_Report.pdf, Date accessed 11/07/09

White, C., 2000, *The Federated Data Warehouse*, March, Information Management Magazine, http://www.information-management.com/issues/20000301/1953-1.html, Date Accessed 12/11/09

Vassiliadis, P. (2004), *Gulliver in the land of data warehousing: practical experiences and observations of a researcher*, Journal of Data Warehousing, 2(3), pp. 1 – 16.

Vassiliadis, P., Simitsis, A. and Baikousi, E., (2009), *A Taxonomies of ETL Activities*, DOLAP' 09, November 6, 2009, Hong Kong: China

Voss, C., Tsikriktsis, N. and Frohlich, M., (2002), *Case research in operations management*, International Journal of Operations & Product Management, Vol. 22, No. 22, pp. 195-219

Walliman, Nicholas, S. R., (2006), *Social Research Methods*, London: Sage Publications Ltd.

Weininger, A., (2002), *Efficient Execution of Joins in a Star Schema*, ACM SIGMOD, June 4-6, Madison, Wisconsin

Whitehorn, M.and Burns, K., (2008), *Best Practices for Data Warehousing with SQL Server 2008*, Available from: http://msdn.microsoft.com/en-us/library/cc719165(v=sql.100).aspx#feedback [Date Accessed: 25 November 2011]

Winter, R. and Strauch, B., (2003), *A Method for Demand-Driven Information Requirements Analysis in Data Warehousing Projects,* In Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track 8 - Volume 8, Page: 231.1

Wixom, B. H. and Watson, H. J., (2001), *An Empirical Factors Affecting Data Warehousing Success*, MIS Quarterly, Vol.25, No. 1, pp. 17-41

Wong, R., K. and Lam, F., (2001), *Modelling and Manipulating Multidimensional Data in Semi-Structured Databases*, World Wide Web, 4: 79-99, Kluwer Academic Publishers

Yin, R. K., (2003), *Application of Case Study Research*, Second Edition, Thousands Oak, Sage Inc

Yin, R. K., (2009), *Case Study Research, Design and Methods*, 4th Edition, Thousands Oak, Sage Inc

Youness, S., (2000) *Professional Data Warehousing with SQL Server 7.0 and OLAP Services*, Birmingham, England: Wrox Press

APPENDIX 1 – DESIGN OF RESEARCH INSTRUMENTS

The propositions outlined below are pursued in this research and provided the background for the design of interview instruments for this research

Abbreviated	Research Propositions - Factors affecting the choice of logical data model for a data warehouse
RP1	High performance of query execution is a factor influencing the choice of logical data model for a data warehouse
RP2	Focus on specific or generic business requirements is a factor influencing the choice of logical data model for a data warehouse
RP3	The goal and scope of the data warehouse is a factor influencing the choice of logical data model for a data warehouse
RP4	Implementation orientation of available resources is a factor influencing the choice of logical data model for a data warehouse
RP5	Consolidation of enterprise data to meet heterogeneous reporting and analytics requirements is a factor influencing the choice of logical data model for a data warehouse

Design of Interview Instruments

The main questions outlined below are engaged to address this research. Each of the research questions aligns with the research propositions above and provides the background for the design of the interview instruments for this research.

Main Research Questions for the Research Participants

MRQ1 – What is the impact of high performance of query execution on the choice of logical data model for a data warehouse?

MRQ2 - What is the impact of specific business requirement on the choice of a logical data model?

MRQ3 – What is the impact of goal and scope of a data warehouse on the choice of data warehouse logical data model?

MRQ4 - What is the impact of implementation orientation of available resources on the choice of a logical data model?

MRQ5 – What is the impact of consolidating common enterprise data to meet different reporting and analytics requirements on the choice of logical data model for a data warehouse?

The main research questions outlined above align with the research propositions and thus provide a way of testing the research propositions. Also, the questions outlined below are grouped to address each of the conversation partners given their knowledge, experience and stake in their organisation's data warehouse initiatives. The questions are further re-aligned and grouped under the relevant research main question.

APPENDIX 2 – DESIGN OF RESEARCH INSTRUMENT FOR GWEALTH

Questions for the GWealth Research Participants

GWRQ1 – What are the problems that the business wants solved that prompted the development of a data warehouse?

GWRQ2 - How is the need to address the business requirements led to choosing a multidimensional data model for the reporting data warehouse?

GWRQ3 - What role did the goal of the reporting data warehouse had in choosing the multidimensional data model for the reporting data warehouse?

GWRQ4 - What role did the scope of the reporting data warehouse had in choosing the multidimensional data model for the reporting data warehouse?

GWRQ5 - What is the end-user ability to understand the data relationships in client reporting data warehouse?

GWRQ6 - What role would you say the staff experience played in choosing the multidimensional data model for the reporting data warehouse?

GWRQ7 – Would you agree or disagree that your staff experience influenced the choice of the data model for your data warehouse?

GWRQ8 - What other data models would you say were considered for the reporting data warehouse by your development team?

GWRQ9 – In what circumstances would your team consider to use different data models than one currently used for client reporting data warehouse?

GWRQ10 - What is the role of performance in choosing the multidimensional data model for client reporting data warehouse?

GWRQ11 - What is the impact of ensuring high performance of query execution on the choice of the multidimensional data model for the reporting data warehouse?

GWRQ12 – What are the non-functional factors that influenced the decision to adopt the multidimensional data model for the reporting data warehouse?

GWRQ13 – How would you describe the users response of executing queries against the reporting data warehouse?

GWRQ14 - In your opinion, what has been the users' experience in analyzing and querying the reporting data warehouse?

GWRQ15 - How diverse is the user base of your data warehouse?

GWRQ16 - How often are different areas of the business on-boarded into the reporting data warehouse?

GWRQ17 - How many areas of the business have been on-boarded onto the reporting data warehouse?

GWRQ18 - How would you describe the data coverage of the reporting data warehouse?

GWRQ19 - How would you describe the level of complexity of business data in the reporting data warehouse?

GWRQ20 - What level of assistance does the architecture team have to offer the business users when looking for information in the client reporting data warehouse?

GWRQ21 – What is your opinion of whether the business users have indicated they found it difficult to navigate the reporting data warehouse?

GWRQ22 - How much weight do you think was given to distributing data to different areas of the business by client reporting data warehouse?

GWRQ23 - How would you characterise the advantages of reporting warehouse data model?

Alignment of the GWealth Research Questions with the Main Research Questions

MRQ1 – What is the impact of high performance of query execution on the choice of logical data model for a data warehouse?

ID	Question
GWRQ10	What is the role of performance in choosing the multidimensional data model for client reporting data warehouse?
GWRQ11	What is the impact of ensuring high performance of query execution on the choice of the multidimensional data model for the reporting data warehouse?
GWRQ12	What are the non-functional factors that influenced the decision to adopt the multidimensional data model for the reporting data warehouse?
GWRQ13	How would you describe the users response of executing queries against the reporting data warehouse?
GWRQ14	In your opinion, what has been the users' experience in analyzing and querying the reporting data warehouse?
GWRQ23	How would you characterise the advantages of reporting warehouse data model?

MRQ2 - What is the impact of specific business requirement on the choice of a logical data model?

ID	Question
GWRQ1	What are the problems that the business wants solved that prompted the development of a data warehouse?
GWRQ2	How is the need to address the business requirements led to choosing a multidimensional data model for the reporting data warehouse?

MRQ3 – What is the impact of goal and scope of a data warehouse on the choice of data warehouse logical data model?

ID	Question
GWRQ3	What role did the goal of the reporting data warehouse had in choosing the multidimensional data model for the reporting data warehouse?
GWRQ4	What role did the scope of the reporting data warehouse had in choosing the multidimensional data model for the reporting data warehouse?
GWRQ5	What is the end-user ability to understand the data relationships in client reporting data warehouse?

MRQ4 - What is the impact of implementation orientation of available resources on the choice of a logical data model?

ID	Question
GWRQ6	What role would you say the staff experience played in choosing the multidimensional data model for the reporting data warehouse?
GWRQ7	Would you agree or disagree that your staff experience influenced the choice of the data model for your data warehouse?
GWRQ8	What other data models would you say were considered for the reporting data warehouse by your development team?
GWRQ9	In what circumstances would your team consider to use different data models than one currently used for client reporting data warehouse?

MRQ5 – What is the impact of consolidating common enterprise data to meet different reporting and analytics requirements on the choice of logical data model for a data warehouse?

ID	Question
GWRQ15	How diverse is the user base of your data warehouse?
GWRQ16	How often are different areas of the business on-boarded into the reporting data warehouse?
GWRQ17	How many areas of the business have been on-boarded onto the reporting data warehouse?
GWRQ18	How would you describe the data coverage of the reporting data warehouse?
GWRQ19	How would you describe the level of complexity of business data in the reporting data warehouse?
GWRQ20	What level of assistance does the architecture team have to offer the business users when looking for information in the client reporting data warehouse?
GWRQ21	What is your opinion of whether the business users have indicated they found it difficult to navigate the reporting data warehouse?
GWRQ22	How much weight do you think was given to distributing data to different areas of the business by client reporting data warehouse?

APPENDIX 2 – DESIGN OF RESEARCH INSTRUMENT FOR ICAPITAL

Questions for the ICapital Research Participants

ICRQ1 – What are the problems that the business wants solved that prompted the development of a data warehouse?

ICRQ2 - How is the need to address the business requirements led to choosing a relational data model for Risk and Finance data warehouse?

ICRQ3 - What role did the goal of Risk and Finance data warehouse had in choosing a relational data model for Risk and Finance data warehouse?

ICRQ4 - What role did the scope of Risk and Finance data warehouse had in choosing a relational data model for Risk and Finance data warehouse?

ICRQ5 - What is the end-user ability to understand the data relationships in Risk and Finance data warehouse?

ICRQ6 - What role would you say the staff experience played in choosing a relational data model for Risk and Finance data warehouse?

ICRQ7 – Would you agree or disagree that your staff experience influenced the choice of the data model for your data warehouse?

ICRQ8 - What other data models would you say were considered for Risk and Finance data warehouse by your development team?

ICRQ9 – In what circumstances would your team consider to use different data models than one currently used for Risk and Finance data warehouse?

ICRQ10 - What is the role of performance in choosing a relational data model for Risk and Finance data warehouse?

ICRQ11 - What is the impact of ensuring high performance of query execution on the choice of relational data model for Risk and Finance data warehouse?

ICRQ12 – What are the non-functional factors that influenced the decision to adopt relational data model for Risk and Finance data warehouse?

ICRQ13 – How would you describe the users response of executing queries against Risk and Finance data warehouse?

ICRQ14 - In your opinion, what has been the users' experience in analyzing and querying Risk and Finance data warehouse?

ICRQ15 - How diverse is the user base of your data warehouse?

ICRQ16 - How often are different areas of the business on-boarded into Risk and Finance data warehouse?

ICRQ17 - How many areas of the business have been on-boarded onto Risk and Finance data warehouse?

ICRQ18 - How would you describe the data coverage of Risk and Finance data warehouse?

ICRQ19 - How would you describe the level of complexity of business data in Risk and Finance data warehouse?

ICRQ20 - What level of assistance does the architecture team have to offer the business users when looking for information in Risk and Finance data warehouse?

ICRQ21 – What is your opinion of whether the business users have indicated they found it difficult to navigate Risk and Finance data warehouse?

ICRQ22 - How much weight do you think was given to distributing data to different areas of the business by Risk and Finance data warehouse?

ICRQ23 - How would you characterise the advantages of Risk and Finance warehouse data model?

Alignment of the ICapital Research Questions with the Main Research Questions

MRQ1 – What is the impact of high performance of query execution on the choice of logical data model for a data warehouse?

ID	Question
ICRQ10	What is the role of performance in choosing a relational data model for Risk and Finance data warehouse?
ICRQ11	What is the impact of ensuring high performance of query execution on the choice of relational data model for Risk and Finance data warehouse?
ICRQ12	What are the non-functional factors that influenced the decision to adopt relational data model for Risk and Finance data warehouse?
ICRQ13	How would you describe the users response of executing queries against Risk and Finance data warehouse?
ICRQ14	In your opinion, what has been the users' experience in analyzing and querying Risk and Finance data warehouse?
ICRQ23	How would you characterise the advantages of Risk and Finance warehouse data model?

MRQ2 - What is the impact of specific business requirement on the choice of a logical data model?

ID	Question
ICRQ1	What are the problems that the business wants solved that prompted the development of a data warehouse?
ICRQ2	How is the need to address the business requirements led to choosing a relational data model for Risk and Finance data warehouse?

MRQ3 – What is the impact of goal and scope of a data warehouse on the choice of data warehouse logical data model?

ID	Question
GWRQ3	What role did the goal of Risk and Finance data warehouse had in choosing a relational data model for Risk and Finance data warehouse?
GWRQ4	What role did the scope of Risk and Finance data warehouse had in choosing a relational data model for Risk and Finance data warehouse?
GWRQ5	What is the end-user ability to understand the data relationships in Risk and Finance data warehouse?

MRQ4 - What is the impact of implementation orientation of available resources on the choice of a logical data model?

ID	Question
ICRQ6	What role would you say the staff experience played in choosing a relational data model for Risk and Finance data warehouse?
ICRQ7	Would you agree or disagree that your staff experience influenced the choice of the data model for your data warehouse?
ICRQ8	What other data models would you say were considered for Risk and Finance data warehouse by your development team?
ICRQ9	In what circumstances would your team consider to use different data models than one currently used for Risk and Finance data warehouse?

MRQ5 – What is the impact of consolidating common enterprise data to meet different reporting and analytics requirements on the choice of logical data model for a data warehouse?

ID	Question
ICRQ15	How diverse is the user base of your data warehouse?
ICRQ16	How often are different areas of the business on-boarded into Risk and Finance data warehouse?
ICRQ17	How many areas of the business have been on-boarded onto Risk and Finance data warehouse?
ICRQ18	How would you describe the data coverage of Risk and Finance data warehouse?
ICRQ19	How would you describe the level of complexity of business data in Risk and Finance data warehouse?
ICRQ20	What level of assistance does the architecture team have to offer the business users when looking for information in Risk and Finance data warehouse?
ICRQ21	What is your opinion of whether the business users have indicated they found it difficult to navigate Risk and Finance data warehouse?
ICRQ22	How much weight do you think was given to distributing data to different areas of the business by Risk and Finance data warehouse?

APPENDIX 3 – GWEALTH OPEN CODES

Transcription of Empirical Data from Research Participant

TRANSCRIBED EMPIRICAL DATA

KO

What is the impact of ensuring high performance of query execution on the choice of the multidimensional data model for the reporting data warehouse?

SMMMBW

I think it is extremely important, I think you need to kind of consider what type of queries the user would want to run before you choose a particular model and, here our data warehouse wasn't that big to be honest. At [redacted], we have millions of transactions coming through every single day and again we found performance wise a dimensional model served the purpose of our needs so we built [redacted] on top of that in [redacted] and again the same we had it here; it seems to work a lot better. That said, the performance is also impacted by the database that you choose to use so, I think the moment that you choose and the hardware that you choosing to implement that model on, very very tightly coupled so we did have problems; I'm not sure if you are aware of the whole [redacted] side of things and so you could choose the best model in the world and you could design it to perfection, but if your hardware is not there to support that model, you're still going to have those problems. If your hardware is not up to the scratch, you will struggle and extending that; how you choose to load it would be another thing. Here we use [redacted] and it's a very high performance tool, ETL tool, I did have previous experience in [redacted] previously and this would not want to say anything about [redacted] as such, but at [redacted] we moved from [redacted] to [redacted] so to give you a real life kind of example, it was taking 22hrs with [redacted] to load all the data that we need into [redacted] warehouse and we have 24 hours in a day, so any problem that came up, you got just 2hrs to deal with it and obviously that was extremely difficult to maintain. We moved it to [redacted] and we got it down to 4 to 6 hrs. So it was quite a significant saving – we found the version we were on in [redacted], the version of [redacted] that they gave us at the time comparable with what [redacted] could do for us was massively different. So all those things, I just want to retreat the point, it is very important to choose the right model, but then if you don't have the tool to stand behind that model, it's kind of worthless; you lost all of the benefit that you might got with it.

Line-By-Line Coding of Transcribed Empirical Data

EMPIRICAL DATA: LINE-BY-LINE CODING

Transcribed Data	Apply Code to Text: Descriptive
The main reason why I chose dimensional structure for a warehouse would be	Performance is the main reason why I chose dimensional structure
performance to actually to bring back data. I always find that works a lot better wh	I always find that works a lot better for high performance
got more complex queries to get across to more tables and I find it that sometime:	when you get complex query to get across
you are inputting things in a very, very fast structure when that table gets too big i	Works better if you are imputing things into fast structure
coming through every single day and again we found performance wisea dimensic	Millions of transactions comes in daily
model served the purpose of our needs so we built … on top of that in …	Dimensional model served our needs performance wise
and again the same we had it here; it seems to work a lot better. That said, the	Dimensional model seem to work better for performance
performance is also impacted by the database that you choose to use so, I think th	Performance is impacted by the database that you choose
moment that you choose and the hardware that you choosing to implement that n	Performance is impacted by the data model that you choose
very, very tightly coupled so we did have problems; I'm not sure if you are aware c	Performance is impacted by the hardware you choose for your data m
whole … side of things and so you could choose the best model in the world and y	You could choose the best model in the world
could design it to perfection, but if your hardware is not there to support that moc	Your hardware must be able to support your data model
you're still going to have those problems. If your hardware is not up to the scratch	You will have problems if hardware cannot support your data model
will struggle and extending that; how you choose to load it would be another thin	It's another thing how you choose to load your data
we use … and it's a very high performance tool, ETL tool, I did have previous	Used high performance ETL tool
experience in … previously and this would not want to say anything about	Would not say anything about previous ETL tool
… as such, but at … we moved from … to … so to give	Moved from one ETL to another
you a real life kind of example, it was taking 22hrs with … to load all the data	Taking 22hrs to load data into a data warehouse
that we need into … warehouse and we have 24 hours in a day, so any problem th	Data loading taking almost all the time in a day
came up, you got just 2hrs to deal with it and obviously that was extremely difficu	only 2 hrs to deal with problems
maintain. We moved it to … and we got it down to 4 to 6 hrs, So it was quite a	Got data loading dow to 6 hrs by moving into another tool
the performance of the micros that we needed for client reporting,	Help with performance of micros for client reporting
It has a massive role, you have to define what data people are looking for, what al	Performance is one of the greatest consideration to take into account
different sources are, what their end goal is, what are you trying to achieve with tl	Performance is very important when choosing your data architecture
business you have to get that very, very clear, is it to provide them with a wareho	I've not worked in a place where users don't care about performance

Transforming Descriptive Codes to Analytic Codes

TRANSFORMING DESCRIPTIVE CODES TO ANALYTIC CODES

Apply Code to Text: Descriptive	Descriptive Code -> Analytic Code	Compare Codes for Saturation
Works better if you are imputing things into fast structure	Multidimensional model works better for fast structures	Multidimensional model works better for high performance
Millions of transactions comes in daily	Client Reporting processes large amount of transactions daily	Multidimensional model works better for high performance
Dimensional model served our needs performance wise	Multidimensional model addressed business performance requirem	Multidimensional model works better for high performance
Dimensional model seem to work better for performance	Multidimensional model works better for high performance	Multidimensional model works better for high performance
Performance is impacted by the database that you choose	Performance is impacted by software and hardware	Performance is impacted by software and hardware
Performance is impacted by the data model that you choose	Performance is impacted by choice of data model	Performance is impacted by choice of data model
Performance is impacted by the hardware you choose for your data model	Performance is impacted by software and hardware	Performance is impacted by software and hardware
You could choose the best model in the world	Performance is impacted by software and hardware	Performance is impacted by software and hardware
Your hardware must be able to support your data model	Performance is impacted by software and hardware	Performance is impacted by software and hardware
You will have problems if hardware cannot support your data model	Performance is impacted by software and hardware	Performance is impacted by software and hardware
It's another thing how you choose to load your data	Performance is impacted by software and hardware	Performance is impacted by software and hardware
Used high performance ETL tool	Performance is impacted by software and hardware	Performance is impacted by software and hardware
Would not say anything about previous ETL tool	There is nothing to say about previous ETL software	Performance is impacted by software and hardware
Moved from one ETL to another	Replaced previous ETL tool with another tool	Performance is impacted by software and hardware
Taking 22hrs to load data into a data warehouse	It took nearly a day to load data into client reporting	Performance is impacted by software and hardware
Data loading taking almost all the time in a day	It took nearly a day to load data into client reporting	Performance is impacted by software and hardware
only 2 hrs to deal with problems	There is not enough time to address data loading issues	Performance is impacted by software and hardware
Got data loading dow to 6 hrs by moving into another tool	Replaced previous ETL tool with another	Performance is impacted by software and hardware
Help with performance of micros for client reporting	Gained improved performance for client reporting	Gained improved performance for client reporting
Performance is one of the greatest consideration to take into account	Performance is an important consideration	Multidimensional model works better for performance
Performance is very important when choosing your data architecture	Performance is an important consideration	Multidimensional model works better for performance
I've not worked in a place where users don't care about performance	Performance is an important consideration	Multidimensional model works better for performance
The answer is as fast as possible	Performance is an important consideration	Multidimensional model works better for performance

Identification of Sub-Category Code from Analytic Codes

IDENTIFICATION OF SUB-CATEGORY CODES

Descriptive Code -> Analytic Code	Compare Codes for Saturation	Sub Category Code
Chose multidimensional model for high performance	Multidimensional model works better for high performance	High Query Performance
Multidimensional model works better for high performance	Multidimensional model works better for high performance	High Query Performance
Multidimensional model works for complex queries	Multidimensional model works better for high performance	High Query Performance
Multidimensional model works better for fast structures	Multidimensional model works better for high performance	High Query Performance
Client Reporting processes large amount of transactions daily	Multidimensional model works better for high performance	High Query Performance
Multidimensional model addressed business performance req	Multidimensional model works better for high performance	High Query Performance
Multidimensional model works better for high performance	Multidimensional model works better for high performance	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by choice of data model	Performance is impacted by choice of data model	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
Performance is impacted by software and hardware	Performance is impacted by software and hardware	High Query Performance
There is nothing to say about previous ETL software	Performance is impacted by software and hardware	High Query Performance
Replaced previous ETL tool with another tool	Performance is impacted by software and hardware	High Query Performance
It took nearly a day to load data into client reporting	Performance is impacted by software and hardware	High Query Performance
It took nearly a day to load data into client reporting	Performance is impacted by software and hardware	High Query Performance
There is not enough time to address data loading issues	Performance is impacted by software and hardware	High Query Performance
Replaced previous ETL tool with another	Performance is impacted by software and hardware	High Query Performance
Gained improved performance for client reporting	Gained improved performance for client reporting	High Query Performance
Performance is an important consideration	Multidimensional model works better for performance	High Query Performance

Identification of Main Category Code from Sub-Category Code

MAIN CATEGORY CODES FOR SUB-CATEGORY CODES

Sub Category Code	Main Category Code
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution
High Query Performance	High Performance of Query Execution

Property of Main Category Code

MAIN CATEGORY CODE: PROPERTY

Main Category Code	Property of Main Category Code
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance
High Performance of Query Execution	High Query Performance

Frequency Count of Analytic Codes

FREQUENCY OCCURRENCE OF ANALYTIC CODE

Analytic Code	Frequency Occurrence
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
High Query Performance	1
....	....
High Query Performance	1
High Query Performance	1
High Query Performance	1
Total	72

APPENDIX 4 – ICAPITAL OPEN CODES

Transcription of Empirical Data from Research Participant

TRANSCRIBED EMPIRICAL DATA

KO

What is the end-user ability to understand the data relationships in Risk and Finance Data Warehouse?

MRRMBC

Not exactly, the approach to the model, for me, the business have looked at the model, but we show that model to the business it was at extremely high level, we showed the client, trade, party, we showed issuer, things the business is interested in we didn't try to define in a manner that you would think about in physical or logical model. We want our model to be about how fast are we able to enter and read information – as simple as that. Physical implementation is all about the access; correct? It wasn't about whether you can understand what you going to use, it wasn't about any of that stuff. It's all about fast inserts and spontaneous inserts and reads. So we have a high level model which is very high level to the major pieces of information and within each of that high level, you are able to drill into the next detail level of that model – correct? And I think we want to the personal level that is the high level to the second level that is already at the detailed level and we kind of arrive at ER model to allow us to keep information which primarily finance is interested in, risk is interested in. We made the attributes that are associated with that, for example, balance and all of the pieces that hang off of it for example the instrument, counterparty of the balance etc.; so we look at it from the business domain and we said finance is interested in balance, risk is interested in sensitivities and all the other pieces of information related to sensitivities – correct? So, the only place that we sort of really weren't closing the deal is on the finest grain of the model is for FLEX- that's a good model.
And in the case of marts for FLEX, we really sat down with the end users in understanding exactly the product information that they are after for reporting and we created that model which is pure mart model, dimensional model; they get three level of dimensions of balances, at trade level, at the party level and at the group level, right? So we have 3 levels and all the facts and dimensions around it; there are some agreed matrixes; there is a matrix saying we can agree 82000 rows every second, we can write 32000 rows every second so we build our model so that we can achieve that.

Line-By-Line Coding of Transcribed Empirical Data

EMPIRICAL DATA: LINE-BY-LINE CODING

Transcribed Data	Apply Code to Text: Descriptive Code
but we show that model to the business it was at extremely high level, we showed the client,	The business looked at the model at high level
trade, party, we showed issuer, things the business is interested in we didn't try to define in a	We showed the business what they are interested in
manner that you would think about in physical or logical model. We want our model to be about	Model not defined to business in typical logical or physical manner
So we have a high level model which is very high level to the major pieces of information and	Major pieces of information is model at high level
within each of that high level, you are able to drill into the next detail level of that model –	The data model allows you to drill to next level of detail
correct? And I think we want to the personal level that is the high level to the second level that is	The model has personal and second level of information
already at the detailed level and we kind of arrive at ER model to allow us to keep information	Arrived at ER to address requirements
which primarily finance is interested in, risk is interested in. We made the attributes that are	The model contains information that finance and risk are interested in
associated with that, for example, balance and all of the pieces that hang off of it for example the	The data model contains balance information
instrument, counterparty of the balance etc.; so we look at it from the business domain and we	The data model also contains instrument, counterparties and balances
said finance is interested in balance, risk is interested in sensitivities and all the other pieces of	Finance is interested in balances and risk is interested in sensitivities
information related to sensitivities – correct? So, the only place that we sort of really weren't	Sensitivities and other related information is fact
closing the deal is on the finest grain of the model is for FLEX- that's a good model.	Addressing the grain of FLEX reporting, it's a good model
And in the case of marts for FLEX, we really sat down with the end users in understanding	Sat down with business users to understand their requirements
exactly the product information that they are after for reporting and we created that model which	Sat down with business users to understand their requirements
is pure mart model, dimensional model; they get three level of dimensions of balances, at trade	Created model that addressed information required by business
level, at the party level and at the group level, right? So we have 3 levels and all the facts and	Created model that addressed information required by business
dimensions around it; there are some agreed matrixes; there is a matrix saying we can agree	Created model that addressed information required by business
82000 rows every second, we can write 32000 rows every second so we build our model so that	Created model that addressed information required by business
we can achieve that.	Created model that addressed information required by business
Yeah, I would say absolutely true, correct? So you have to create data architecture, all the access	Model created to address business requirements
on the system right? You have to understand all the loading and the reading access of	You need to understand all your requirements
comparable systems if you don't articulate that, you end up with a system that is un-useable,	You need to understand your requirements to build useable system
that cannot grow overtime, yes I agree with that statement	You need to understand all your requirements
that very well. So I would say in the case of RFDW, we set out the left side and the right side as	Data model contain right side and left side information
you might have known; the left side is the information address driving just-in-time information,	Left side contains just-in-time information

Transforming Descriptive Codes to Analytic Codes

TRANSFORMING DESCRIPTIVE CODES TO ANALYTIC CODES

Apply Code to Text: Descriptive Code	Descriptive Code -> Analytic Code	Compare Codes for Saturation
The business looked at the model at high level	Model reviewed with business at high level	Model addressed generic business requirements
We showed the business what they are interested in	Model addressed generic business requirements	Model addressed generic business requirements
Model not defined to business in typical logical or physical manner	Model reviewed with business at high level	Model addressed generic business requirements
Major pieces of information is model at high level	Model addressed generic business requirements	Model addressed generic business requirements
The data model allows you to drill to next level of detail	Model enables drill down of information	Model addressed generic business requirements
The model has personal and second level of information	Model enables drill down of information	Model addressed generic business requirements
Arrived at ER to address requirements	Model addressed generic business requirements	Model addressed generic business requirements
The model contains information that finance and risk are interested in	Model addressed generic business requirements	Model addressed generic business requirements
The data model contains balance information	Model addressed generic business requirements	Model addressed generic business requirements
The data model also contains instrument, counterparties and balances	Model addressed generic business requirements	Model addressed generic business requirements
Finance is interested in balances and risk is interested in sensitivities	Model addressed generic business requirements	Model addressed generic business requirements
Sensitivities and other related information is fact	Model addressed generic business requirements	Model addressed generic business requirements
Addressing the grain of FLEX reporting, it's a good model	Model addressed generic business requirements	Model addressed generic business requirements
Sat down with business users to understand their requirements	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Sat down with business users to understand their requirements	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Created model that addressed information required by business	Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp
Model created to address business requirements	Model addressed generic business requirements	Model addressed generic business requirements
You need to understand all your requirements	Business requirements must be understood	Model addressed generic business requirements
You need to understand your requirements to build useable system	Business requirements must be understood	Model addressed generic business requirements
You need to understand all your requirements	Business requirements must be understood	Model addressed generic business requirements
Data model contain right side and left side information	Data model contains generic information	Model addressed generic business requirements
Left side contains just-in-time information	Data model contains generic information	Model addressed generic business requirements

Identification of Sub-Category Code from Analytic Codes

IDENTIFICATION OF SUB-CATEGORY CODES

Descriptive Code -> Analytic Code	Compare Codes for Saturation	Sub Category Code
Model reviewed with business at high level	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model reviewed with business at high level	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model enables drill down of information	Model addressed generic business requirements	Generic Requirements
Model enables drill down of information	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Created multidimensional to focus on reporting require	Multidimensional model enables you to focus on sp	Specific Requirement
Model addressed generic business requirements	Model addressed generic business requirements	Generic Requirements
Business requirements must be understood	Model addressed generic business requirements	Generic Requirements
Business requirements must be understood	Model addressed generic business requirements	Generic Requirements
Business requirements must be understood	Model addressed generic business requirements	Generic Requirements
Data model contains generic information	Model addressed generic business requirements	Generic Requirements
Data model contains generic information	Model addressed generic business requirements	Generic Requirements

Identification of Main Category Code from Sub-Category Code

MAIN CATEGORY CODES FOR SUB-CATEGORY CODES

Sub Category Code	Main Category Code
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Specific Requirement	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement
Generic Requirements	Business Requirement

Properties of Main Category Code

MAIN CATEGORY CODE: PROPERTIES

Main Category Code	Property of Main Category Code
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Specific Requirement
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements
Business Requirement	Generic Requirements

Frequency Count of Analytic Codes

FREQUENCY OCCURRENCE OF ANALYTIC CODE

Analytic Code	Frequency Occurrence
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
Generic Requirements	1
...	...
Generic Requirements	1
Generic Requirements	1
Total	381

APPENDIX 5 –TEST OF SIGNIFICANT RELATIONSHIP

Contingency Table of Logical Data Models and Observed Frequency of Analytic Codes

Logical Data Model (LDM)	OUTCOME					**Total**
	ACV1	ACV2	ACV3	ACV4	ACV5	
Multidimensional Data Model (**MDM**)	72	84	65	81	22	324
Relational Data Model (**RDM**)	87	381	197	166	436	1267
Total	159	465	262	247	458	1591

Observed and Expected Frequency of Analytic Codes and Logical Data Models

Logical Data Model (LDM)	OUTCOME (Observed & Expected)					**Total**
	ACV1	ACV2	ACV3	ACV4	ACV5	
Multidimensional Data Model (**MDM**)	72 (32.38)	84 (94.70)	65 (53.36)	81 (50.30)	22 (93.27)	324
Relational Data Model (**RDM**)	87 (126.62)	381 (370.30)	197 (208.64)	166 (196.70)	436 (364.73)	1267
Total	159	465	262	247	458	1591

$$X^2 = \sum \frac{(O - E)2}{E} = 157.5$$

Where:

O = Observed code frequency

E= Expected code frequency

157.5 *= Chi Square value for ACV1 – ACV5*

$$X^2 = \sum \frac{(72-32.38)2}{32.38} + \frac{(84-\ 94.70)2}{94.70} + \frac{(65-53.36)2}{53.36} + \frac{(81-50.30)2}{50.30} + \frac{(22-93.27)2}{93.27}$$

$$+ \frac{(87-126.62)2}{126.62} + \frac{(381-370.30)2}{370.30} + \frac{(197-208.64)2}{208.64} + \frac{(166-196.70)2}{196.70}$$

$$+ \frac{(436-364.73)2}{364.73}$$

=48.488 + 1.21 + 2.54 + 18.74 + 54.46 + 12.40 + 0.31 + 0.65 + 4.79 + 13.92 =157.5

Chi Square Distribution Table

Chi Square Distribution Table										
DF	**0.995**	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	**14.860**
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928

Chi Square Distribution Table										
DF	**0.995**	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Percentage Contribution of the Analytic Code Variables

Logical Data Model (LDM)	Percentage (%) Contribution of ACVs' For MDM & RDM					**% Total**
	ACV1	ACV2	ACV3	ACV4	ACV5	
Multidimensional Data Model (**MDM**)	22.22%	25.93%	20.06%	25%	6.79%	20.4
Relational Data Model (**RDM**)	6.87%	30.07%	15.55%	13.10%	34.41%	79.6%
% Total	9.99%	29.23%	16.47%	15.52%	28.79%	100%