

**Implementing Business Intelligence Processes in a
Telecommunications Company in South Africa**

D.P. DU PLESSIS

22079068-2009

Thesis submitted for the degree Doctor of Philosophy in Information
Technology at the Vaal Triangle Campus of the North-West University

Promoter: Prof. J.H. Kroeze

May 2012

Acknowledgements

My sincere thanks and appreciation go to the following people for their contribution to the successful completion of this thesis:

My promoter, Prof. J. H. Kroeze, for the guidance and leadership he provided during the writing of this thesis. He spent much of his valuable time mentoring and guiding me through the entire research project.

Mr. A. Venter, an executive at the telecommunications company used for the research - for arranging the necessary permission to undertake the research at his company.

Prof. P. Pretorius, from the North-West University, for financing my attendance at the 2010 Doctoral Consortium of the European Conference on Information Systems (ECIS).

Prof. J. Tolmie, from the University of the Free State, for awarding me a scholarship to attend one of the data warehousing courses of the world leader in BI – Dr. R. Kimball - in Chicago, USA.

Ms. M. Esterhuizen, from the library of the North-West University, Vaal Triangle Campus, for her professional services and support with the identification of information sources.

Prof. C. De Villiers, from the University of Pretoria, for helping me to find a highly qualified and dedicated promoter.

My wife, Lenie, and sons, Breyton and Dylan, for their continued love and support, as well as their patience and understanding.

All the honour and glory goes to my Saviour Jesus Christ.

Opsomming

Hierdie tesis fokus op probleme wat ondervind is met die implementering van Besigheidsintelligensie (BI) in 'n telekommunikasie maatskappy in Suid Afrika. Die drie hoof uitdagings wat deur hierdie studie aangespreek word, is: die hoë BI implementasie-koste, die BI-geletterdheid van BI-eindgebruikers (besigheidsmense weet nie noodwendig hoe om BI-sagteware te gebruik nie) en die gebrek aan tyd om 'n BI-oplossing te implementeer (die maatskappy moes onmiddellik begin om aan die regering se doelwitte te voldoen).

Hierdie studie stel 'n konseptuele raamwerk met twee modelle en een argitektuur voor. Hierdie raamwerk kan 'n maatskappy met bogenoemde uitdagings deur die implementasie van 'n BI-oplossing lei.

Summary

The focus of this thesis is on problems experienced with the implementation of Business Intelligence (BI) in a telecommunications company in South Africa. The three main challenges addressed by the study were: the high BI implementation cost, the BI literacy of the BI end-users (business people do not necessarily know how to use BI tools), and the limited time to implement a BI solution (the company had to immediately start delivering on government targets).

This study presents a conceptual framework that includes two models and a BI architecture. This framework may guide a company with the above-mentioned challenges through the implementation of a BI solution.

Table of contents

Acknowledgement	iii
Opsomming	iv
Summary	v
Table of contents	vi
List of figures	xiii
List of tables	xiv
Glossary definitions	xv
Keywords	xvii
CHAPTER 1 The problem statement and research methodology	1
1.1 Introduction.....	1
1.2 Definition of business intelligence.....	1
1.3 Origins and background of study	2
1.4 The importance of business intelligence	6
1.5 Motivation for the study.....	7
1.5.1 Business intelligence implementation challenges	8
1.6 Problem statement.....	9
1.7 Main research question	10
1.7.1 Secondary research questions	10
1.8 Central theoretical statement.....	11
1.9 Method of investigation	11
1.9.1 Literature review	13
1.9.2 Case study	14
1.9.3 Action research	16
1.9.4 Survey	17
1.10 Contribution to the field of information technology	18
1.11 Overview of the content of this study	18
1.12 Conclusion	21

CHAPTER 2	The BI concept	23
2.1	Introduction	23
2.2	What is data warehousing?	24
2.3	History of data warehousing	25
2.4	Difference between the OLTP system and a data warehouse	26
2.5	Data warehouse design techniques.....	27
2.5.1	Dimension and fact tables	27
2.5.2	Dimension types.....	33
2.5.3	Fact tables	36
2.6	The importance of business intelligence	40
2.7	Business intelligence architecture	41
2.7.1	Data source.....	42
2.7.2	Data acquisition.....	42
2.7.3	Enterprise data.....	45
2.7.4	Data enhancement	47
2.7.5	Data mart.....	48
2.7.6	Data access.....	48
2.8	Conclusion	49
CHAPTER 3	The double wave data warehouse lifecycle model.....	51
3.1	Introduction	51
3.2	The case study research methodology.....	52
3.2.1	Data generation methods.....	52
3.3	Order process case study.....	53
3.3.1	Challenges with existing BI SDLCs	53
3.3.1.1	Agile BI.....	53
3.3.1.2	Inmon’s CLDS	54
3.3.1.3	The business dimensional lifecycle model.....	55
3.3.1.4	Requirement phase.....	56
3.3.1.5	Dimensional modelling and physical design.....	57

3.3.1.6	Technical architecture design and product selection/installation.....	58
3.3.1.7	End-user specification and development	58
3.3.1.8	Data staging design and development.....	58
3.3.1.9	Deployment.....	59
3.3.1.10	Maintenance and growth.....	59
3.3.2	Difficulties with the business dimensional lifecycle model.....	59
3.3.3	Development of the DWDW lifecycle model.....	61
3.3.4	The order process	62
3.4	Data analysis	70
3.4.1	Data categorisation.....	70
3.4.2	Wave 1	72
3.4.3	Wave 2	73
3.5	Conclusion	74
CHAPTER 4	Refinement of the double wave data warehouse lifecycle model.....	76
4.1	Introduction.....	76
4.2	Action research methodology	76
4.3	Background	78
4.4	Diagnostic stage (action research)	79
4.4.1	Business requirement	80
4.4.2	Credit management process	81
4.5	Planning stage (action research).....	85
4.5.1	Credit management business requirement	86
4.5.2	Improving the DWDW lifecycle model.....	86
4.6	Intervention stage (action research)	86
4.6.1	Wave 1	87
4.6.1.1	Receive critical requirement	87
4.6.1.2	Data profiling	88
4.6.1.3	Extract, merge and load (EML)	88
4.6.1.4	Limited processing power.....	89

4.6.1.5	Limited server random access memory.....	90
4.6.1.6	Limited network speed.....	91
4.6.1.7	The wrong disk speed and storage combination	91
4.6.1.8	Loading the flat file during wave 1 EML process.....	92
4.6.1.8.1	Transactions	92
4.6.1.8.2	Master data.....	96
4.6.1.9	EML process	96
4.6.1.10	Use BI front-end tool on flat file.....	100
4.6.1.11	Test and refine the results	100
4.6.1.12	Is the business requirement once-off?.....	101
4.6.2	Wave 2	101
4.6.2.1	Build a dimensional model	101
4.6.2.2	Document the product.....	103
4.6.2.3	Transform and load process	104
4.6.2.4	Dimension tables.....	105
4.6.2.5	Fact tables	106
4.6.2.6	Build new BI front-end or build solution into existing BI front-end	107
4.6.2.7	Post-delivery maintenance	107
4.7	Evaluation stage (action research)	108
4.8	Reflection stage (action research).....	108
4.9	Conclusion	108
CHAPTER 5	Double wave data warehouse architecture.....	110
5.1	Introduction.....	110
5.2	Diagnostic stage (action research)	111
5.2.1	Different components of a BI architecture.....	111
5.2.2	Data acquisition.....	112
5.2.3	Data storage	113
5.2.4	Information delivery	113
5.3	Planning stage (action research).....	114

5.3.1	Limitations of existing BI architecture when using the DWDW lifecycle model	114
5.3.2	Intervention stage (action research)	115
5.3.2.1	Conformed dimension ensures smooth movement of BI solution from development environment to production environment.....	116
5.3.3	DWDW architecture also caters for disaster recovery	121
5.4	Evaluation stage (action research)	132
5.5	Reflection stage.....	132
5.6	Conclusion	133
CHAPTER 6	Maturing a BI solution: The need for the BI literacy and culture maturity model	134
6.1	Introduction.....	134
6.2	Developing a BI division	135
6.2.1	The spare infrastructure requirement	137
6.2.2	The spare infrastructure BI solution.....	140
6.3	BI culture as part of corporate culture.....	142
6.4	The role of BI literacy in a company's BI culture.....	144
6.5	BI literacy and culture maturity model	145
6.6	Data analysis	146
6.6.1	Data categorisation.....	146
6.7	Conclusion	152
CHAPTER 7	Business intelligence strategic interactive and incremental framework.....	153
7.1	Introduction.....	153
7.2	Background	154
7.3	A new BI strategy	155
7.3.1	Business intelligence vision	157
7.3.2	Data	158
7.3.3	Information	159
7.3.3.1	Installation and fault management	161
7.3.3.2	Billing	163
7.3.3.3	Dismantlement	165

7.3.4	Analytic.....	166
7.3.4.1	Analysis of the telephone market.....	166
7.3.4.2	Market segmentation.....	169
7.3.4.3	Marketing based on the market segmentation.....	170
7.3.4.4	Planning infrastructure based on market segmentation.....	170
7.3.5	Knowledge	171
7.3.6	Wisdom.....	172
7.4	BI literacy and culture maturity model	174
7.5	Conclusion	175
CHAPTER 8	Evaluation of the DWDW lifecycle model	176
8.1	Introduction.....	176
8.2	Interview schedule	177
8.3	Sampling methods.....	178
8.4	Compiling the interview schedules	180
8.5	Pilot interview schedules	182
8.6	Transcription of interview data	183
8.7	Data analysis	183
8.8	Interview with BI team	184
8.8.1	Experience of BI professionals	185
8.8.2	Interview data analysis.....	188
8.8.3	Benefits of the DWDW lifecycle model	192
8.8.4	DWDW lifecycle model business requirement specification	193
8.8.5	DWDW lifecycle model database design techniques	193
8.8.6	ETL processes of the DWDW lifecycle model.....	194
8.9	Analysis of BI team interview data.....	195
8.9.1	Quality of information received from the BI solution	199
8.9.2	Performance of the BI solution	200
8.9.3	Analysis of super-user interview data.....	200
8.10	Conclusion	201

CHAPTER 9	Conclusion	202
9.1	Introduction.....	202
9.2	Summary of the thesis.....	205
9.3	Future research.....	206
REFERENCES	209

List of figures

Figure 1.3 Cost of implementing a BI solution.....	5
Figure 1.5 Implementation cost of BI.	7
Figure 2.5.1.1 Differences between dimension table and fact table.	28
Figure 2.5.1.2 Simple data warehouse design.....	29
Figure 2.5.1.3 Explanation of bitmap index.....	32
Figure 2.7 Data warehouse and BI architecture.....	41
Figure 2.7.2.1 The ETL process.....	43
Figure 2.7.2.2 Data flow tread.	44
Figure 2.7.3.1 The ODS design.....	47
Figure 3.3.1.3 The business dimensional lifecycle model.	55
Figure 3.3.4.1 Bucket system.....	65
Figure 3.3.4.2 Flat file table for order bucket reporting.....	67
Figure 3.3.4.3 Dimensional model developed for the order bucket system.....	69
Figure 3.4.1 The double wave warehouse lifecycle model.....	71
Figure 4.4.2 Telecommunications collection process.	84
Figure 4.6.1.8.1.1 Flat file with test data.	93
Figure 4.6.1.8.1.2 Data cube for debtors' management solution.	95
Figure 4.6.1.9 Using the delta field for table comparison.....	99
Figure 4.6.2.1 Data model for the debtors' management solution.....	102
Figure 4.6.2.5 Fact table is populated with the primary keys of the dimension tables.	106
Figure 5.2.1 Three components and services of BI architecture.....	112
Figure 5.3.2.1.1 DWDW architecture.	118
Figure 5.3.2.1.2 UML for credit management BI solution.	120
Figure 5.3.3 Code to create a BI solution for debtors' management.	122
Figure 6.2.1 Fixed line telecommunications network.....	139
Figure 6.2.2 Infrastructure dimensional model for spare infrastructure.	141

Figure 6.6.1.1 BI literacy and culture maturity model.....	148
Figure 6.6.1.2 Data warehousing maturity model.....	151
Figure 7.3a Yin Yang symbol.....	156
Figure 7.3b Business intelligence strategic iterative and incremental framework.....	157
Figure 7.3.3.1 Technical infrastructure management data model.....	160
Figure 7.3.4.1 Actuate OLAP data cube for customer billing information.....	168
Figure 8.4 Organogram for the people interviewed.....	181

List of tables

Table 2.4 Customer dimension table to demonstrate normalisation of a database.	27
Table 2.5.1 Surrogate key combination	31
Table 2.5.2.1 Type 1 SCD (before update).	33
Table 2.5.2.2 Type 1 SCD (after update).	34
Table 2.5.2.3 Type 2 SCD.....	35
Table 2.5.2.4 Type 3 SCD.....	36
Table 2.5.3.1 Cumulative fact tables.....	37
Table 2.5.3.2 Snapshot fact tables.....	39
Table 2.7.3 Differences between ODS and DW.	46
Table 7.3.6 Telephone service repairs for January 2004.	173
Table 8.8 Qualifications of the BI team interviewed	184
Table 8.8.1.1 Years' experience in this company	185
Table 8.8.1.2 Distribution of BI experience.....	186
Table 8.8.1.3 Experience BI team had of other SDLCs.....	187
Table 8.8.2 Evaluation of the DWDW lifecycle model, used by the BI team	188
Table 8.9 Qualifications of the super-users interviewed.....	194
Table 8.9.1 BI solution's utilisation in years	197

Table 8.9.2 Evaluation of the performance of the BI solution built using the DWDW lifecycle model	196
---	-----

Glossary definitions

BDLM	Business Dimensional Lifecycle Model
BI	Business Intelligence
BISIIF	Business Intelligence Strategic Iterative and Incremental Framework
BO	Business Objects
DM	Data Mart
DP	Distribution Point
DRP	Disaster Recovery Plan
DW	Data Warehouse
ERP	Enterprise Resource Planning
ETL	Extract, Transform and Load
GIS	Graphical Information System
GUI	Graphical User Interface
HCI	Human Computer Interaction
IN	Intelligent Network
IT	Information Technology
MD	Master Data
MDF	Main Distribution Frame
MOLAP	Multi-Dimensional On-Line Analytical Processing
MS	Microsoft

MTTI	Mean Time To Install
MTTR	Mean Time To Repair
ODBC	Open Database Connection
ODS	Operational Data Store
OLAP	On-Line Analytical Processing
OLTP	Online Transaction Processing
PABX	Private Automated Business Exchange
RAM	Random Access Memory
RCD	Rapidly Changing Dimensions
ROI	Return on Investment
ROLAP	Relational On-Line Analytical Processing
SAN	Storage Access Network
SCD	Slowly Changing Dimensions
SDC	Street Distribution Cabinet
SDLC	Software Development Lifecycle
SQL	Structured Query Language
UML	Unified Modelling Language

Keywords

Business intelligence

Data warehousing

Software development lifecycle

Corporate culture

Business intelligence culture

Telecommunications

CHAPTER 1

The problem statement and research methodology

1.1 Introduction

This chapter explains the definition of Business Intelligence (BI), discusses its importance, motivates the study undertaken for this thesis and explains the research strategy used. The use of case studies, action research and interviews are explained. The research problem, research question and the hypothesis will be discussed. The chapter concludes with a listing and summary of all later chapters presented in the thesis.

1.2 Definition of business intelligence

There are several definitions for BI. Three have been selected from the literature and will be discussed and considered in this thesis:

- Viviers *et al.* (2005:1) define BI as - “a broad category of application programs and technologies for gathering, storing, analysing, and providing access to data stored in a data warehouse to help enterprise users make better business decisions.”

- According to Wally (2003:1), “BI is the process of getting enough of the right information in a timely manner and usable form to analyse it, so that it can have a positive impact on business strategy, tactics or operations.” Wally (2003:1) also states that BI applications include the “activities of decision support, query and reporting, online analytical processing (OLAP), statistical analysis, forecasting and data mining”.
- Pisello (2005:1) states that “business intelligence solutions hold great promise to automate and consolidate the analysis, presentation, reporting and compliance capabilities necessary to free enterprise data for actionable insight”.

Based on the above definitions, the assumption can be made that BI is a strategic and operational information management tool used for decision support in an organisation.

The company used in this study had to develop a BI solution in a very short period of time to facilitate very important business decisions. The origins of the study will be discussed in the next section.

1.3 Origins and background of study

This study was undertaken in a South African telecommunications company. This new company was created as part of a privatisation exercise from some of the government’s state departments. In 1994, just after the privatisation of the Department of Post and Telecommunications, the new democratic government

of South Africa was the sole shareholder of the only “fixed line telecommunications company” in South Africa. The Government had vibrant discussions with all the relevant parties on how telecommunications might be restructured to create an even distribution of access to telecommunications services to all the people in the country. This resulted in a White Paper on Telecommunications policy, which was released in March 1996 (Anon, 1996:1).

The major proposal contained in the White Paper was that the owner and operator of the fixed telephony infrastructure would be granted a limited period, the so-called “exclusivity period”, with regard to the monopoly in the provision of basic telecommunications services. This exclusivity period was to last for five years, but could be extended to six years if the telecommunications operator met network rollout and service targets. The rollout targets included doubling its subscriber access lines by 2.7 million, installing 120,000 new public telephones, connecting 3,200 villages for the first time and providing service to more than 20,000 priority customers such as schools and clinics (Anon, 1996:1). The exclusivity period was intended to allow the company to expand the network as rapidly as possible in order to facilitate universal access and to move towards universal service. The agreement left the telecommunications’ provider with the challenge to plan and manage the implementation targets set by government, while at the same time preparing for competition once the exclusivity period expired.

A BI solution was needed by the telecommunications' provider, that provided information on spare infrastructure. There were several challenges whilst implementing this BI solution. The three main challenges addressed by this study were: the high BI implementation cost, the BI literacy of the BI end-users (the business people did not know how to use the BI tool), and the limited time to implement the BI solution (the company had to immediately start delivering on government targets).

The implementation of a BI solution can be very challenging. Sumathi and Sivanandam (2006:145) state that almost every BI project follows the 2:2:50 patterns. This means that the project costs an average of \$2 million, takes an average of two years to complete, and has an expected 50% failure rate. A study done by Madsen (2010:8) in 1000 global companies in different industries, revealed that a large company - similar to the company used for this study - pays \$1,586,826 to implement a BI solution (see figure 1.3).

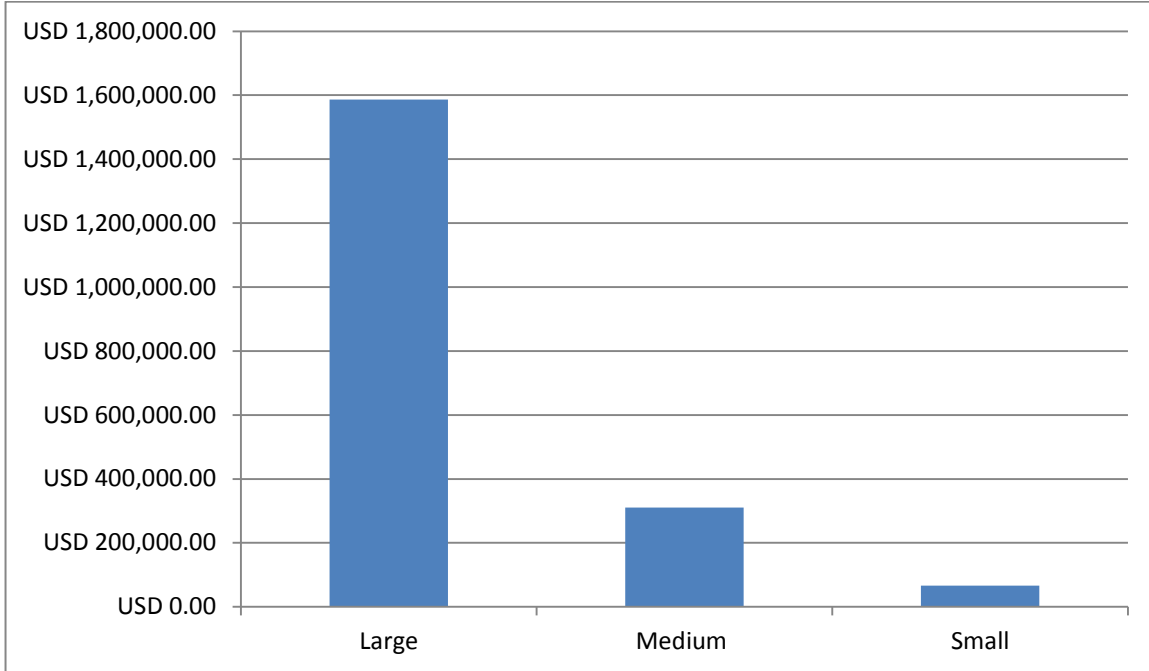


Figure 1.3 Cost of implementing a BI solution (Madsen, 2010:8).

Hart (2006:11) states - based on a research done in 24 large South African companies which included two telecommunications companies - that BI implementation in South Africa costs between R20 and R100 million. It is, therefore, difficult or almost impossible for a small company to implement a BI solution (Hart 2006:11).

The company concerned had limited funds to develop a BI solution. The BI team, therefore, had to find ways to minimise the cost of implementing the BI solution. This BI solution was needed immediately, and could not wait two years for completion. Failure, therefore, could not be tolerated.

1.4 The importance of business intelligence

According to Williams and Williams (2004:206), data warehousing and BI initiatives have been IT-driven in the past, with the focus on how technology delivers information to the BI user community. Today, BI is focused on how well the BI solution is utilised by the business user.

Burton *et al.* (2006:1) state that the management of business and operations in larger organizations is becoming more challenging and complex, and it is getting worse because of difficulties in the world economy in 2008 and 2009. Managing this complexity means that BI departments in organisations are called upon to provide BI-related capabilities for understanding where and how value can be created in the business. This is done in order to respond quickly to market changes and opportunities in the contemporary business world. These macro business changes require that organisations view BI in different ways.

The advantages derived from BI range from simple cost avoidance like saving on labour cost to competitive advantage information such as rapidly identifying hot selling items and responding positively by avoiding out-of-stock conditions (Pisello, 2005:1). When looking at these mentioned benefits of a BI solution, it becomes clear that such a solution is critical for the success of a business, although its implementation does not come without challenges.

1.5 Motivation for the study

Sumathi and Sivanandam's (2006:145) argument that almost every BI project follows the 2:2:50 pattern, made the BI team in the above-mentioned telecommunications company realise that more research was needed to overcome the challenges mentioned above. Madsen (2010:8) states that 70% of the total cost of a BI solution is labour cost (see figure 1.5). Madsen's study was done globally in 1000 companies in different industries.

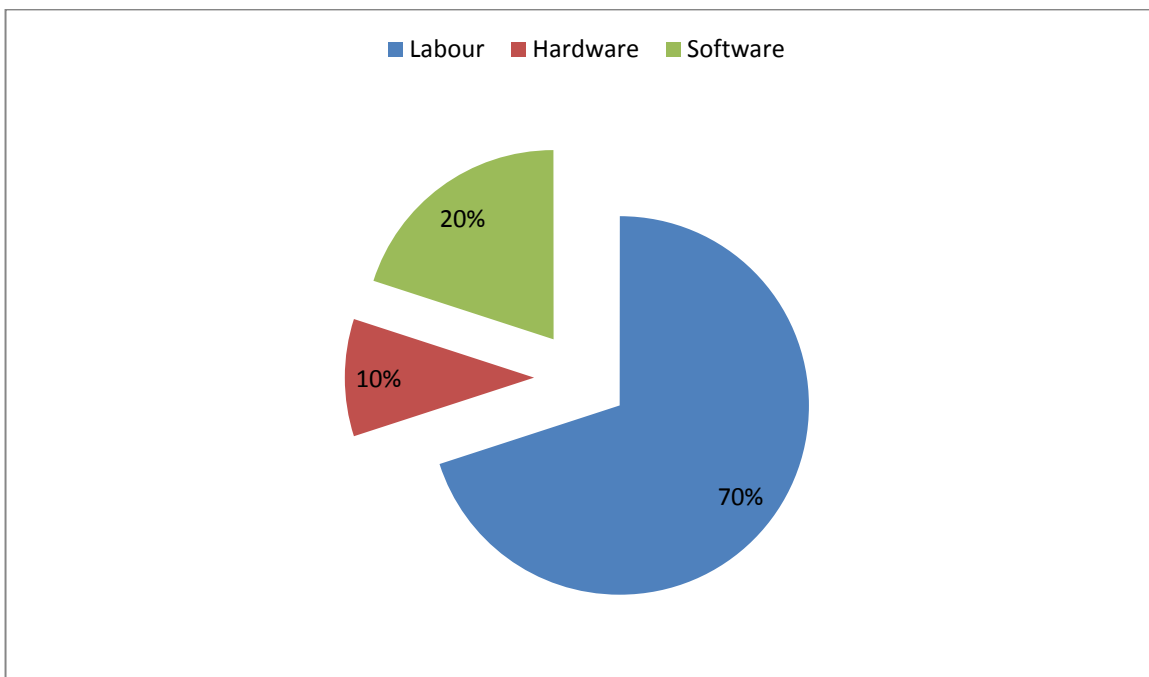


Figure 1.5 Implementation cost of BI (Madsen, 2010:8).

This means that only 30% of the cost of a BI solution is IT hardware and software cost. Madsen (2010:9) states that the only way cost of hardware and

software can be managed, is by looking at solutions of different vendors of BI tools. The company used for this study already had a BI tool. The cost-saving challenge, therefore, was on the BI labour side.

Hwang and Xu (2007:32) state that the failure rate of BI is 60% to 90%. For the telecommunications company in this study, failure was not an option. For the survival of the company, a BI solution was needed to manage and monitor the targets set by government. Meeting these targets posed various implementation challenges.

1.5.1 Business intelligence implementation challenges

Inmon (2003:1) argues that companies have different challenges causing delays in the implementation of a BI solution. One of these challenges is the inability to spend money on these implementations. This could be one of the main reasons for the slow uptake of BI. However, the benefit of a successful BI implementation is so significant that the academic research community should search for methods to improve the success rate of implementing BI (Inmon 2003:1). This includes recognising hot-selling items quickly enough to respond to customer demand and avoiding "out-of-stock" conditions.

The high BI implementation cost would obviously affect the implementation of BI solutions in large companies. According to White (2007:1), although companies overwhelmingly believe in the benefits derived from BI implementation, high implementation costs have prevented two thirds of them from seeking a comprehensive BI solution.

Companies in South Africa are challenged by the fact that a BI solution is competing with other Online Transaction Processing (OLTP) systems for budgets and training, preventing the creation of a BI culture in the company concerned (Du Plessis & McDonald, 2007a:218). The reason for this dilemma is that front offices are using OLTP systems and that the companies want to ensure that these systems are used properly and efficiently. Inadequately used OLTP systems could have a negative impact on customer satisfaction in the front offices. Customers do not want to wait in long queues or to be delayed and could in response go to the competition. It is therefore important to spend money on an OLTP system. Ramesh and Bhattiprolu (2006:367) state that OLTP systems host transactions which are the bread and butter of the company. BI is, however, meant to enhance OLTP systems by allowing companies to report on data gathered in OLTP systems.

A high percentage of available funds are spent on buying new hardware for OLTP and training staff on OLTP. This obviously results in a culture that supports and benefits the OLTP environment.

1.6 Problem statement

The implementation of BI in a South African telecommunications company is very difficult or in some instances impossible when considering the following factors:

- High cost of BI when following the software development lifecycle models available in industry

- Low level of BI literacy and culture in this industry, especially in South Africa
- The limited time for companies that are already doing business

1.7 Main research question

How can BI be implemented successfully in a South African telecommunications company, considering the high cost of implementing these kinds of solutions, the limited time for implementation, and the negative culture for BI that may exist in the company?

1.7.1 Secondary research questions

- 1) Where did data warehousing originate from?
- 2) Why is a data warehouse and BI solution needed?
- 3) How can a BI and data warehouse solution be implemented successfully in a company with a limited budget for it?
- 4) How can a BI culture that supports the optimal utilisation of a data warehouse and BI solution be instilled?
- 5) How can a complete enterprise data warehouse be implemented and grown, following an incremental approach?

1.8 Central theoretical statement

A BI solution in a telecommunications company in South Africa can be implemented using a new BI SDLC and a literacy and culture maturity model:

- The suggested BI development lifecycle model concentrates on the incremental implementation of BI, with the first focus on answering the critical business questions and the second focus on optimising the final solution
- The BI literacy and culture maturity model concentrates on creating a culture within the company that ensures the use of the solution, and maximises the value added to the business

1.9 Method of investigation

The research approach was a combination of qualitative research, with much of it relating to case studies and action research, reinforced by interviews done in the same telecommunications company, with feedback from peers at international conferences and the faculty of the European Conference on Information Systems (ECIS) Doctoral Consortium in 2010. The research has been published in four peer reviewed papers, the contents of which are presented in this thesis. The publications have been edited and revised for the purpose of this thesis. The research focus of particular papers will largely be treated in different chapters, as follows:

Chapter 1 & 6	DU PLESSIS, D. & MC DONALD, T. 2007a. Challenges in building and maturing of a telecommunications business intelligence solution in a developing country. (In Khosrow-Pour, M. (ed.) Managing worldwide operations and communications with information technology. Proceedings of 2007 Information Resources Management Association International Conference, Vancouver, B.C., Canada, May 19-23. p. 217-221.)
Chapter 3 & 4	DU PLESSIS, D. & MC DONALD, T. 2007b. The double wave data warehouse lifecycle model. Proceedings of the Eighth International Conference on Data Mining. Published in book format: Data mining VIII: Data, text and web mining and their business applications, edited by Zanasi, A., Brebbia C.A. & Ebecken, N.F.F. New Forest: WIT Press. p. 105-114.
Chapter 5	DU PLESSIS, D. & KROEZE, J. 2010. The double wave data warehouse hardware architecture for business intelligence implementation. The International Business Information Management Conference. Published in Proceedings of the 15th IBIMA Conference, Cairo, Egypt, November 6-7.

Chapter 7	DU PLESSIS, D. & MC DONALD, T. 2007c. Strategic framework to implement a telecommunications business intelligence solution in a developing country. Proceedings of the Ninth International Conference on Enterprise Information Systems, Funchal, Portugal, June 12-16. p. 227-232.
-----------	---

1.9.1 Literature review

The aim of the current study is to investigate possible ways of implementing BI successfully in a company, considering the high cost of implementing this kind of IT solution, the level of BI culture, and the urgency for the solution. A literature review was conducted and the literature available on BI was examined. This was compared with a case study in the above-mentioned company, and the results were refined by means of an action research process to develop a Software Development Lifecycle (SDLC) model for the implementation of BI. The effectiveness of the models was tested by a survey done in the telecommunications company used for the entire study.

Most of the literature and sources used were obtained from journals and scientifically verifiable research documents. Additional resources were found using online resources and a variety of academic databases. A qualified

librarian helped to find the relevant sources. Sections 1.9.2 to 1.9.4 will discuss the different research strategies followed for this study.

1.9.2 Case study

Yin (2003:1) defined a case study as “an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident.” Oates (2008:146) states that case studies can be used to develop a new concept, theory, framework or model. In this study the researcher analysed a historical case study, and based on the results, developed the DWDW (Double Wave Data Warehouse) lifecycle model. Oates (2008:144) states that case studies are different in their approach to time. A historical study examines what happened in the past, while a contemporary study examines what is happening presently. Cohen *et al.* (2005:160) states that the difference between historical research and contemporary research is that with historical research the data already exist, while a contemporary study needs to create new data.

In this thesis, a historical case study was used to create the DWDW lifecycle and the BI literacy and culture maturity model. Action research was then used to improve the DWDW lifecycle model. Action research will be explained later.

Rowley (2002:1) states that “typically case study research uses a variety of evidence from different sources, such as documents, artefacts, interviews and

observation, and this goes beyond the range of sources of evidence that might be available in a historical study”.

Yin (2003:215) states that there are three types of case studies:

- An exploratory study is used to define questions or the hypothesis to be used in a research project
- A descriptive study concentrates on the detailed analysis of a particular phenomenon, as well as its context
- An explanatory study compares theories and models in a case study to theories and models found in the literature.

This study used a combination of these three types of studies. An exploratory study was done to define the research question, a descriptive study explains the refinement process of the models, and the explanatory study compares the new models with models found in the literature study.

Yin (2003:215) states that “case studies are one approach that support deeper and more detailed investigation that is normally necessary to answer ‘how’ and ‘why’ questions”. The research question of this study categorises under the ‘how’ questions. That means that the research question for this study starts with ‘how’ (see section 1.7).

The following documents were available for the case study research:

- Business case
- Project plan
- Requirement documents
- Design documents of the BI solution
- System architecture document

These are internal documents and the company does not want to disclose them, and, therefore, the researcher did a survey to test the proposed models. Oates (2008:187) states that interviews are often used to test case studies and ethnographies. Chapter 3 will explain the case study in detail. The survey will be discussed later.

1.9.3 Action research

Baskerville (1999:1) describes action research as a research method that was already established in twentieth century and used in the medical and social sciences since then. Towards the end of the 1990s it became popular for research projects focusing on Information Systems. Chiasson *et al.* (2009:34) state that action research is a research strategy used to develop a solution whilst solving a practical problem and simultaneously creating new knowledge.

Action research was the research strategy used to improve the implementation processes developed during a case study research strategy. Chapters 4 and 5 will discuss this study in detail.

1.9.4 Survey

Surveys were done to test the success of the DWDW lifecycle model. Oates (2008:185) states that there are two kinds of surveys, namely:

- Questionnaires
- Interviews

Kumar (2005:126-127) states that the selection between an interview schedule and a questionnaire depends on the following criteria:

- The nature of the investigation
- The geographical distribution of the study population
- The type of study population
- The quality of data depends upon the quality of the interviewer
- The quality of data may vary when many interviewers are used

Interviews were selected because the target group was 23 people working in the same building. There are three types of interviews according to Oates (2008:186): structured, semi-structured and unstructured interviews. Structured interviews were selected where the same set of questions was asked of the different interviewees. The questions asked were a combination of open-ended

and closed questions. The reason for using open-ended questions was to give interviewees the opportunity to reflect on their real experience.

1.10 Contribution to the field of information technology

This study will illustrate that a BI solution can be implemented incrementally by using the proposed BI lifecycle model. It will also illustrate that BI literacy can be enhanced and a culture can be created which supports the use of the BI solution - by following the proposed literacy and culture maturity model for BI. The knowledge accumulated will supplement the current ICT body of BI knowledge.

1.11 Overview of the content of this study

The content of this thesis concentrates on creating a BI SDLC that can be used in a company with a limited BI budget and a limited BI implementation time.

The chapter layout is as follows:

Chapter 1: Introduction, problem statement, research goals, methodology

The background, motivation, goals and method of investigation are discussed in this chapter.

Chapter 2: The BI concept

The biggest source of a BI solution is the data warehouse, and, therefore, the data warehouse is discussed as part of the BI solution. An overview of the history of data warehousing will be given, explaining the origins of data warehousing and its evolution thereafter. The two concepts - data warehousing and BI - are explained in detail, before the architecture of a data warehouse and BI solution are reviewed.

Chapter 3: The development of the double wave data warehouse lifecycle model

Ponniah (2001:72) states that a data warehouse cannot be developed following the traditional software development lifecycles for OLTP systems. The development of data warehouses needs to be done by using a data warehouse SDLC. This chapter concentrates on discussing SDLC models that exist in the field of BI. Their advantages and disadvantages are discussed. A historical case study of the order process was used to create the DWDW lifecycle model.

Chapter 4: Refinement of the double wave data warehouse lifecycle

This chapter explains the action research strategy followed to finalise the proposed DWDW lifecycle model.

Chapter 5: Double wave data warehouse architecture

This chapter will explain the BI architecture currently used in industry. It will also explain why it was not suitable to use whilst implementing BI using the DWDW lifecycle model.

The new DWDW architecture, developed by action research, and with the detailed elements thereof, is introduced. The chapter also explains the benefits that this new architecture offers to companies with limited IT budgets.

Chapter 6: BI literacy and culture maturity model

Cooper and Reimann (2003:45) define a model as a tool to simply represent complex structure in order to understand it better.

Users sometimes have very little understanding of the need for a BI solution. It is often hard to understand what needs to be fulfilled in order to improve quality and to create the culture. This is because people have different perceptions of the concept. Some people have no understanding, whilst others have had bad experiences. Business people originating from companies where the BI project has failed previously, have a negative perception of BI. The BI literacy and culture maturity model, developed by means of action research, creates an understanding of the literacy and cultural level at which the company was positioned and presents recommendations for growth.

Chapter 7: Business intelligence strategic iterative and incremental framework results

Two new models have been introduced in the previous chapters. The focus was on putting a SDLC in place to facilitate the implementation of BI in a company with a limited budget and on creating a BI culture in a company that promotes the use of a BI solution in the company. This chapter presents these two models in a framework that guides the incremental implementation of BI in a company with limited budget, limited implementation time and a low culture for BI.

Chapter 8: Evaluation of the DWDW lifecycle model

This chapter discusses the interviews done to evaluate the DWDW lifecycle model, and the BI solution implemented using the DWDW lifecycle model.

Chapter 9: Summary, conclusions and recommendations

Conclusions on the implementation of BI using a strategic, iterative and incremental framework will be presented in this chapter. In addition, recommendations concerning the use of the framework will be made.

1.12 Conclusion

This chapter concludes the problem statement for this research project. Background was given on the importance of BI within a company. A research

question was derived from the problem statement, and was expanded to five secondary research questions. The research methodology used for this study was discussed, and the chapter concluded with a review of the chapter layout of this thesis.

Chapter 2 will give a thorough background on the history of BI and the different development techniques for BI currently used in industry.

CHAPTER 2

The BI concept

2.1 Introduction

The biggest component of a BI solution is the data warehouse, and, therefore, an outline of the history of data warehousing will be given. This will explain the origins of data warehousing and its evolution thereafter. The two concepts - data warehousing and business intelligence - will be explained in detail, before the architecture of a BI solution is discussed. It is important that these two concepts be discussed, because a model will be suggested to improve the implementation of a BI solution in subsequent chapters.

This chapter will also explain the design and build technique, processes and architecture based on the existing data warehouse lifecycle models. The design and build technique, processes and architectural approach used in this chapter, will be explained using the business dimensional lifecycle model (Kimball, 2006:34). This is important, because of the quest to get an understanding on whether the design and build technique, processes and architecture would be influenced when using the new SDLC to be developed.

2.2 What is data warehousing?

Inmon (2005:234) defines a data warehouse as “a subject oriented, integrated, non-volatile, and time variant collection of data in support of management decisions.”

According to Demarest (2002:1), a data warehouse “is a data repository that is needed to store information collected from diverse business systems for decision support”. According to Demarest (2002:1), it is important that the network elements of the company’s network should be considered when determining where the data warehouse server should be. It should be close to all big source systems, but the location of most of the end-users also plays a role. There will be large volumes of data flowing from the source systems to the data warehouse, but the end-user should also be able to extract data from the data warehouse (Dyche, 2000:136).

According to Oppel (2009:358), data in a data warehouse are stored as star schemas consisting of a fact table in the middle, surrounded by dimension tables. Some of the tables contain atomic data, which are data of the highest level of detail in the organisation - for example transactional data. Other tables contain aggregate data, which are summaries or totals representing longer periods of time. Fact tables and dimension tables are discussed in more detail later in this chapter.

2.3 History of data warehousing

Haisten (1999:132) believes that the data warehouse originated from a study done in the 1970s where researchers were seeking database architectural design guidelines. These researchers realised that there should be a different approach between a transactional database and an analytical database. The transactional database needs to be optimal for capturing transactions and should, therefore, be normalised. The database of the data warehouse should be de-normalised to simplify and improve query and analysis processing. Agosta (2000:208) states that a database is de-normalised for the following two reasons:

- To reduce the number of joins processed in average queries, to ensure good database performance
- For mapping the physical database structure more closely to the user's dimensional business model and to structure tables based on how the user will ask the business questions, so allowing common access paths to be tuned to relevant data, ensuring data are easily accessible and understood

During the 1970s and 1980s, processing power (big processors on servers and personal computers) and storage capacity (hard drives on servers and personal computers) were limited, forcing developers to restrict transactional system users from using any reporting on the system. This restriction was, however, not the final option, because a reporting solution was still a requirement, and this is where the data warehouse originated in 1991 (Inmon, 2005:289).

After the reporting solution was moved from the transactional system, it continued to grow until it became difficult for the hardware (servers) to process the load. Developers, therefore, started looking at new database design techniques that supported the de-normalisation of the data warehouse, not only to make it simple to use, but also to use minimal resources for delivering information to the data warehouse end-user (Inmon, 2005:289).

2.4 Difference between the OLTP system and a data warehouse

There are different design techniques between an OLTP system and a data warehouse. With an OLTP system, the main focus is on eliminating duplications, while the data warehouse focus is on simplifying the queries for end-users.

Table 2.4 (below) represents a customer dimension table in a data warehouse. In the table, Johannesburg, Pretoria and Cape Town are repeated several times. In an OLTP system, this customer table will be split up into two or three tables to avoid duplication. This process of elimination of duplications is called normalisation of the database. Normalisation also prevents different spellings of the same word, e.g. city and street names.

Customer ID	Customer name	Customer ID	City	Street	Street Num
1	K Nel	3452	Cape Town	Nelson Mandela	1
2	H van Wyk	1232	Johannesburg	Nelson Mandela	67
3	P de Wee	9825	Johannesburg	Nelson Mandela	33
4	I Cloete	1524	Johannesburg	Nelson Mandela	87
5	H Smith	4354	Cape Town	Main	7
6	J Kotse	2098	Pretoria	Schoeman	9
7	P Schoeman	4387	Pretoria	Schoeman	7

Table 2.4 Customer dimension table to demonstrate normalisation of a database.

2.5 Data warehouse design techniques

2.5.1 Dimension and fact tables

Kimball (2006:87) divided the tables of a data warehouse into two families: fact tables and dimension tables. Fact tables contain all numeric values that can be aggregated, for example monetary values like rand value (R3.56) and number of stock items. The customer number is also numeric, but is excluded, because it cannot be summed or counted. Fact tables are explained in more detail in section 2.5.3. Data in the dimension table describe the facts in the fact table. For example, the name and address of the customer that purchased a product is kept in the customer dimension table, while the number of products

and the total cost is stored in the fact table (see figure 2.5.1.1). Dimension tables are explained in more detail in section 2.5.2.

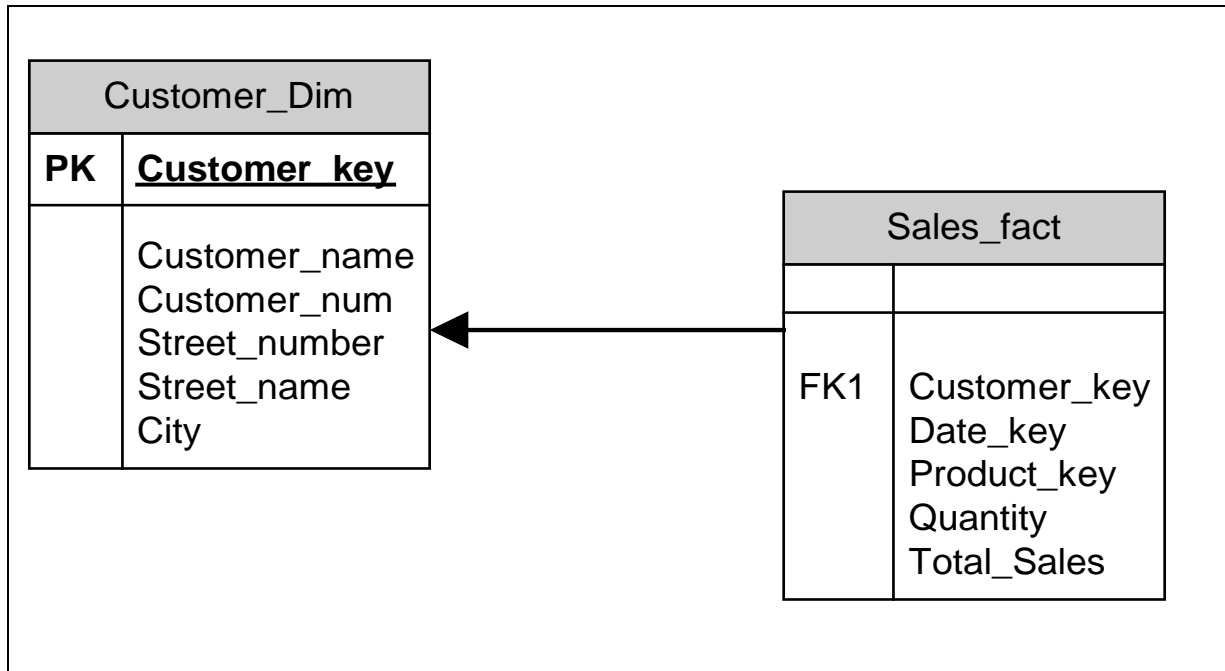


Figure 2.5.1.1 Differences between dimension table and fact table.

Examples of dimension tables are: (a) Date dimension, (b) Customer dimension and (c) Product dimension.

The fact table is normally surrounded by different dimension tables (see figure 2.5.1.2).

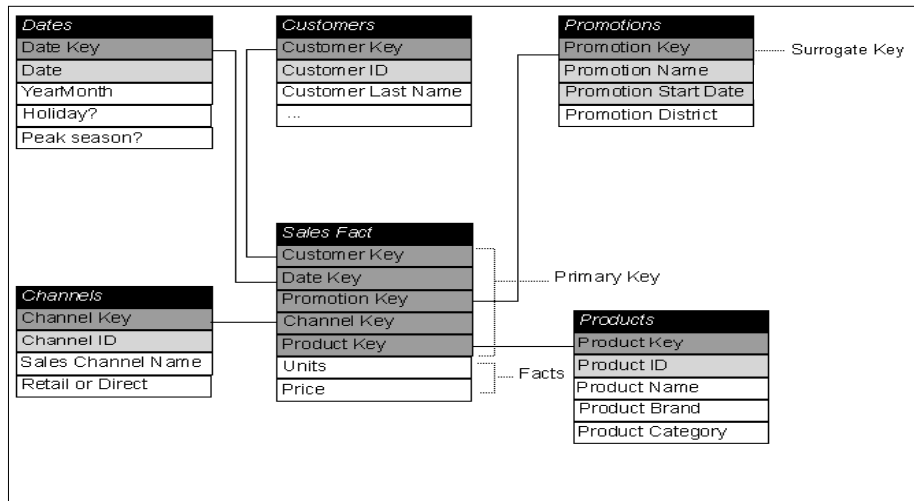


Figure 2.5.1.2 Simple data warehouse design (Kimball, 2006:223).

Dimension tables are connected to the fact table using primary keys in the dimension table and foreign keys in the fact table. All foreign keys in the fact table make up the primary key for the fact table. The relationships between the dimension tables and a fact table are one (in the dimension table) to many (in the fact table). The naming convention is normally as follows: the foreign key in the fact table and the primary key in the

dimension table are named after the dimension, so making maintenance much easier in the future. When the fact table is open, it is easy to see which foreign key links to which dimension table. Primary keys and foreign keys are replaced by surrogate keys (replacement keys) for the following reasons (Simsion & Witt, 2005:193):

- They are much smaller (two to four bytes instead of ten or more).
- Surrogate keys are not affected by business changes, for instance if a customer number is changing, the history of the previous customer number is retained because the history is kept in the surrogate key combination [e.g. Mr P.L. van Wyk's purchases for the last two years can be seen in table 2.5.1 below]. However, surrogate keys are normally concealed from the end-user and are illustrated in table 2.5.1 only for the argument. Date key, Customer key and Product key contain the surrogate keys. The results from the different tables are represented in different colours. When looking at the customer number column, it is clear that the customer number has changed. Customer name and ID number show that the history regarding the customer number has been kept and a new customer key has been created.

Date Dimension		Customer Dimension				Product Dimension		Fact Table
Date key	Date	Customer key	Customer number	Customer name	Customer ID	Product key	Product name	Quantity
12	08/09/2008	120	563283F	P L van Wyk	12345	297	Coke	1
54	10/03/2009	317	754391D	P L van Wyk	12345	297	Coke	4

Table 2.5.1 Surrogate key combination.

- They reduce memory utilisation of the data warehouse server because surrogate keys use integer, and integer is smaller in size than varchar or text data types.
- Surrogate keys simplify bitmap indexing (Stackowiak *et al.*, 2007:101). Chen (2002:221) states that bitmap indexing was created for data warehousing. The foreign key of a fact table normally has a bitmap tree to ensure that queries do not run through all the records of a huge fact table, but are sent straight to the data required for the query. This bitmap indexing is used to optimise query and analysis. Figure 2.5.1.3 (below) explains the bitmap for colours in a colours dimension table.

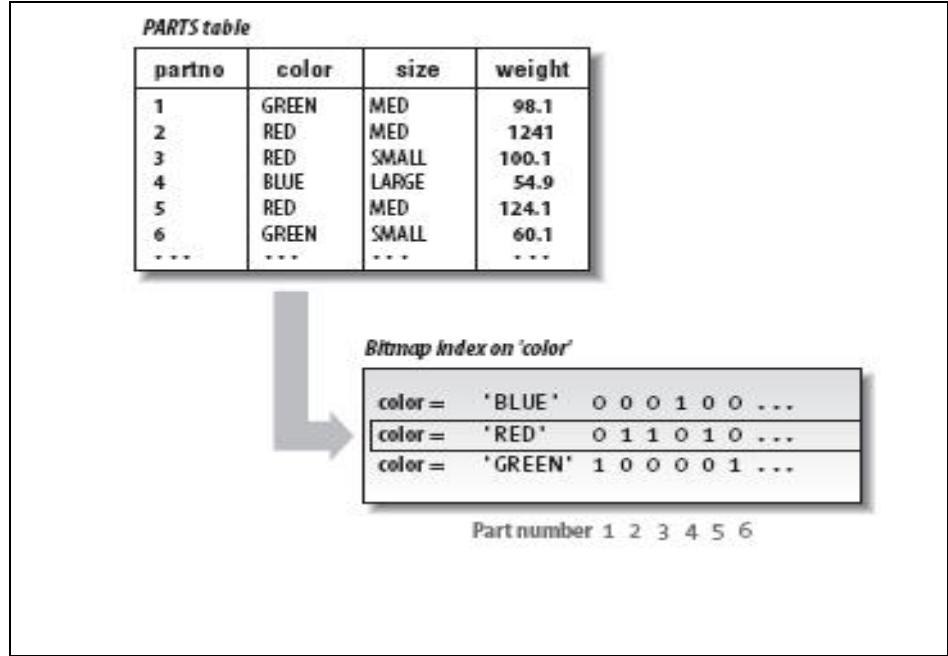


Figure 2.5.1.3 Explanation of bitmap index (Chen, 2002:221).

In a data warehouse production area, indexing is removed when doing the Extract, Transform and Load (ETL) updates because the insert command of SQL is slowed down by bitmap indexing. The ‘on’ and ‘off’ switch of the bitmap indexing is normally automated as part of the ETL code. The ETL will be explained in more detail later in this chapter.

2.5.2 Dimension types

Malinowski and Zimanyi (2009:186-188) state that there are two dimension types: slowly changing dimensions (SCD), and rapidly changing dimensions (RCD). For this study, the focus will only be on the SCD type. RCD is not discussed. Data captured in dimensions in the telecommunications company studied, are not changing often, and were therefore only making use of SCD.

The slowly changing dimension type consists of four sub-types (Malinowski & Zimanyi, 2009:186-188):

- **Type 1**

The rows of the table get replaced with the latest information. Attributes reflect only the most recent data.

Product Key	Product Desc	Department	SKU Number (Natural Key)
13	Hp Laptop	Finance	ABC9922-Z

Table 2.5.2.1 Type 1 SCD (before update).

If the data in the department column in table 2.5.2.1 change from finance to HR, finance is overwritten by HR and no history is kept regarding the fact that the laptop was initially ordered for finance (see table 2.5.2.2).

Product Key	Product Desc	Department	SKU Number (Natural Key)
13	Hp Laptop	HR	ABC9922-Z

Table 2.5.2.2 Type 1 SCD (after update).

The dimension will in future only reflect HR in the department column.

- **Type 2**

One of the primary goals in any data warehouse is to keep history. According to Wightman (2008:263), the type 2 SCD is the most commonly used dimension in a data warehouse, and will handle the change of the department as mentioned in the type 1 SCD, as follows (see table 2.5.2.3 below):

Product Key	Product Desc	Department	SKU Number (Natural Key)
13	Hp Laptop	Finance	ABC9922-Z
14	Hp Laptop	HR	ABC9922-Z

Table 2.5.2.3 Type 2 SCD.

When the data in the department column are changed from finance to HR in the source system of the data warehouse, then the type 2 SCD will add a new row to the table when loading the data into the data warehouse. The new surrogate key created with the insert of the new row will be written into the fact table against the foreign key of the date a record was created using this new row in the dimension table.

- **Type 3**

Kimball and Caserta (2004:213) describe a type 3 SCD as a dimension where history is kept by adding columns (see table 2.5.2.4 below). When the laptop moves from finance to HR, the column with finance is not overwritten. A new column is added, where the new department (HR) is stored.

Product Key	Product Desc	Old Department	New Department	SKU Number (Natural Key)
13	Hp Laptop	Finance	HR	ABC9922-Z

Table 2.5.2.4 Type 3 SCD.

- **Hybrid**

The Hybrid SCD is normally a combination of the above-mentioned SCD types. This sub-type is used where the business requires that if data in one column of the dimension table change, they are overwritten with the new information. However, if the other column is changing, a new row is added and the old row is kept to report on history.

2.5.3 Fact tables

Kimball (2006:64) identified two types of fact tables in a data warehouse:

- Cumulative fact tables

These tables keep historical data over a period of time. A good example is where information of product sales is kept in a fact table. Every item

on a sales order is a new record in the fact table. The quantity and sales price are summed to get the total sales per day, per month or per year, depending on the requirement (see table 2.5.3.1).

Date Key	Product Key	Customer Key	Price Key	Quantity	Rand Value Total
24	136	28	129	4	122.44
25	532	38	128	7	594.00
25	243	28	129	3	176.33
26	120	78	398	5	236.65

Table 2.5.3.1 Cumulative fact tables.

The Date keys 24, 25 and 26, when linked to the date dimension, will be three consecutive days. With a cumulating fact table, the quantities of consecutive days can be added as the total sales for these three days.

- Snapshot fact tables

This kind of fact table cannot be summed over time, because it is a snapshot of, for example, total store items for a specific date. Volumes could increase or decrease the next day, depending on items that came in or went out of the store. Facts in this fact table are semi-additive, meaning they cannot be summed across different dates. They can, however, be summed across different stores for a specific date. In table 2.5.3.2 (below), date key 25 appears twice with different numbers in the store key column, meaning that it is the same day with different stores, and therefore quantities and rand values can be added for these two rows (Kimball, 2006:73).

Date Key	Product Key	Store Key	Price Key	Quantity	Rand Value Total
24	24	28	129	5	555.55
25	24	28	128	1	111.11
25	24	38	128	3	176.33
26	73	78	398	5	236.65

Table 2.5.3.2 Snapshot fact tables.

2.6 The importance of business intelligence

Liautaud and Hammond (2001:24) state that BI is about finding ways to use intelligence to cultivate customer loyalty, drive profits and concentrate on finding ways to outsmart competition. They describe the relationship between data, information and intelligence, as follows:

- Data are raw and unadorned. For example, a single record in a transactional database records the purchase amount of an item from a website by a consumer at a particular location.
- Information is data that have been cleaned, synthesised and aggregated. Data are transformed into information by sorted, filtering and interoperating it.
- Intelligence is derived from information and underwrites an organisational state that may be characterised as collective intelligence. Intelligence results from a full appraisal of information, past actions and options.

Liautaud and Hammond (2001:23) further believe that companies basically make two kinds of decisions: the big strategic decisions made by the few higher level executives and the many small decisions made by all the employees on a daily basis. For the strategic decisions a lot of information is gathered, deep analysis is conducted, and options are carefully reviewed. Companies usually invest what is needed to gather the required intelligence. Liautaud and Hammond (2001:23), however, demonstrate that a company's

performance is largely determined by all the smaller decisions made every day by its employees. BI is the solution supporting these two kinds of decisions.

2.7 Business intelligence architecture

Inmon (1995:125) divides the data warehouse architecture into the following six components (see figure 2.7 below): (a) Data Source, (b) Data Acquisition, (c) Enterprise Data, (d) Data Enhancement, (e) Data Mart and (f) Data Access. These components are discussed below:

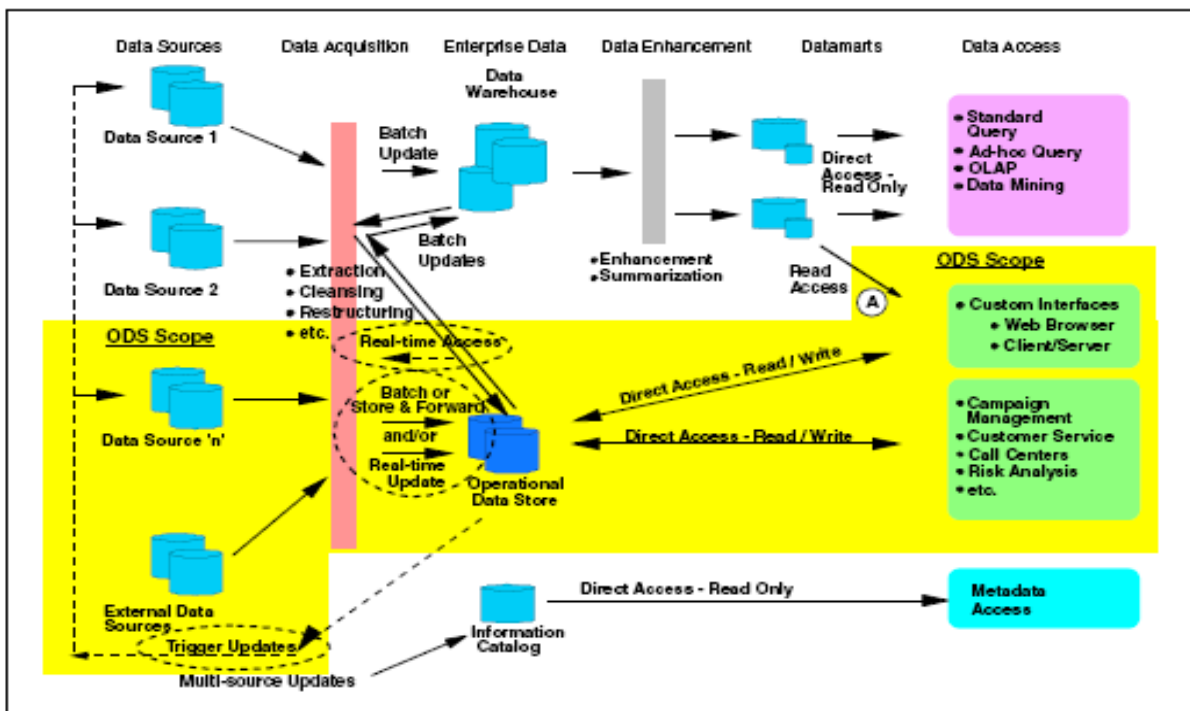


Figure 2.7 Data warehouse and BI architecture (Baragoin *et al.*, 2001:42).

2.7.1 Data source

Ponniah (2001:263) believes that when considering building a data warehouse, it is important to understand the requirement well enough to identify the correct source systems. The OLTP systems, where the day-to-day transactions are captured, are referred to as the source systems. Source systems can be Enterprise Resource Planning (ERP) systems, laboratory systems, production systems, sales systems, and so forth (Jones, 2008:15-34).

2.7.2 Data acquisition

In the data acquisition process, data are cleaned, transformed and moved into the data warehouse or data mart (Cologon & Cohen, 2008:589-590). The data mart will be explained in more detail in section 2.7.5. This is the most time-consuming process when building a data warehouse (Imhoff *et al.*, 2003:399), and is known by the data warehouse professionals as the Extract, Transform and Load (ETL) process (see figure 2.7.2.1). It is developed by making use of the ETL tool. The first part of the process concentrates on extracting the data from the source system.

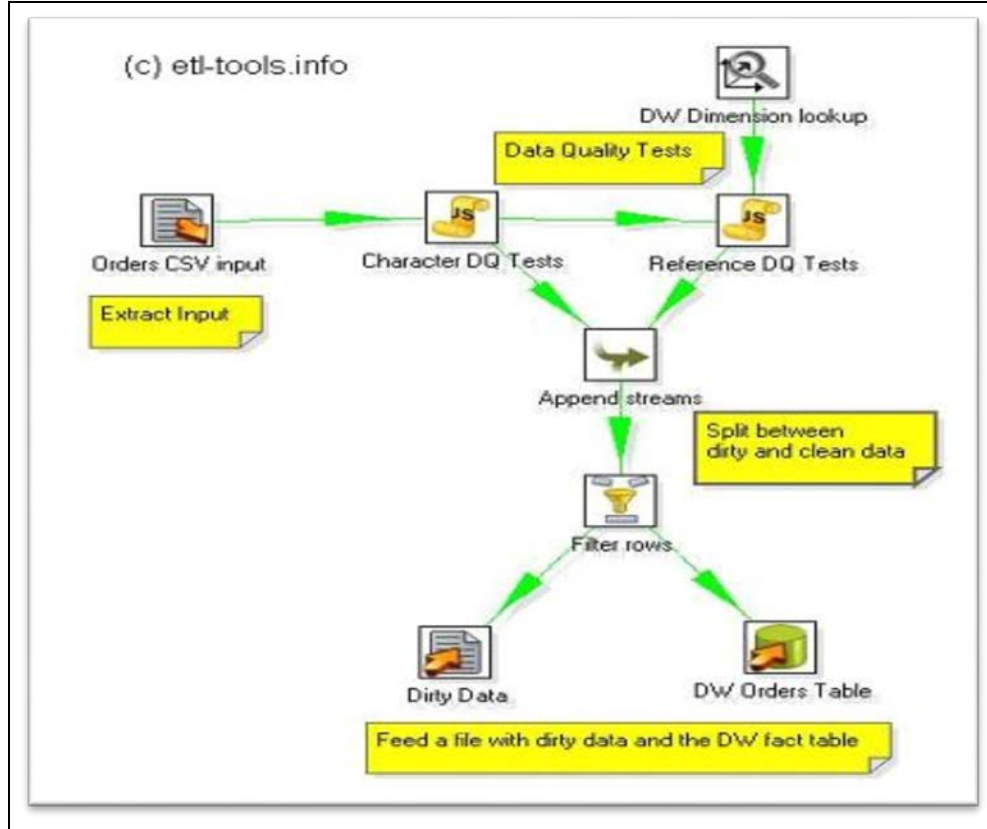


Figure 2.7.2.1 The ETL process (Baragoin *et al.*, 2001:211).

The data are temporarily stored in a database called the staging area. The staging area is where data are transformed and cleaned before being loaded into the data warehouse (Imhoff *et al.*, 2003:399). The cleansing part of the process concentrates on cleansing exercises such as changing data types, while the transform part of the process transforms data that were in, for example, a third normal form - to a de-normalised structure. To put this process into production, it is scheduled to run automatically at an agreed frequency rate (Baragoin *et al.*, 2001:123).

Kimball (2004:14) refers to the steps of an ETL process as:

- Extract
- Cleans
- Conform
- Deliver

See figure 2.7.2.2 below.

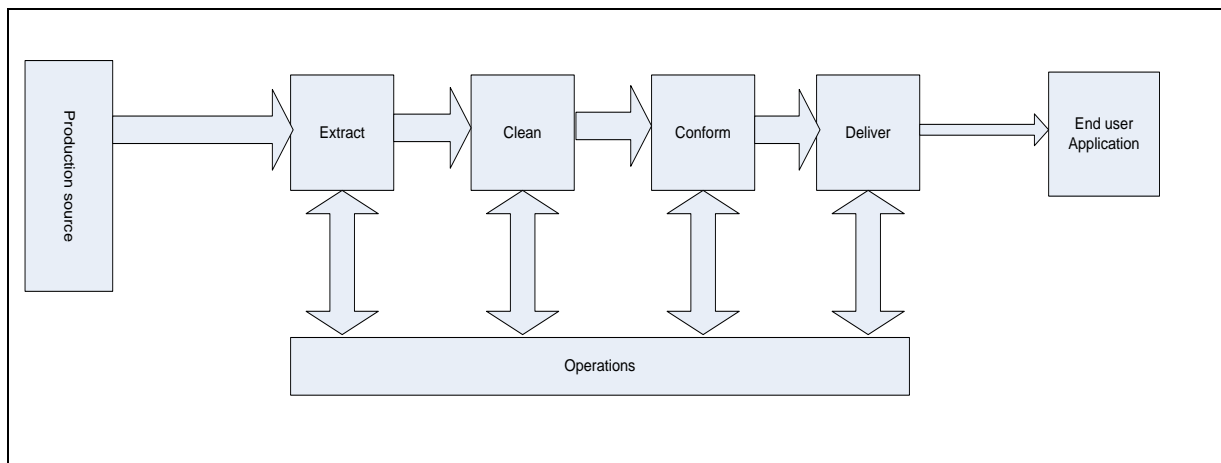


Figure 2.7.2.2 Data flow tread (Kimball, 2004:14).

To load the dimension tables is straightforward. One writes a “distinct select query” on the customer table in the staging area, and the result is loaded onto a customer dimension table. When loading facts, the keys of the different dimension tables should be looked up and written into the fact table along with the different facts (Kimball, 2006:240).

2.7.3 Enterprise data

This part of the architecture consists of the data warehouse and the operational data store (ODS) (Kimball, 2006:240). The database of the data warehouse is de-normalised, to simplify the queries for the data warehouse user.

The data warehouse is used for strategic information, which normally comes from trend analysis. Therefore, there is no need for it to be loaded with small intervals, especially during office hours. The data warehouse is normally loaded during the night (Kimball, 2006:240). The ODS is used for operational information.

Baragoin *et al.* (2001:13) compiled a table (see table 2.7.3 below) to explain the differences between ODS and a data warehouse.

ODS	DW
Data of high quality at detailed level and assured availability	Data may not be perfect, but adequate for strategic analysts; data do not have to be available with small latency
Contains current and near-current data	Contains historical data
Real-time and near real-time data loads	Normally batch data loads
Mostly updated at data field level (even if it may be appended)	Data are appended, not updated
Typically detailed data only	Contains summarised and detailed data
Modelled to support rapid data updates (3NF)	Variety of modelling techniques used, typically 3NF for DW and dimensional for data marts, to optimise query performance
Transactions similar to an OLTP system	Queries process, larger volumes of data
Used for detailed decision-making and operational reporting	Used for management reporting and long-term decision making
Used at the operational level	Used at the managerial level

Table 2.7.3 Differences between ODS and DW (Baragoin *et al.*, 2001:13).

It is important that the information of the ODS is as close as possible to real-time. The ODS, therefore, needs to be updated with small latencies. The design of the ODS should accommodate these small latencies, because there is no time to do big transformations and look-ups. All information of a specific subject is stored in the same way it was stored in the OLTP system (see figure 2.7.3.1 below). It is not optimal for query and analysis, but the main aim is to move the information to the ODS as soon as possible. The ODS, however, does not carry

history for more than one day. The queries on the ODS, therefore, do not have a lot of data to handle, and the process is still very fast (Inmon, 2005:143).

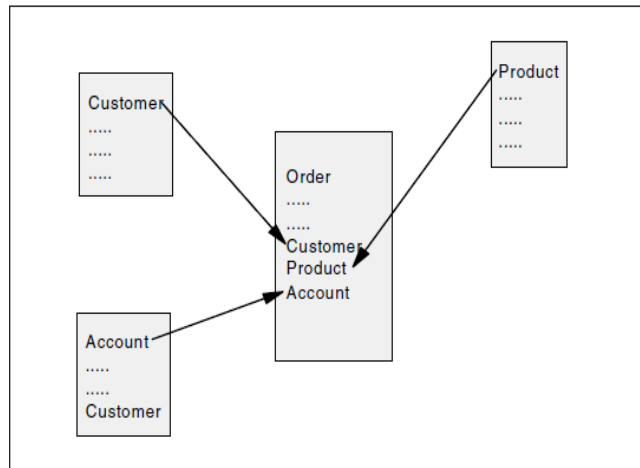


Figure 2.7.3.1 The ODS design (Inmon, 2005:143).

2.7.4 Data enhancement

With data enhancement, the focus is on making the BI solution user-friendly. A Graphical User Interface (GUI) is implemented to hide the complexity of the underlying database design. The intention is to make use of the BI solution simpler for the end-users (Inmon, 2005:123).

2.7.5 Data mart

The design of the data mart is exactly the same as a data warehouse. The only difference is that a data mart is department specific, while a data warehouse contains information across departments (Kimball, 2006:128).

2.7.6 Data access

The data access area of the architecture represents the tools used by the end-user to extract and analyse information stored in the data warehouse and ODS. These tools fall within the general rubric of ‘business intelligence tools’.

BI tools include the following families of tools (Kimball & Ross, 2002:218):

- Query Tools

These tools allow the user unfamiliar with SQL (Structured Query Language) to write a query using a GUI against the warehouse and get a result.

- On-Line Analytical Processing (OLAP)

These tools enable users to look at the data from different ‘angles’. They use a multi-dimensional database referred to as a ‘data cube’. Abello *et al.* (2002:554) state that a data cube is a data structure used by OLAP tools to simplify the analysis of multidimensional data sets. There are three kinds of OLAP tools:

- Relational OLAP (ROLAP)
- Multi-Dimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP, a combination of MOLAP and ROLAP)
- Data Mining Tools

These tools automatically search for patterns in the dataset coming from the data warehouse or any other source. These systems often consist of complex statistical formulae. Unlike a data mining tool, an OLAP tool answers existing questions in a company.

2.8 Conclusion

In this chapter a literature study was done on BI and its role as a tactical and strategic tool. The chapter further concentrated on the data warehouse in terms of it being the biggest source of BI. The data warehouse was discussed, focusing on the history and architecture thereof. The history explains the origin of the data warehouse. The difference between OLTP systems and a data warehouse design, were discussed. Functions of components of the BI architecture were discussed in detail, to create a good understanding of the architecture currently used in companies. Understanding the current architecture will simplify the introduction of a new architecture in chapter 5.

Different data warehousing design techniques were also introduced. These techniques will be used in subsequent chapters to explain a new SDLC developed for implementing a data warehouse in a telecommunications company with a limited BI budget.

CHAPTER 3

The double wave data warehouse lifecycle model

3.1 Introduction

In chapter 2 the history, design techniques and architecture of a BI solution were discussed. The historical review examined the origins of data warehousing. Arguments about the feasibility of reporting from an OLTP system were given, which led to the conclusion that a BI solution is needed in business today.

The main process for developing software is called a system development life cycle (SDLC) (Unhelkar, 2008:46-47). This chapter will discuss the SDLC which was predominantly used in the telecommunications industry in South Africa previously. The BI team in the company used for this study, discontinued focus on the known SDLC because of the challenges faced in chapter 1 (limited time and budget to implement a BI solution).

“The order process BI case studies” from the company have been analysed to develop the new SDLC for BI (The DWDW lifecycle model). This chapter will explain the case study research process which was followed to deliver the DWDW lifecycle model.

The design techniques and architecture of BI - as discussed in chapter 2 - will be used to explain the challenges with the existing BI SDLCs, and to demonstrate the functionality of the DWDW lifecycle model.

3.2 The case study research methodology

Oates (2008:141) states that a case study focuses on a specific instance of a business or other initiative and studies it in depth by using a variety of data generation methods, such as observations and document analysis. The main aim of this is gaining as much insight as possible into the 'life' of the specific case. This research process started by investigating the one BI project that was implemented successfully in the above-mentioned company. The 'order process case study' was used to identify the reasons for the success of the project. This investigation revealed that this BI solution was built following different implementation processes. These processes were studied and a new model, called the DWDW lifecycle model, was created.

3.2.1 Data generation methods

The data creation method used for this study was document analysis, interviews and observation. The researcher started by documenting the entire case study as it was documented in the project documents. Documents used for this purpose were:

- Project scope
- Business case

- Requirements
- Design
- Project plan

3.3 Order process case study

The case study was divided into two discussions. The first concentrated on the challenges experienced with existing SDLCs, while the second concentrated on the development of the DWDW lifecycle model.

3.3.1 Challenges with existing BI SDLCs

3.3.1.1 Agile BI

Collier (2011: 5) states that before using agile BI processes, the following two misconceptions need to be considered:

- Successful agility requires that all shareholders (BI and business people) undergo a mind-set change. The agile processes are completely different, and people should have a lot of discipline to ensure the successful completion of a BI project.
- Collier (2011:5) states that “agile doesn’t always mean faster project completion. Agile practices do steer teams to focus on the high-value and riskiest features earlier”.

The company needed a solution that could guarantee quick delivery on information requirements. Agile processes, on the other hand, do not address

the challenges of companies with a limited BI Budget. Agile processes, therefore, were not considered to build a BI solution for this company.

3.3.1.2 Inmon's CLDS

Inmon (1995:1) believes in the use of a data-driven method to develop a BI solution. He calls this method the CLDS (the reverse of SDLC), because this method, unlike other SDLCs, begins with data and ends with requirements. This method starts with building a central data store for one subject area which is populated from OLTP systems. As the analytical ability of the new data warehouse is discovered, demand for an integrated data store for another subject area grows. This process then repeats itself until a complete data warehouse has been developed.

Building a BI solution using this CLDS often takes years (Han & Kamber, 2006:67). One way to convince companies to invest in a BI solution is by adding evident value as you build a solution. From the case study, the argument was that, with the CLDS, value is only added right at the end of the project, and therefore, it was not suitable for this company.

With Kimball's (2006:34) Business Dimensional Lifecycle Model or BDLM (see figure 3.3.1.3), the data warehouse project starts with determining user requirements. Kimball (2006:34) states that a clear understanding of business requirements is the most important factor contributing to successful use of this model.

The limitation of the BDLM was that it was aimed at implementing a complete BI solution by modelling the different business processes. The answer to critical business questions cannot be dealt within a day or two. The limitations of the model are discussed in more detail in section 3.3.1.3.

3.3.1.3 The business dimensional lifecycle model

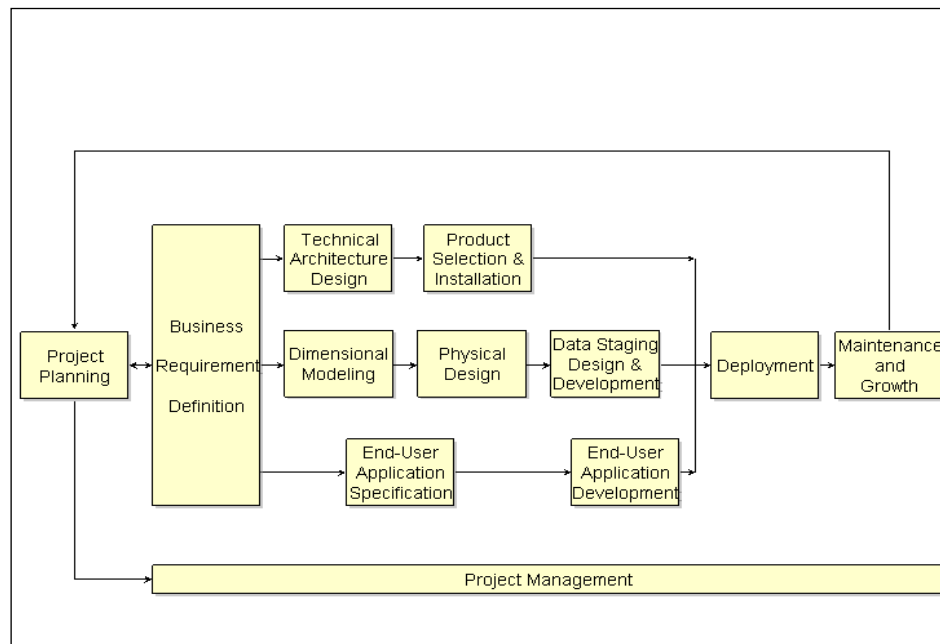


Figure 3.3.1.3 The business dimensional lifecycle model (Kimball, 2006:34).

The company used for this study deviated from using the BDLM (Kimball, 2006:34), because it was difficult to use in companies battling with insufficient funding for BI and with limited time to implement it. This had previously contributed to the failure of BI implementation in the company, and the

implementation process was further complicated by the prevalence of low literacy levels and a poor BI culture. BI literacy and culture will be discussed in chapter 6. The BDLM consists of seven successive phases, and from the analysis of the case study, the following challenges were experienced in the respective phases.

3.3.1.4 Requirement phase

Siu (1997:166) states that the requirement phase is crucial in the success of the software development project, because, if not managed well, requirement tends to grow to a level where it becomes difficult to complete in one project. Poor management of requirements can create expectations that cannot be fulfilled in a single project. This can result in a situation where BI end-users feel that BI has failed to fulfil their business requirements.

The business requirement phase is represented by the large vertical rectangle in figure 3.3.1.3 (above) and normally takes up a fair portion of project time. This is because during the first iteration of this phase BI is not always known to the BI end-users. This means that a lot of time can be spent educating end-users about BI before the requirements can be collected.

Pekowsky (2004:38) stated that splitting up the requirements into more than one project helps software implementation teams to manage more substantial software project requirements. By using this approach people will not develop a negative perception towards BI, in the sense that they will understand that project delay does not mean project failure (Du Plessis & McDonald,

2007a:217). Based on the importance of the business requirements and the quest for the management of user expectations, in reality it becomes difficult to implement a BI solution in one big project, as suggested by the BDLM (Kimball, 2006:34).

3.3.1.5 Dimensional modelling and physical design

Westerman (2001:63-64) states that data modelling is the process whereby business questions are used to determine the dimensionality required in the data warehouse, and to build that requirement into a data model.

It is, however, very difficult to present the dimension models to business people, because when they look at data, they want to see spreadsheets answering their business questions. They normally do not understand dimensional models well enough to give approval for the build process (Schrage, 2000:88).

In fact, business people prefer to drive urgent business questions with the hope that they will be answered quickly. Unfortunately, the BDLM concentrates on the business process as a whole, and not on certain critical business questions (Brackett, 1996:312). This means that more money is needed to complete the entire process, which could further conflict with expectations.

3.3.1.6 Technical architecture design and product selection/installation

Martin (2008:46) states that this part of the model concentrates on the process of putting hardware and software in place, and it is normally only part of the first iteration of the SDLC. The technical architecture prescribed by the BDLM (Kimball, 2006:34), is meant to support a complete data warehouse implementation. A complete data warehouse implementation is expensive and needs a sound budget. The technical architecture is, therefore, not suitable for an incremental implementation approach.

3.3.1.7 End-user specification and development

According to Gordon (2008:180), this phase of a SDLC concentrates on designing and building the front-end of the BI solution. The front-end of the BI solution gets used by business people, and they are, therefore, more willing to participate in this design and build, than in the data warehouse design (Kimball, 2006:34).

3.3.1.8 Data staging design and development

Martin (2008:49) stated that this part of the SDLC concentrates on the design and build of the ETL processes. This part of the model is labour intensive and very time consuming, and, therefore, it further delays answering business questions.

3.3.1.9 Deployment

Martin (2008:46) states that, during this phase, ETL processes and Graphical User Interfaces (GUI) that have been designed and built, are put into production. This phase sometimes only happens two to three years into the process, meaning that value is only added at this time. In the company concerned, a SDLC was needed where value can be added within a much shorter period of time.

3.3.1.10 Maintenance and growth

The last phase of the BDLM is ongoing maintenance and growth of the data warehouse. Anderson and Kerr (2002:138) state that the needs of customers are changing over time, and, therefore, business is forced to change. If business is changing, data and information change too. Ongoing maintenance and growth, therefore, require alignment between the customer's needs and business information.

3.3.2 Difficulties with the business dimensional lifecycle model

When examining the challenges faced when using the BDLM of Kimball (2006:34), the need for a different SDLC becomes apparent.

Kimball & Caserta (2004:437) state that the right information is not the only challenge. The information needs to be delivered to the business decision-makers at the right place and at the right time for maximum business value.

Where a company is still busy building the data warehouse, it is impossible to create this maximum business value. The fact that information is not available instantaneously or with a very small lead-time, can make it difficult for business people to make informed decisions (Lavin, 1992:183).

For example, if a telecommunications company needs to make a decision on the viability of installing a prepaid landline telephone platform, the following information is needed to make an informed decision:

- How many customers apply for a post-paid service, but fail credit vetting?
- Is network infrastructure available at the addresses of these customers?

If this information is not available to the decision-makers, it becomes difficult or impossible to make an informed decision. If a data warehouse project exists in the company, it will also not help to change the scope of the project by providing this required information upfront. This is because following the steps in the SDLC in figure 3.3.1.3 took several months before the required information became available.

The problem with the BDLM in figure 3.3.1.3 is that it takes time to fulfil the needs of the business-critical information requirement (BCIR). Information in business is often needed for a 'window of opportunity' or legal compliance, which cannot wait for a project to finish (Du Plessis & McDonald, 2007b:105). When this need arises, there is often not enough time to go through a full SDLC. The requirement can also sometimes be a once-off with the information

not required again, as with the information required to decide on the implementation of the prepaid platform, discussed above. The SDLC should enable the business to make an informed decision within a day or two, but should also allow the data warehouse and BI expert to build the requirement into the formal structure, if this is needed for on-going decision making.

Section 3.3.3 (below) describes the development of a new SDLC. A case study of a BI project implemented to manage the order process was analysed, and the results were used to develop the DWDW lifecycle model.

3.3.3 Development of the DWDW lifecycle model

After the sales team went out and sold telephone services as prescribed by government, as discussed in chapter 1, it was important to track the orders through the order system. Not only had the number of orders in the order system increased rapidly, but it was also important to get the orders through the system as soon as possible, to ensure that the services were provided to the customer to achieve the targets set by government.

While implementing the BI solution, the BI team realised that it was impossible to use any of the existing SDLCs, because all of them were very complex to follow, and there was no time to wait for a BI solution to be finalised. There were two objectives for this BI project. Firstly the company needed the data almost instantaneously to measure the targets set by government. The second objective was to have an optimised BI solution for on-going reporting and analysis. The implementation of the first objective

(instantaneous data) had a limited time scope, while the implementation of the second objective (optimisation of BI solution) needed a lot of time. The project was, therefore, broken up into two phases. The first phase concentrated on providing the information needed to measure the company against the targets set by government, and then after the BI solution was running satisfactorily, the optimisation was done.

In order to stay in line with all the existing software development projects, the team decided to use the waterfall model of Boehm and Papaccio (1988:103) as a baseline to create the software development steps of the different phases of the project. The order process will be discussed in section 3.3.4 (below), to give a clear understanding of the business requirement. Thereafter, there is a discussion on how the DWDW lifecycle model was developed from this case study.

3.3.4 The order process

The order process consisted of the following steps:

1. Capturing paper order on the system
2. Credit vetting of new customers
3. Allocating access network
4. Allocating core network

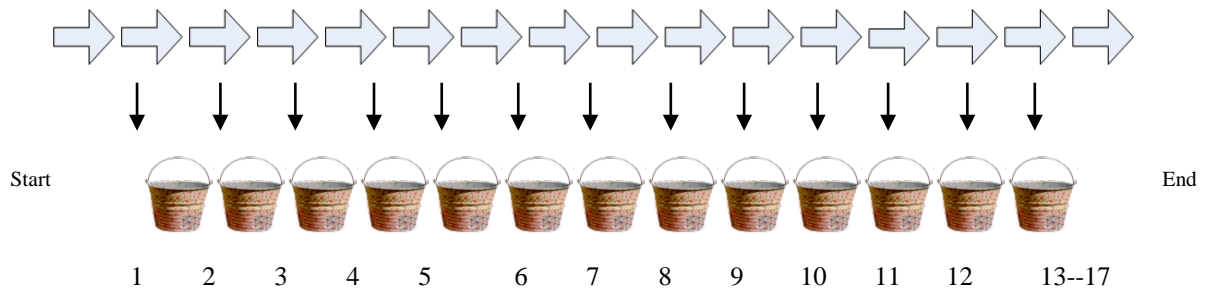
5. Receiving connection fees
6. Order sent to exchange to create a new port and telephone number
7. Order sent to main distribution frame for cross-connection
8. Order sent to technical team for installation
9. Finished order sent to billing team to start the billing process.

The nine step process would have been sufficient in the perfect world, but all orders could not go through the process without getting stuck somewhere in the process. These orders were dropping off the 'order pipeline'. Therefore, the orders that got stuck in the system were grouped into - what were termed - buckets. Some examples of these drop-off conditions, were as follows:

1. Sometimes addresses were incorrect on the paper order, and the capturing of the order could not be finished
2. The information on the network was incorrect and no network existed for the address
3. The customer was not available for the installation of the service.

To make provision for all of these conditions, the process below was designed and all the different drop-out conditions were called buckets. Figure 3.3.4.1, below, represents the business requirement.

When the requirement occurred to track the orders through the OLTP system, the orders were already in the system. That meant that it was a business-critical information requirement. There was no time to go through the data warehouse development lifecycle. A solution was needed immediately. The BI team consequently designed the front-end solution for the order-tracking system, and called it the bucket system.



Bucket number	Bucket description
1	Street address or X and Y coordinates not completed
2	Postal or billing address not completed
3	Wait for credit vetting results
4	Credit vetting score lower than required score, send letter to customer
5	Waiting for payment
6	Waiting to allocate access network
7	No distribution side access network available
8	Waiting to allocate exchange side network
9	No exchange side access network available
10	Waiting for cable repairs
11	Waiting for allocations of a port on the exchange
12	Waiting for telephone number to be allocated
13	No core network available at the exchange
14	Waiting to run cross connection on the main distribution frame
15	Waiting for installation
16	Customer not at home
17	Installation completed, wait for billing

Figure 3.3.4.1 Bucket system.

The BI team provided the information required in a flat file (see figure 3.3.4.2 below), because there was no time to go through one of the existing SDLCs. The ETL process transformed the different conditions into buckets. That meant that depending on the order's condition, a bucket number was placed next to the order in the flat file. This bucket number was known as the current bucket, and the snapshot date was known as the date that the order went into the particular bucket. This was only for the first run of the flat file. For the second run, there was a current bucket and a previous bucket, and the number of days spent in the current bucket. The number of days in a bucket was calculated day-by-day from the first snapshot date on which an order went into a specific bucket. This flat file was accessed by an Excel pivot table. The advantage of this flat file was that it could be used while the bucket process was refined. Changes that occurred because of the refinement were easy to handle because these changes did not require a total redesign of the dimensional model. It required only small changes to the ETL and the flat file, which could be done in minutes.

Order_bucket_flat_file	
	Date Order_number Order_type_code Order_type_desc New_Bucket Old_Bucket Days_in_buket Total_days_outstanding Telephone_number Port_num Customer_num Customer_name Strip_line_number Region_name Town_name Suburb_name Street_number Street_name Zip_code DP_name X_Y_coordinates DP_size DP_num SDC_Dside SDC_name

Figure 3.3.4.2 Flat file table for order bucket reporting.

This bucket process was helping business to concentrate on orders that were the longest in a particular bucket. The orders were also divided amongst the different sections. Core network sections were responsible for core network buckets, the access network sections were responsible for the access network

buckets, and finance was responsible for the financial buckets, and so forth. The responsible section would log into the order system and attend to the order. This bucket system was not only used to follow the orders through the order system, it was also used in the call centre for customers phoning in to find out where their orders were in the process.

After this bucket system had been in place for a month and the BI team was happy that all information needed was included and correct, the BI team started to model this system into a proper dimensional model and BI solution (see figure 3.3.4.3 below). The advantage of this quick BI solution was that the requirement was now understood very well and there was a proper interim solution in place that fulfilled the immediate need. After the interim solutions were signed-off, the flat file was designed into a proper dimensional model, and the ETL-process was now changed to load the tables of the dimensional model (see figure 3.3.4.3).

Designing the star schema was easy, because all the information needed was already in one spreadsheet. There was only one fact element that the business people were interested in, and that was the number of orders. Therefore, the only fact table was the order fact. When looking at the different characteristics of the two different kinds of fact tables, as described in chapter 2, it becomes clear that the order fact table is a semi-additive snapshot fact table, because it presents the orders at a certain point in time and cannot be added over time.

Identifying the different dimension tables was also relatively easy, because the remaining fields were grouped into a logical dimension table. All the fields

forming part of the address were clustered into an address dimension. All the dimension tables were type 2 SCD.

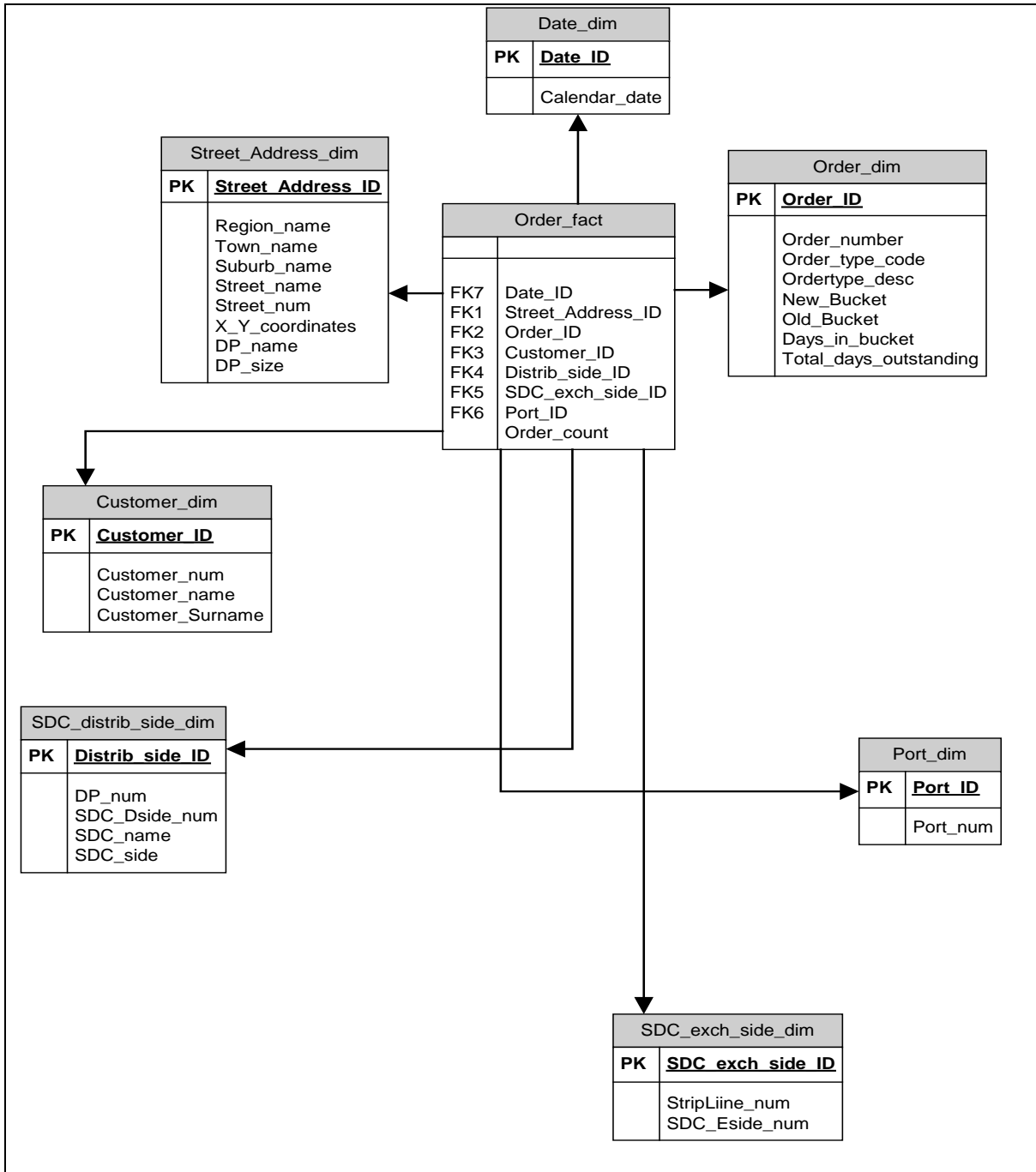


Figure 3.3.4.3 Dimensional model developed for the order bucket system.

All this was done without the user knowing that it had happened.

3.4 Data analysis

Cohen *et al.* (2009:316-329) states that, when analysing a case study, care should be taken that the most relevant events are used for examination. The event deemed relevant to this case study was the importance of quick availability of information. The other event was the optimisation of the solution. These two events were also considered for the data categorisation in section 3.4.1 below.

Kumar (2005:114-176) states that researchers should avoid concentrating too much on pre-conceived ideas or theories in analysing the data when using the case study research strategy. Reinard (2007:236-296) asserts that this is the data analysis approach, where the theory emerges from the data and which means that the theory is grounded in the data. Cohen *et al.* (2009:316-329) state that grounded theory is an inductive approach and not a deductive approach. These two approaches will be discussed in more detail in Chapter 8.

3.4.1 Data categorisation

Goulding (2002:74-76) states that open coding prescribes that, when labelling data units, the labels should not come from the literature, but from the data. The labels derived from the case study used for this research, were the two phases to implement a BI solution for the order process (see section 3.3.4). The BI team divided the implementation of the BI solution into two phases in order

to address the limited time available to implement the BI solution, and parts of the waterfall model were used in both phases to implement the model. These phases were called waves of the DWDW lifecycle model. Figure 3.4.1, below, was used to represent the two phases, and the detailed step involved in each step represented the results coming from the case study. This newly developed model was then named the DWDW lifecycle model. Data around the two waves were labelled into the different steps of wave 1 and wave 2. Sections 3.4.2 and 3.4.3, below, will discuss the detail involved in each step.

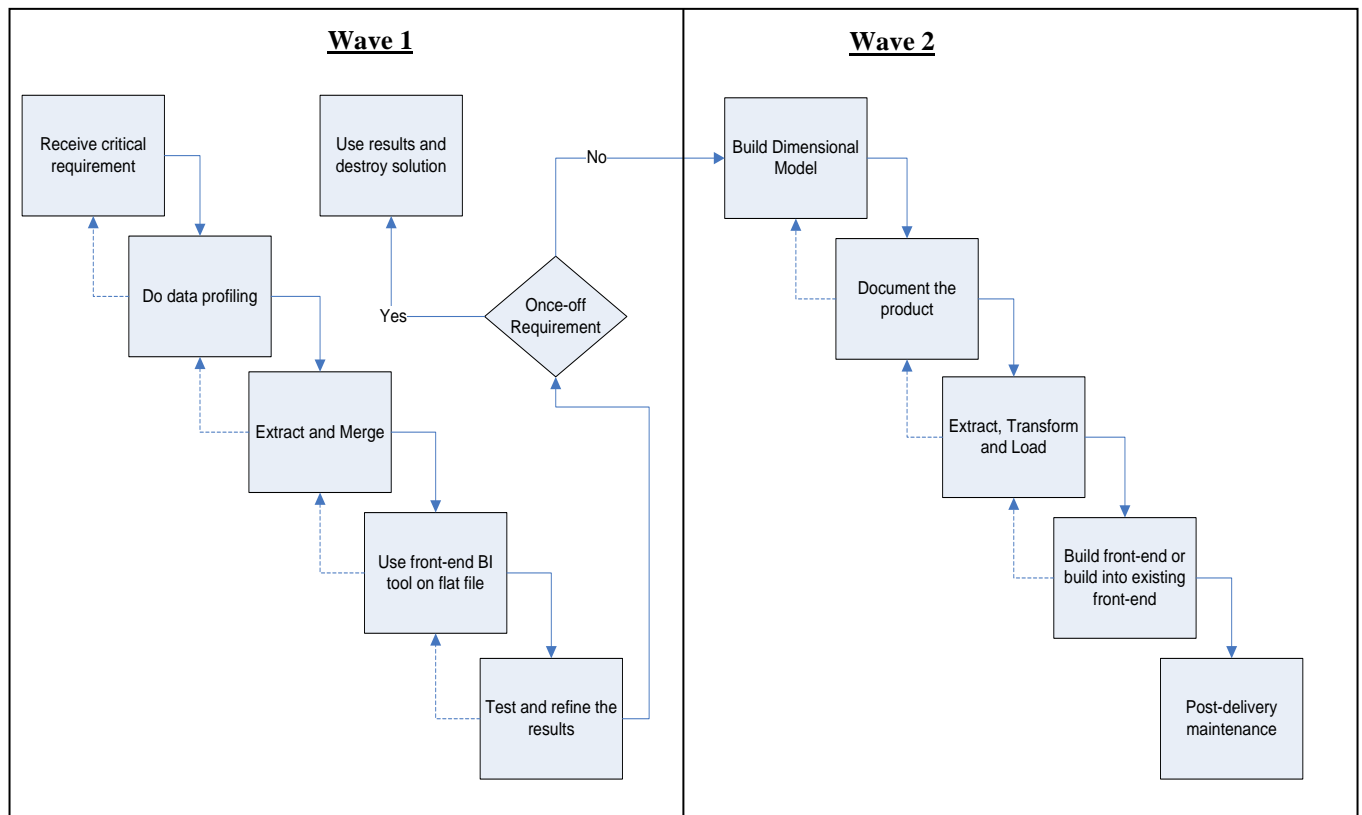


Figure 3.4.1 The double wave data warehouse lifecycle model.

3.4.2 Wave 1

The first wave of the DWDW lifecycle model concentrates on ensuring immediate answers to the business question. There can be little or no time to wait for a project to finish. In wave 1 there is no focus on the query performance. The main objective here is achieving a result in a day or two. Query optimisation can take months or years and is therefore left for wave 2.

The database in wave 1 is in a flat file structure. A flat file structure consists of several big tables. The advantage of a flat structure, according to Hoffman (2003:172-173), is that not much time is needed to build this kind of database. The tables are in a spreadsheet format. Harts (2007:100) states that BI end-users find it much easier to understand and use data stored in spreadsheet format.

The biggest challenge of wave 1 is to find the correct source of information. The users (data capturers) of source systems are normally only familiar with the screen where the data are entered and not the tables where they are stored. Data can sometimes be inserted into one table by the front-end screen, and that insert activates triggers or store procedures that insert data in other tables or databases. Finding the correct source can be challenging at times and can result in long hours of data profiling. Data profiling will be discussed in more detail in chapter 4.

Once the right sources are found, the extract, merge and load (EML) process is built. Historically the extract, transform and load (ETL) process was the only

process known by data warehouse and BI professionals. The proposed DWDW lifecycle model uses the ETL process only in the second wave. The EML process extracts information from the different sources, merges it into one file, and loads it into a staging area where it can be accessed by a BI front-end tool. The extract and merge parts of the EML process can be used in wave 2 to build the ETL process.

Wave 1 of the DWDW lifecycle model concentrates on producing information quickly. However, it also creates an easy platform where new business questions lead to a process of building a complete data warehouse, when a large dedicated budget is not available. By answering the business question, the BI team is building credibility, because it is constantly adding value.

3.4.3 Wave 2

During wave 2 in the DWDW lifecycle model, the data warehouse is optimised for query and analysis. This wave concentrates on properly modelling the business requirement into the final data warehouse structure. The structure built in wave 2 is a de-normalised database structure with fact and dimension tables. The biggest challenge of wave 2 is to avoid duplication of dimensions. It might happen that the new source that needs to be added hosts a more complete set of attributes of a specific dimension, for example customer dimensions. Customer information is normally kept in the financial system for payments and in the order system for placement of orders. If there are some sales that do not go through the order process, there may be customers in the

financial system who are not listed in the order system. It sometimes makes sense to load the information from both systems to avoid possible problems.

After the data model is built in wave 2, it is ready for population. Because the right sources were already selected in wave 1, the biggest focus of the ETL process in wave 2 is the 'transform' part of the process. Look-up queries should be built to write the foreign keys into the fact table.

Once the data warehouse or new part of the data warehouse is populated, the front-end can be adapted to include the new information. As soon as the end-users are happy, the documentation can be finalised to ensure that on-going maintenance can be done.

3.5 Conclusion

This chapter has used a case study research strategy to develop the DWDW lifecycle model. Oates (2006:146) states that case studies can be used to build new theories, concepts, framework or models. These theories can then be applied by researchers to new situations, making use of other research strategies. This case study was used to develop the concept of the DWDW lifecycle model. This model will also be tested by means of structured interviews in Chapter 8.

The DWDW lifecycle model has improved the time it took to deliver on information requirements from business. The budget for BI hardware was still

a problem, and chapter 4 will explain the study that was done to address this, by optimally utilising all the IT hardware available in the company.

CHAPTER 4

Refinement of the double wave data warehouse lifecycle model

4.1 Introduction

In chapter 2, the design and build technique, processes and architecture that exist in the field of BI, were reviewed. Chapter 3 introduced the DWDW Lifecycle Model, and explained how it was developed using a case study research strategy.

This chapter will explain the action research strategy followed, in order to refine and improve the DWDW Lifecycle Model, while also attending to critical business requirements. It will first present a background on action research methodology, and the reasons for using this research methodology. It will then conclude with a detailed discussion of the action research process used and of the improvements made to the DWDW lifecycle model. Sections 4.4 to 4.8 will discuss the first action research cycle followed for this study (Oates, 2008:157).

4.2 Action research methodology

Baskerville (1999:1) described action research as an established research method used in the social and medical sciences, since the mid twentieth

century. Towards the end of the 1990s its use began growing in popularity in scholarly investigations of information systems. Chiasson *et al.* (2009:34) state that action research is a research strategy used to develop a solution, while solving a practical problem and simultaneously creating new knowledge.

Blum (1955:43) discussed two stages in action research - a diagnostic stage, followed by a therapeutic stage. During the diagnostic stage, the researcher studies the social situation. On the other hand, the therapeutic stage involves change of the models up to a stage where the research objectives or hypothesis are satisfied. For the DWDW lifecycle model, the action research strategy was followed, changes to the model were introduced, and the results were studied to introduce more changes in order to improve the model until it fulfilled the research question.

Oates (2008:158) conceptualised action research into the terms of what he called F, M and A:

- ‘F’ stands for framework of ideas
- ‘M’ stands for problem-solving methodology
- ‘A’ stands for area of application (the company where the action research is done)

In this study, the framework ‘F’, in the form of the DWDW lifecycle model, was established in the previous chapter. The DWDW lifecycle model was used in the credit management section of the company as the framework of ideas for the study. Problems were solved using practical solutions for the practical problems, within the telecommunications company used for this study (M).

Sections 4.3 to 4.6 describe the problems experienced with the model, as well as solutions to those problems. This study was done in a telecommunications company in South Africa (Area 'A').

Oates (2008:157) quotes Susman & Evered (1978) and states that action research often expresses the following five stages:

- Diagnostic
- Planning
- Intervention
- Evaluation
- Reflection

The research cycle has been repeated with BI projects in the above-mentioned company. Under each one of these stages, the detail will be explained regarding the steps followed to achieve the research results. The research cycle starts with the background of the research project, whereafter each step of the project is reviewed.

4.3 Background

In 1996 the South African government compiled a white paper on telecommunications policy, which was released in March of that year (Anon, 1996:1). The major proposal in the white paper was that the owner and operator of the fixed telephony infrastructure would be granted a limited period

- the so-called exclusivity period - of monopoly in the provision of basic telecommunications services. This exclusivity period was to last for a period of five years, but could be extended to six years if the telecommunications operator met network rollout and service targets. The rollout targets included doubling the number of subscriber access lines to 2.7 million, installing 120,000 new public telephones, connecting 3,200 villages to the telecommunications network for the first time and providing service to more than 20,000 priority customers such as schools and clinics. The exclusivity period was intended to facilitate network expansion as rapidly as possible, in order to facilitate the move towards universal access. The agreement left the telecommunications provider with the challenge to plan and manage the implementation targets set by government, while at the same time preparing for competition once the exclusivity period expired.

4.4 Diagnostic stage (action research)

During the action research project of this study, the research team consisted of the researcher and five BI professionals. Systems thinking was used as the diagnostic tool. Coghlan and Brannick (2009:95) state that systems thinking and action research complement each other when investigating a real life situation. The systems thinking process suggests that the system/process is represented in a diagrammatic form (Coghlan & Brannick, 2009:95).

For this study, the DWDW lifecycle model was represented in diagrammatic form. The business requirement for the credit management process was first

documented, whereafter a diagram was developed which covered the entire credit management process. Thus, this study has two diagrams. The first is the DWDW lifecycle model that represents the BI SDLC. The second diagram is the collection process within the credit management section. The diagnostic stage reported on the business requirement that exists in the credit management and on the detailed steps within the two waves of the DWDW lifecycle model, which were used to fulfil the business requirement.

4.4.1 Business requirement

The company used for this study needed a BI solution to identify customers with a poor credit record in areas where there were customers with a good credit record and who were waiting for a new service. Although the replacement of a customer with a poor credit record with a new customer would not increase the number of lines, it would help to maintain or improve the cash flow of the company. Replacing customers who could not afford the service and who stayed in the under-serviced area with customers who could afford the service and lived in the same area, was not the only solution, however, as one of the targets was to install new services in under-serviced areas. Furthermore, the unemployment rate in the under-serviced areas of the country was high, and this made it very difficult for people in these areas to afford a telephone service (South African Institute of Race Relations, 2000:359). To accommodate these people, the company had to relax its credit policy.

4.4.2 Credit management process

Graham and Coyle (2000:57) state that a credit policy is a formal document approved by the directors, or an informal document controlled by the financial director. The credit policy includes credit limits allowed by the company, as well as credit vetting and monitoring. ITC and Experian are the two organisations providing a credit vetting service in South Africa.

Relaxing the vetting policy would have a negative influence on debtors, which would lead to an increase in bad debts. Bad debts would periodically be written off and would then be given to a lawyer to collect. Hope (2006:165) states that bad debts are a cash-flow indicator. Reduction in bad debts represents good cash flow management (Salehi & Abedini, 2009:402). In this study, bad debts were increasing, which meant that cash flow was under pressure. Bad debts were collected, and a credit note was captured on the system to credit the bank account of the company as “bad debts recovered”.

Gray (2004:131-132) states that bad debts are risky to any company, and are therefore one of the indicators used by investors to evaluate their investment risk. Bad debts would inevitably decrease the share price of the company.

When looking at the circumstances of this particular telecommunications company, it became clear that it was important to ensure proper control within the debtors’ management process. The average collection time in days (debtors’ days) was the financial management ratio used by the company to

manage debtors and prevent bad debts. Armstrong (2006:605) presents the formula for debtors' days as follows:

$$\text{Debtors' days} = \text{Debtors} / \text{sales} \times 365$$

In order to understand debtors' management, the payment process will be explained. Before a customer's service was installed, a R250 deposit plus installation fee was required for a customer using domestic calls only, and R3,000 plus installation fee for those customers also needing an international service. The customer's service was suspended for outgoing calls if the outstanding balance reached the value of the deposit paid for more than 30 days. If it exceeded 60 days, the service was suspended for incoming calls too. When the outstanding balance was higher than the deposit for 90 days, the service was terminated.

A BI solution was needed to do the collections, following the collection process described below. The agents doing these collections were receiving incentives for successful collection efforts. The process for collections worked as follows (see figure 4.4.2):

- Collection agent runs a debtors' management report
- The agent temporarily lifts the suspension of the telephone of the debtor
- Agent makes a call to the debtor and negotiates payment dates

- Payment dates are inserted on the billing system, and the telephone number is suspended again
- Payment dates are extracted into the data warehouse for collection reporting
- Agent uses the collection reporting to follow-up on promised payment dates (reminder call is made a day before agreed payment date).

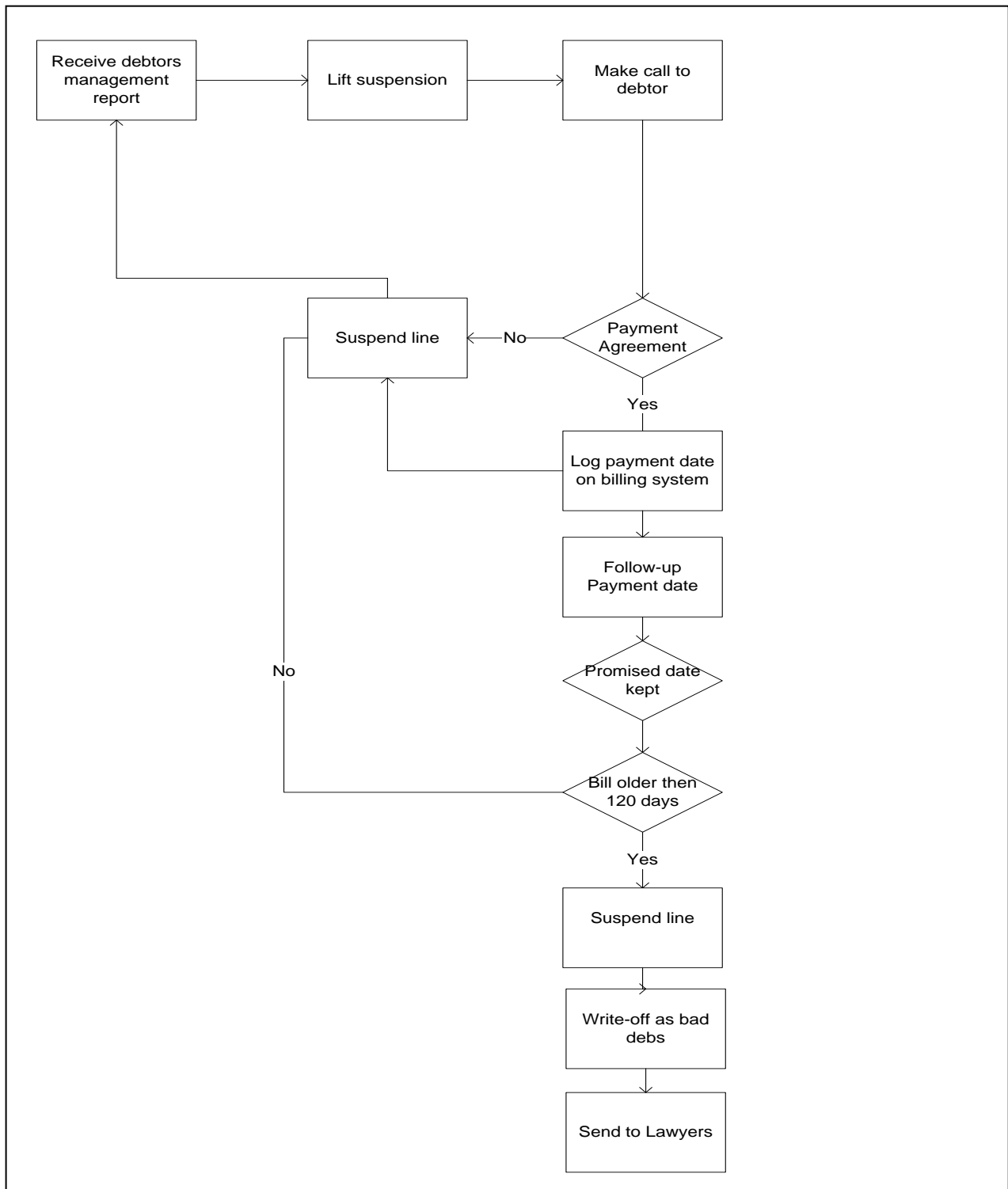


Figure 4.4.2 Telecommunications collection process.

Incentives were only paid to agents when debtors kept their promises. This was to avoid the payment of incentives when payment was not received. Because of the small percentage of debtors that paid on the exact date previously agreed upon, an arrangement was made that when payment was made not more than five days late, the agent still received the agreed incentive.

These circumstances considered, the BI solution needed to identify customers with a poor credit record on the infrastructure, and who were preventing customers with good credit records from joining the infrastructure. The company also needed a solution to manage poor-paying customers in areas where there were no customers waiting for a service. This BI solution needed to be implemented following the DWDW lifecycle model. Section 4.5 (below) will explain the planning that was done for this implementation.

4.5 Planning stage (action research)

During this stage the research team used the DWDW lifecycle model (figure 3.4.1) as the framework to improve the BI implementation processes, while the credit management process in figure 4.4.2 was used to reduce the bad debts in the company. This planning stage concentrated on two areas:

- Credit management business requirement
- Improving the DWDW lifecycle model

4.5.1 Credit management business requirement

Business needed a solution to calculate the incentives to be paid to the collection agents. The solution also needed to manage poor-paying customers in areas where there were no customers waiting for a service, and the solution should be able to track debtors based on different outstanding payment segments.

4.5.2 Improving the DWDW lifecycle model

The DWDW lifecycle model developed from the case study in chapter 3 was developed only from one case study. The research team therefore planned to use action research to optimise the lifecycle model using a real life situation. The BI solution, developed for this action research study, tested and improved every single step of the DWDW lifecycle model introduced in chapter 3. The detailed steps are covered under wave 1 and wave 2 of the DWDW lifecycle model and will be explained in the intervention stage of the action research cycle, discussed below.

4.6 Intervention stage (action research)

The intervention stage was used to implement a BI solution based on the business requirement in the credit management environment, while improving the DWDW lifecycle model.

4.6.1 Wave 1

4.6.1.1 Receive critical requirement

The requirement was to have a BI solution where debtors could be categorised according to the age of their oldest bill. There should be categories for 30 days, 60 days, 90 days and more than 90 days. Debtors would be allocated to these categories, based on their oldest bill. It was also important to know when a debtor was in the more than 90 day category, and what the amount of debt was for the 90, 60 and 30 day categories. The credit management department concentrated primarily on the 'more than 90 days debtors'. However, it sometimes happened that the debtor paid older debts, but that current debts were growing rapidly. In some instances, the debtor's telephone service was not suspended as the credit management policy required. A debtor could also have more than one telephone service and could sometimes pay for one service and not for the other one. It was, therefore, important to report debtors on an account number and not on a telephone number.

While collecting the business requirements, it was realised that it was easier for business people to explain their requirements in a report or spreadsheet. It was therefore decided that the business would give their requirement in spreadsheet format. That helped the BI team to design the front-end whilst collecting the requirements. Excel was used to collect requirements, because it was the reporting/BI tool known to the business. Any other BI tool, like 'Business Objects' or 'Qlikview' can be used for collecting the requirements, but it

improves the process when the BI tool is known to the business and it does not require a capital investment.

4.6.1.2 Data profiling

Olson (2003:122) states that the data profiling process ensures that the correct source system is used for a BI solution. In this study, the billing system was up-to-date for extracting billing information, but a better source system was needed for information on the suspension of telephone numbers. Data profiling revealed that the exchange database was the best source for suspension information and that the telephone number was the only common field by which the two sets of information could be combined. The common field is used to merge data coming from different source databases or tables into one single flat file.

4.6.1.3 Extract, merge and load (EML)

EML in wave 1 of the DWDW lifecycle model provided a flat file by extracting source data and merges into one file for reporting purposes. The extract and merge process puts the data in a flat file for the first wave of the lifecycle model. The data in this study were not only coming from different tables in the database of the billing system, but also from the exchange database. The suspension type, user code and name of user that had changed the status, came from the exchange database. All data sets were merged to create a flat file to report from - see figure 4.6.1.8.1.1.

While implementing the EML process, the BI team was facing the following challenges, because of the limited budget available for BI:

- Limited processing power
- Limited server random access memory (RAM)
- Limited network speed
- The wrong disk speed and storage combination

4.6.1.4 Limited processing power

When the BI team started the BI implementation, there was not a budget for BI servers, and the team made use of the old hardware (servers) that were replaced in the OLTP environment. It was important for the company to use the best servers for the OLTP systems used by the front offices and call centres. The company did not want to battle with slow response of a system, if a customer was waiting at the counter or holding the line when doing business with a call centre. Therefore, money available for hardware was spent in the OLTP environment.

The BI team, therefore, decided that the hardware resources available in the company needed to be used optimally to improve the BI solution. However, a data warehouse is normally updated at night to avoid slowing down the OLTP system. Furthermore, SQL queries are very resource-intensive and were not allowed on the OLTP system database during working hours. Permission was thus granted to the BI team to use the OLTP hardware after hours, because

people working on the OLTP systems were only using the systems during normal working hours. Sections 4.6.1.5 to 4.6.1.8 will explain how these resources were utilised.

4.6.1.5 Limited server random access memory

When an ETL process is extracting data from more than one table, the data are first loaded into RAM for the transformation of the data before they are either loaded into a staging table, or the data warehouse table. Even though RAM is not very expensive, it was not always possible to add RAM because the operating system that was used (Windows Server standard edition) had limited the server to only using 4 gigabytes of RAM. Upgrading to the Enterprise edition was very expensive, and therefore the BI team had to optimally use what was available in the company.

The BI team was challenged with the fact that if not enough RAM is available, the server started using virtual RAM. With virtual RAM, the server starts writing the above-mentioned data to the temporary files on the hard drives of the server. Some IT professionals refer to the use of virtual RAM as paging. When the system goes into paging, it slows down the query completely, because the processor needs to read and write continuously to hard drive. Therefore, BI professionals need to minimise paging.

4.6.1.6 Limited network speed

The network speed discussed in this study refers to Ethernet speed. When building the ETL processes, it is important that the BI team understand the network speed between the data warehouse and the source systems. To avoid the movement of unnecessary detailed data, the facts were summed; the detailed records were not moved, only the aggregated data. Section 4.6.1.9 (below) will explain this approach, and how it was used with the DWDW lifecycle model.

4.6.1.7 The wrong disk speed and storage combination

The BI team requested the OLTP system developers to use smaller disks on the server or Storage Access Network (SAN), because disks that are rotating faster perform better than big disks. Bijaoui and Hasslauer (2007:61-65) state that every time big files are moved around, the disk access pattern comes close to sequential access. Bijaoui and Hasslauer (2007:61-65) state that sequential access occurs when the server is reading or writing disk blocks in retain order.

On the other hand, there are just a certain number of slots for disks on the server. That means that if 300 gigabyte (GB) disks are used and there are only 6 slots available, the storage of the specific server amounts to 1800 GB, while if slower disks are used - for example 1 terabyte (TB) with 6 slots - it amounts to 6 TB.

4.6.1.8 Loading the flat file during wave 1 EML process

For this study the BI team came across two kinds of record:

- Transactions
- Master data

4.6.1.8.1 Transactions

Transaction records are normally a combination of different master data (MD) elements. For instance, a sales transaction consists of Customer MD, Product MD, Price MD, and so forth. These records are easy to select from the OLTP system, because a transaction cannot be changed. If a transaction is incorrect, it needs to be reversed by another transaction, and both transactions are reported on. Therefore, only new records need to be loaded into the data warehouse, which simplifies the ETL process.

Cust_num	Cust_name	Tel_num	Debt or_class	Current_Date	User_code	Suspension_desc	Credit_rating_score	Current_amount	30_day_s_amount	60_day_s_amount	90_days_amount	90_plus_amount	Province	City	Street_name	Street number
123865	J van Rooyen	4015643	90 days	13/02/2006	vanrj	Both Ways	232	0	0	146	2879	0	Natal	Durban	Main	25
123654	A Spies	4016543	60 days	13/02/2006	spiea	Outgoing	132	374	1234	1234	0	0	Northern Cape	Upington	Brink	1
123059	N du Plessis	4017123	30 days	13/02/2006	dupln	Outgoing	212	112	123	0	0	0	Western Cape	Cape Town	Hoof	22
123913	D Kock	4017000	30 days	13/02/2006	kockd	Outgoing	123	113	221	0	0	0	Natal	Durban	Middel	4
123190	S Burger	4017002	60 days	13/02/2006	burgs	Outgoing	278	129	3785	738	0	0	Natal	Durban	van der Walt	123
123345	B Botha	4017004	90 plus	13/02/2006	bothb	Both Ways	238	0	0	0	2885	5674	Natal	Durban	Main	1
123356	P Piek	4017006	90 plus	13/02/2006	piekp	Both Ways	239	0	0	0	2886	5675	Northern Cape	Upington	Brink	3
123366	G De Wee	4017008	30 days	13/02/2006	deweg	Outgoing	264	124	275	0	0	0	Western Cape	Cape Town	Hoof	5
123378	J van Rooyen	4017010	30 days	13/02/2006	vanrj	Outgoing	251	251	284	0	0	0	Natal	Durban	Middel	7
123862	A Spies	4017012	30 days	13/02/2006	spiea	Outgoing	258	118	269	0	0	0	Natal	Durban	van der Walt	9
123567	N du Plessis	4017014	60 days	13/02/2006	dupln	Outgoing	265	125	3781	734	0	0	Natal	Durban	Main	35
123987	D Kock	4017016	90 days	13/02/2006	kockd	Both Ways	235	0	0	123	2882	0	Northern Cape	Upington	Brink	37
123349	S Burger	4017018	90 days	13/02/2006	burgs	Both Ways	252	0	0	149	3249	0	Western Cape	Cape Town	Hoof	39

Figure 4.6.1.8.1.1 Flat file with test data.

The BI team decided to design and build the front-end as represented in figure 4.6.1.8.1.2 below. At this stage the source data were not available. The data warehouse team, therefore, used test data to complete the design and build the front-end. The file hosting the test data, was kept to serve as a template for the flat file with the real data.

Debtor Profile	Suspension_desc	Current_amount	30_days_amount	60_days_amount	90_days_amount	90_plus_amount
Current	Blocked Both Ways	9 786.00	0.00	0.00	0.00	0.00
	International Block	18 724.00	0.00	0.00	0.00	0.00
	Open	0.00	0.00	0.00	0.00	0.00
	Incoming Only	0.00	0.00	0.00	0.00	0.00
30 days	Blocked Both Ways	0.00	0.00	0.00	0.00	0.00
	International Block	0.00	0.00	0.00	0.00	0.00
	Open	0.00	0.00	0.00	0.00	0.00
	Incoming Only	5 899.00	8 273.00	0.00	0.00	0.00
60 days	Blocked Both Ways	0.00	0.00	0.00	0.00	0.00
	International Block	0.00	0.00	0.00	0.00	0.00
	Open	0.00	0.00	0.00	0.00	0.00
	Incoming Only	6 611.00	45 103.00	23 774.00	0.00	0.00
90 day	Blocked Both Ways	0.00	0.00	0.00	0.00	0.00
	International Block	0.00	0.00	1 598.00	0.00	0.00
	Open	0.00	0.00	0.00	0.00	0.00
	Incoming Only	0.00	0.00	0.00	0.00	0.00
90 plus days	Blocked Both Ways	0.00	0.00	0.00	0.00	45 404.00
	International Block	0.00	0.00	0.00	0.00	0.00
	Open	0.00	0.00	0.00	0.00	0.00
	Incoming Only	0.00	0.00	0.00	0.00	0.00

Figure 4.6.1.8.1.2 Data cube for debtors' management solution.

Figure 4.6.1.8.1.2 (above) is an example of the front-end solution. The rows represent the different debtors' categories and the suspension descriptions, while the columns indicate the aging of the outstanding amounts. There were other dimensions like customer number, customer name, and telephone number available in the report, which are not currently shown in figure 4.6.1.8.1.2 - but these are available to the customer using the live report.

4.6.1.8.2 Master data

Master data were used in the data warehouse to report on infrastructure available. That means that a snapshot is taken of the infrastructure every day. The different components of the infrastructure could sometimes change, and only those changes are loaded into the data warehouse. Master data could change and result in changes in the data warehouse.

4.6.1.9 EML process

When doing the EML process, the first step of the process is to extract the information from the source system. As discussed in chapter 2, the EML process is done with an ETL tool, for example IBM Data Stage or Business Objects Data Integrator. These BI tools use Open Database Connection (ODBC) and other native database connections in order to connect to the source system on the one end for extracting the information from the OLTP systems. They also connect to the data warehouse to load the transformed/merged data into the data warehouse.

When a big flat file needs to be loaded on a data warehouse during wave 1 of the DWDW lifecycle model, the file was first compiled on the OLTP server. A staging database was created on the OLTP server. That means that the merging of data was done on the OLTP side before it was moved to the data warehouse, and therefore, the merging SQL query is using the resources in the OLTP environment. To deliver the first BI solution, the merged data were moved to the data warehouse after hours. This process is known as the data take-on.

After the data take-on was done, the staging database was used for ongoing loads. The data were used to compare against the OLTP system to identify changes in master data and to identify new records. Only the changes and the new records were moved over to the data warehouse.

That means that the OLTP server was used (after hours) to do the most resource-intensive part of the EML process. This sped up the EML process substantially. It helped because the company had 600,000 customers. That means that the customer table in the OLTP system had approximately 600,000 records. If each customer makes two calls per day, 1.2 million records are created within a single day. One record consists of several fields. In figure 4.6.1.8.1.1 (above), there are 17 different fields. It is resource- and time-intensive to compare each field of every record. The BI team, therefore, implemented two tables in the staging database. These were the old staging table that represents the data as they were the previous day, and the new staging table that represents the data as they are on the day the EML process runs.

An extra field was created for each of these two tables. The additional field was called the delta field. The delta field was used to put all the data of the different fields of an entire record into one text string. This field was used to compare the two tables, in order to identify the changes. The changes were moved into the transfer table, which was moved to the data warehouse. In the data warehouse, the record transferred was added to the flat file table.

D e l t a	Cust_num	Cust_name	Tel_num	Current_Date	User_code	Province	C i t y	Street_name	Street number
123865J van Rooyen	401564313/02/2006	vanrj	Natal	Durban	Main	25			
123654A Spies	401654313/02/2006	spiea	Northern Cape	Upington	Brink	1			
123059N du Plessis	401712313/02/2006	dupln	Western Cape	Cape Town	Hoof	22			
123913D Kock	401700013/02/2006	kockd	Natal	Durban	Middel	4			
123190S Burger	401700213/02/2006	burgs	Natal	Durban	van der Walt	123			
123345B Botha	401700413/02/2006	bothb	Natal	Durban	Main	1			
123356P Piek	401700613/02/2006	piekp	Northern Cape	Upington	Brink	3			
123366G De Wee	401700813/02/2006	deweg	Western Cape	Cape Town	Hoof	5			
123378J van Rooyen	401701013/02/2006	vanrj	Natal	Durban	Middel	7			
123862A Spies	401701213/02/2006	spiea	Natal	Durban	van der Walt	9			
123567N du Plessis	401701413/02/2006	dupln	Natal	Durban	Main	35			
123987D Kock	401701613/02/2006	kockd	Northern Cape	Upington	Brink	37			
123349S Burger	401701813/02/2006	burgs	Western Cape	Cape Town	Hoof	39			
123123B Botha	401702013/02/2006	bothb	Natal	Durban	Middel	17			

Figure 4.6.1.9 Using the delta field for table comparison.

When master data were changed in the OLTP system, they were identified by comparing the old delta field with the new delta field of the tables in the staging database (figure 4.6.1.9 above). These changes are moved to the data warehouse. In the data warehouse, the changed record is just added to the flat file table. Because the date is normally part of the entire record, it can be stored as a new record, and does not need to overwrite the old record.

4.6.1.10 Use BI front-end tool on flat file

BI front-end tools like Actuate, Business Objects and Essbase, can use a flat file as a data source (Scheps, 2008:58). The main aim of the first wave of the DWDW lifecycle model was to fulfil the immediate information need and not to optimise the solution. The front-end was already designed when the business requirements were received from the BI end-user. The main aim of this phase was to link the front-end with the source data - which at this stage was in a flat file format.

4.6.1.11 Test and refine the results

Testing the interim solution required a lot of data profiling. At this stage of the DWDW lifecycle model, the sources of the necessary data for the flat file could still change. Version control of the ETL jobs was considered to be very important. There should be at least two versions of the ETL at any given time during the development of the solution (old and new versions). Most ETL tools make provision for development comments.

4.6.1.12 Is the business requirement once-off?

Not all the information used in wave 1 is an ongoing requirement. If it is a once-off requirement, the flat file can be destroyed. However, ETL jobs must be stored and could be re-used for new developments, or the flat file could be recreated, if required in future.

4.6.2 Wave 2

4.6.2.1 Build a dimensional model

Wave 2 concentrates on creating a star schema design. Spenik and Sledge (2003:847) refer to a star schema as a database performance optimisation strategy. Figure 4.6.2.1 (below) represents the debtor management star schema design (data mart).

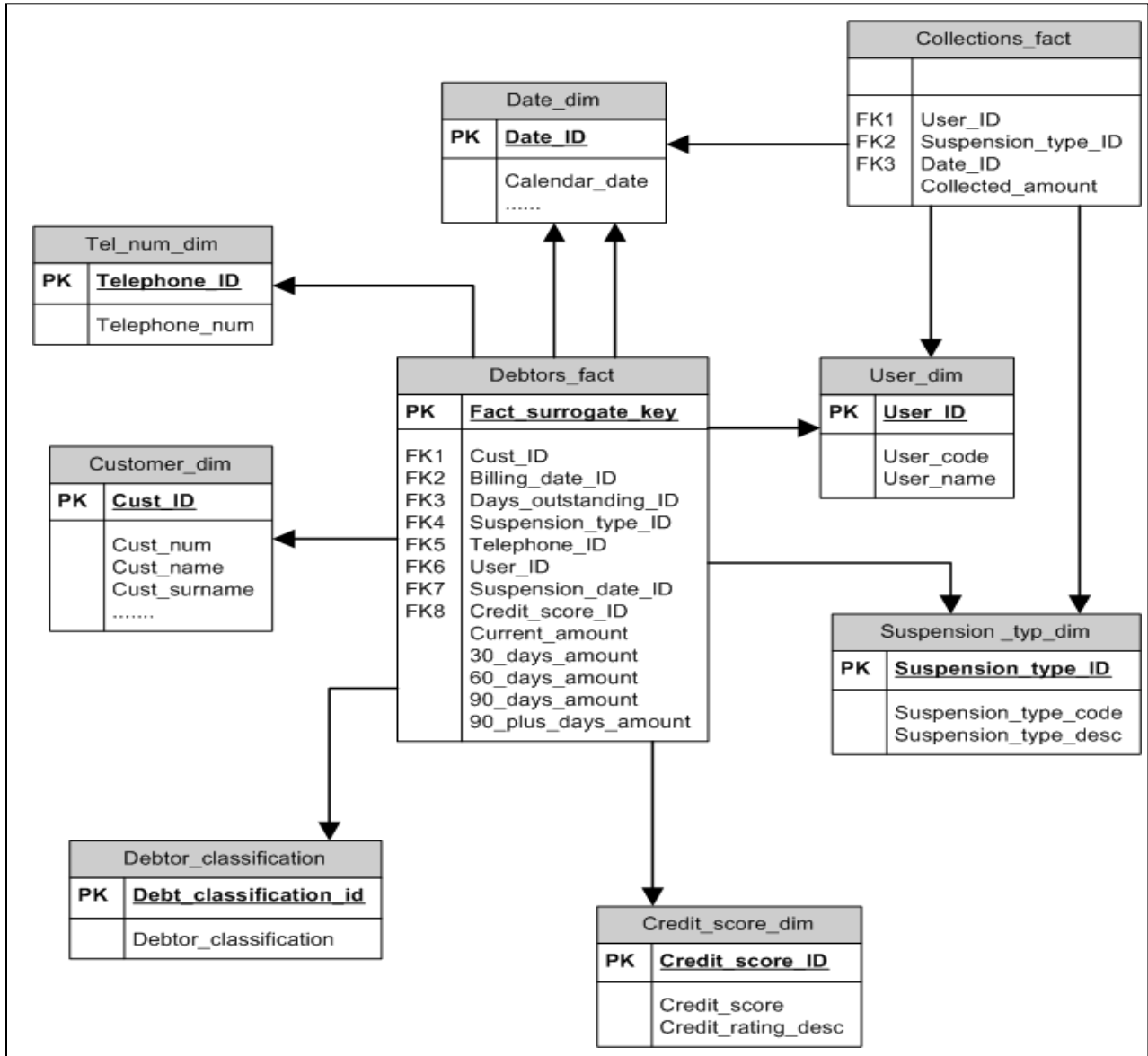


Figure 4.6.2.1 Data model for the debtors' management solution.

The star schema design in data warehousing consists primarily of two kinds of tables - fact and dimension tables. Kimball and Ross (2002:12) defined a fact table as the primary table in a star schema, where the numerical performance measurements of the business are stored. This is illustrated in figure 4.6.2.1 (above) in the Debtors_fact and Collections_fact. The Debtors_fact stores the

outstanding amounts for debtors split up in the four different outstanding time frames. The Collections_fact stores the rand financial value collected from debtors. Both of these fact tables have foreign keys that link to the different dimension tables.

4.6.2.2 Document the product

Schach (2005:489-490) states that the technical documentation of any software product needs to be in place for post-delivery maintenance. Documents for the design of the system are kept to ensure maintenance can be done by anybody else, even if these people were not initially involved in the software projects. Design documents include the graphical design of the data warehouse database/data mart tables, and the relationship between them, as well as the Structured Query Language (SQL) for the ETL. These design documents include the design decisions. When designing a data warehouse/data mart, the architect reaches a point where a decision needs to be made between different design options. Pfleeger and Atlee (2009:296-297) argue that the reason why decisions are captured in design documents is to explain to a new maintenance architect why the specific design option was taken.

In the study a design decision was made to put debtors' aging facts and collection amounts into two different fact tables. It is technically correct to either put them into different fact tables, or into one table. The architect decided to use two separate fact tables, because the reports needed by the BI end-user would need these sets of information separately. These two queries would be much faster with this design because the tables will be much smaller.

4.6.2.3 Transform and load process

The second step of ETL used in wave 2 of the DWDW lifecycle model, concentrates more on the transform and load of the traditional ETL process mentioned previously. The first wave resolves extraction and merging, which concentrates on changing data types from different source systems into the same data type. It also links data sets from different sources for loading into the flat file. The wave one flat file can now be used as a staging area for transformation in wave two. Transformation then transforms data from a normal form into a denormalised form. When loading a dimension table in a dimensional database, it is easy to do a distinct select SQL statement and load the data directly into the dimension table, while the ETL tool creates a new surrogate key for every new row loaded into the table.

With loading of the fact table, surrogate keys from the dimension tables need to be loaded into the fact table as foreign keys. This means that several lookups need to be done for the transformation from a flat file to a dimensional database.

When loading a data warehouse star schema database, one needs two kinds of ETL processes. One ETL process is for loading the dimension tables, and one is for loading the fact tables. These are now discussed.

4.6.2.4 Dimension tables

This study only used the type 2 SCD. Therefore, the type 2 SCD will be discussed in detail. When loading data into the star scheme, the BI team started with the dimension tables. For the data take-on, the flat file that was created in wave one for reporting, was used. This flat file was used because all the history was already loaded into it. If the customer surname, for instance, changes because the woman concerned married, this history was kept in the flat file against the date that it changed. This flat file was also used because it was already on the data warehouse server. The staging database in the OLTP environment was used to load the dimension tables. For the type 2 SCD star schema data take-on, the ETL process is making use of a distinct select SQL statement. The fact that changed records were also identified in wave 1 and brought over to the flat file in the data warehouse, makes it easy to load the type 2 SCD.

After the data take-on was done, the flat file table was empty and was only used as a staging table for the on-going loads of the data warehouse.

As the data were loaded into the dimension table, surrogate keys were numerically allocated to the rows loaded into the data warehouse - starting from 1. The flat file table is emptied after every load. The fact tables will now be discussed.

4.6.2.5 Fact tables

After all the dimension tables were loaded, the fact table was loaded using the surrogate keys of the dimension tables.

Customer_dim		p	
Customer_key	Customer_number	Customer_name	ID_number
1	265	Jan Cloete	5208025493008

Date_dim	
Date_key	Date
5	09/08/2010

Product_dim		
Product_key	Product_number	Product_name
3	8735	Telephone

Sales_fact			
Customer_key	Date_Key	Product_key	Number_sold
1	5	3	1

Figure 4.6.2.5 Fact table is populated with the primary keys of the dimension tables.

That means that if a sale of one telephone is captured in a fact table, the name and customer number (the dimensions) are represented in one single number (the surrogate key). Figure 4.6.2.5 (above) represents the different dimension

records for the above-mentioned transaction and demonstrates how the keys of the different dimensions were loaded into the fact table.

4.6.2.6 Build new BI front-end or build solution into existing BI front-end

This phase is required when the new front-end is built using different software from that used for the front-end in the first wave. This phase was used because the business people were not familiar with the new BI front-end tools. The BI team, therefore, used MS Excel as the front-end tool, as this was familiar to the personnel in the study. Excel was then systematically replaced with the new BI tool as the BI maturity of the company improved (Du Plessis & McDonald, 2007c:231).

4.6.2.7 Post-delivery maintenance

It is known that a good BI solution continuously expands (Inmon, 2005:53). Post-delivery maintenance in a BI environment can, therefore, become a big challenge. In the event of adaptive and corrective maintenance, it might require that the tables be reloaded. When a new dimension is added, it is sometimes unavoidable to reload the fact table.

Naming conventions play an important role in the maintenance of a data warehouse. This allows the maintenance team to identify the business area that the fact or dimension table belong to, as well as the conformed dimensions and fact tables shared amongst the different business areas.

4.7 Evaluation stage (action research)

The evaluation phase was initially used to test the improvement every time the model was changed. This was in order to determine whether the change had a positive influence, no influence, or a negative influence on the time to implement the data warehouse/data mart, as well as the time to load the data into the data warehouse or data mart. The last part of the evaluation stage used interviews to evaluate the success of both the BI solution built, and the DWDW lifecycle model used to build it. Chapter 8 will discuss these interviews in more detail.

4.8 Reflection stage (action research)

Oates (2008:157) states that the reflection stage of an action research cycle is meant to reflect on what has been achieved in terms of the practical outcome of the research process and to determine a possible need for another cycle.

The first cycle of the action research study has developed the detailed steps of wave 1 and wave 2 of the DWDW lifecycle model, but it also identified a new challenge with the hardware architecture used for BI in the company. Chapter 5 will discuss the details of the second action research cycle followed to address the challenge experienced with hardware architecture.

4.9 Conclusion

This chapter discussed the process that was followed to improve the DWDW lifecycle model, while attending to a real life business requirement. The next

chapter will discuss the second action research cycle followed to develop a new hardware architecture that was used with the DWDW lifecycle model.

CHAPTER 5

Double wave data warehouse architecture

5.1 Introduction

Chapter 3 and 4 introduced the DWDW lifecycle model for BI and described its use in the telecommunication company. The DWDW lifecycle model deviates from traditional SDLCs in the sense that a new technical architecture for BI was created.

This chapter will explain the SDLC architecture which existed in the company used for this study. It will also explain why it was not suitable to use whilst implementing BI using the DWDW lifecycle model.

The action research cycle followed to develop the new DWDW lifecycle model will be discussed, and the benefits of this new architecture for companies with a limited IT budget will be introduced.

5.2 Diagnostic stage (action research)

The diagnostic stage of the second action research cycle investigated the architecture used in the company and compared it to architectures from the literature review done. The architecture used for this study will be discussed by referring to the components of the BI architecture developed by Ponniah (2001:129).

5.2.1 Different components of a BI architecture

Ponniah (2001:129) classified the different elements of a BI architecture into the following three major areas (see figure 5.2.1 below):

- Data acquisition
- Data storage
- Information delivery

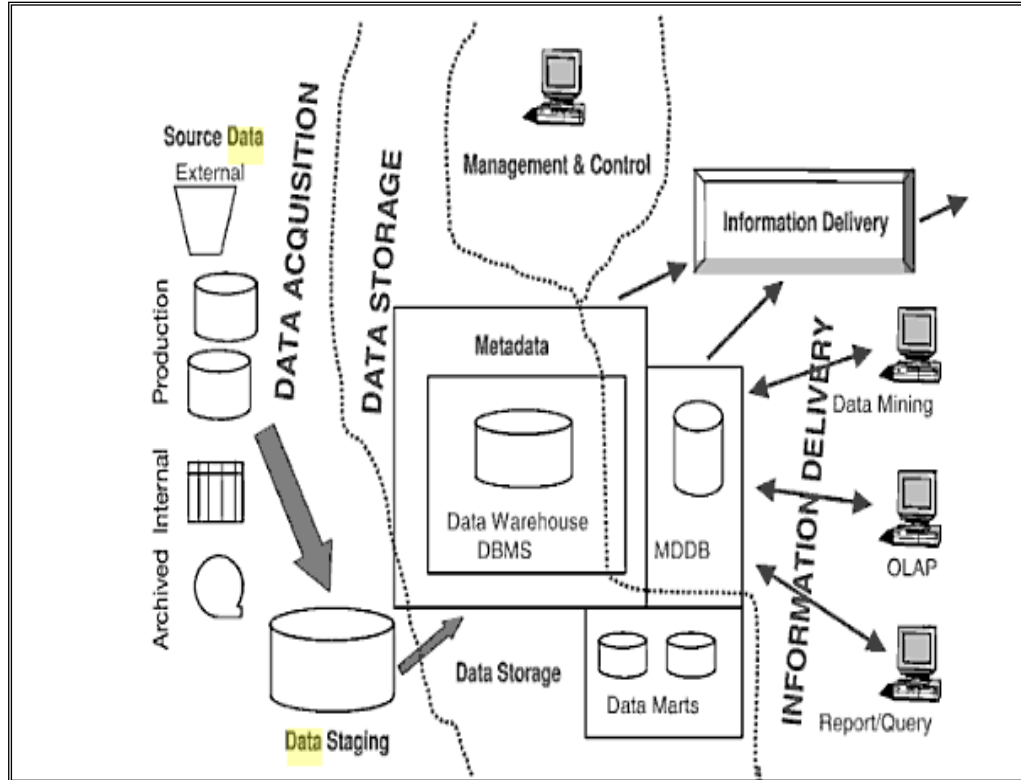


Figure 5.2.1 Three components and services of BI architecture (Ponniah, 2001:129).

5.2.2 Data acquisition

Silvers (2008:139-141) states that data acquisition requires data from different sources of company information - for example from production and financial systems. Some of the sources can be external systems and data files like customer satisfaction survey files. These different data sources - as well as the extract process of the Extract, Transform and Load (ETL) element - represent the data acquisition area (Kelly, 2007:235).

5.2.3 Data storage

The data storage area focuses on loading the selected data into the data warehouse, data mart or data cube. Rainardi (2008:71) states that designing a data warehouse, data mart or data cube can be very time-consuming and, therefore, increases the time spent on delivering a BI solution. It is therefore very difficult to respond to critical business questions when following a star schema design. In contrast, the DWDW SDLC requires a flat file database. This kind of database does not require any design because the data are only merged and stored. This dramatically speeds up the delivery of the solution to the BI end-user. Optimisation is done at a later stage of the project.

5.2.4 Information delivery

Sears and Jacko (2009:38) refer to the process of developing the front-end of a software system as “defining the problem”. Information delivery concentrates on defining the business problem and is about the presentation of information and the interaction of the BI end-users with the BI system. The science of this interaction is called Human Computer Interaction (HCI) (Kerren *et al.*, 2007:1). Hassan (2008:230) states that “HCI is regarded as the intersection of computer science, behavioural sciences, design and several other fields of study”. The front-end of a BI system concentrates on designing a graphical interface that simplifies accessing the company’s data via a computer screen.

5.3 Planning stage (action research)

5.3.1 Limitations of existing BI architecture when using the DWDW lifecycle model

The architecture detailed above was the one used in the company where this study was done. The biggest limitation of this BI architecture, as presented by Ponniah (2001:129), lies in the data storage area. Data are stored in a data warehouse or data mart. Both of these data repositories represent a star schema database, which takes a lot of time to design and build.

Rankins *et al.* (2003:547) believe that any architecture for software development normally needs at least the following two environments:

- Development environment
- Production environment

Loveland *et al.* (2005:16) state that it is important that these two environments look the same and use the same software, to eliminate any software compatibility problems. Neely (2006:175) argues that software compatibility refers to how the software system recognises and integrates with other software. Software incompatibility on the same platform can result in functionality that works, for example, in the development environment, but not in the production environment.

It is very difficult to run only one environment for development and production, because during this process the development environment is often unstable. Developers often change software code, and the environment is frequently restarted because of defects in the code. The architecture only reflects one environment, because the other environment is normally identical to the one represented in the architecture. With the DWDW lifecycle model, it was planned to accommodate a flat file structure and a star schema structure in the data storage area on the development server. The production environment only has a star schema structure. The elements of the new architecture (DWDW architecture) are discussed under the intervention stage of the action research cycle in section 5.3.2 below.

5.3.2 Intervention stage (action research)

Whilst addressing challenges experienced with the architecture of Ponniah (2001:129) when using the DWDW lifecycle model, the new DWDW architecture was created. The name of the new architecture is concordant with the lifecycle model name. The DWDW architecture, like other architectures for software development, required at least two environments: development and production (see figure 5.3.2.1.1 below). The DWDW architecture differed from other SDLCs in that the development environment does not comprise the same elements as the production environment. The development environment made provision for waves 1 and 2 of the DWDW lifecycle model. This means that some of the reports used by the BI end-user were still running in the

development environment. The development environment, contrary to other development environments, catered for star schema and flat file databases.

Business people were not aware of the environment the report is running in, except that response times were slower in the development environment than in the production environment. The main reason for the slow response was the instability of the development environment and because the solution was not yet optimised. However, the focus of wave 1 of the DWDW lifecycle model concentrated on delivering a solution to the BI end-user. Optimisation of the solution only occurred in wave 2.

While the BI solution was used with the front-end connected to the flat file in the development environment, the star schema was built in another database in the same development area. On completion of the database design in the development environment, the database was moved to the production environment. The front-end was then disconnected from the flat file and moved to the new database tables in the production environment.

5.3.2.1 Conformed dimension ensures smooth movement of BI solution from development environment to production environment

When moving the star schema of the newly developed solution to the production environment, the new design was joined to the rest of the data warehouse in the production environment through conformed dimensions. Kimball and Ross (2002:394) state that keys and row headers of conformed

dimensions need to be exactly the same. Song and LeVan-Shultz (1999:379) argue that conformed dimensions are normally used to analyse facts in different data marts. In this study, however, conformed dimensions were used to link the existing BI solution with the new one. Therefore, the dimension used as a conformed dimension had to exist in the production environment and had to be part of the new BI requirement in the development environment. If the customer dimension was, for example, needed as the conformed dimension, it was copied from the production environment to the development environment and regularly updated from there during development. This was done to ensure that the two dimensions were always synchronised in order to smooth the linking process when moved into production.

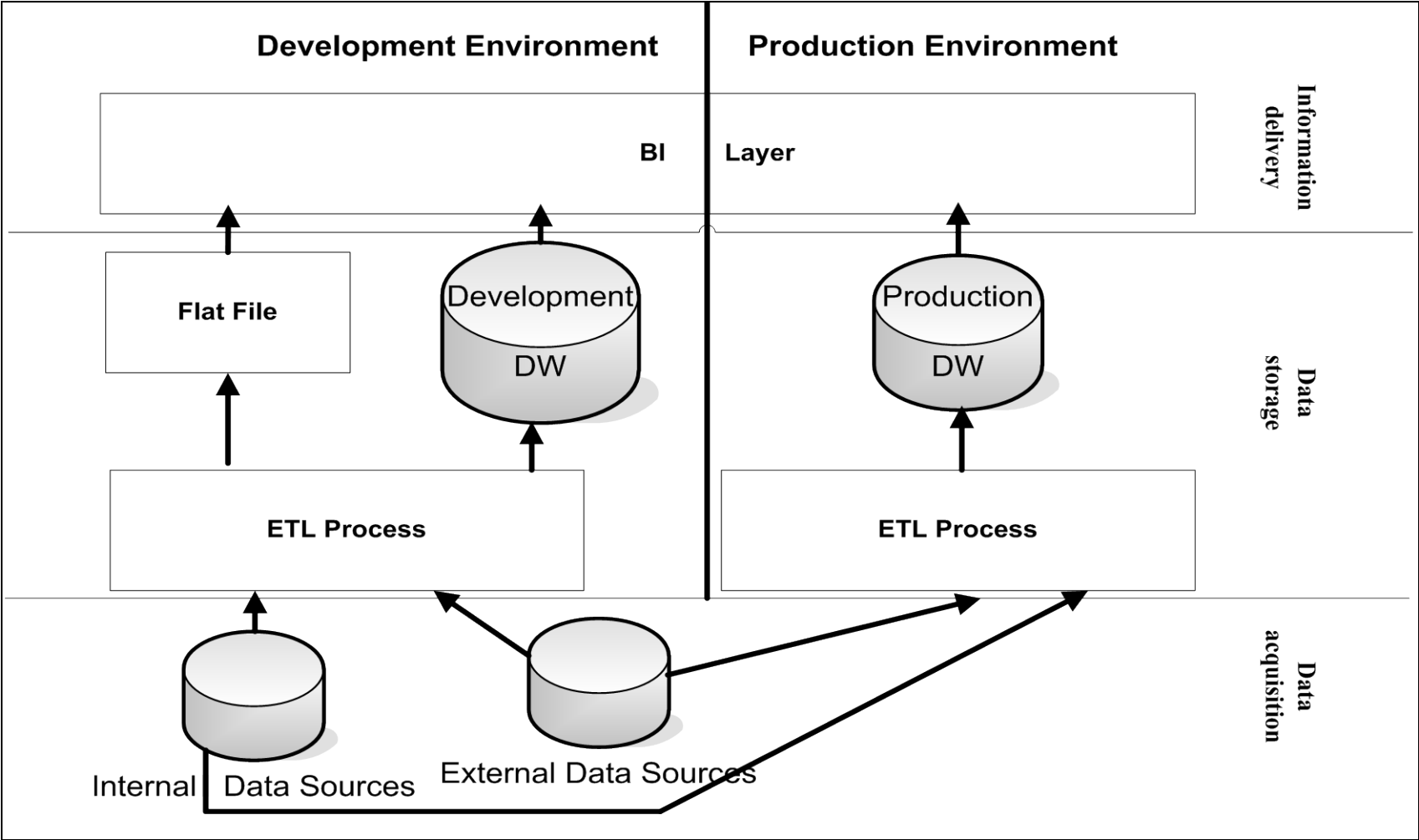


Figure 5.3.2.1.1 DWDW architecture.

When the BI solution was moved from the development environment to the production environment, the database in the development environment was stored in a Unified Modeling Language (UML) format (figure 5.3.2.1.2). Bennett *et al.* (2001:5-8) state that UML is the standard language used to design and document the artefacts of a software system. Database design tools using UML have a reverse engineering function, which is used to create a UML file from an existing database.

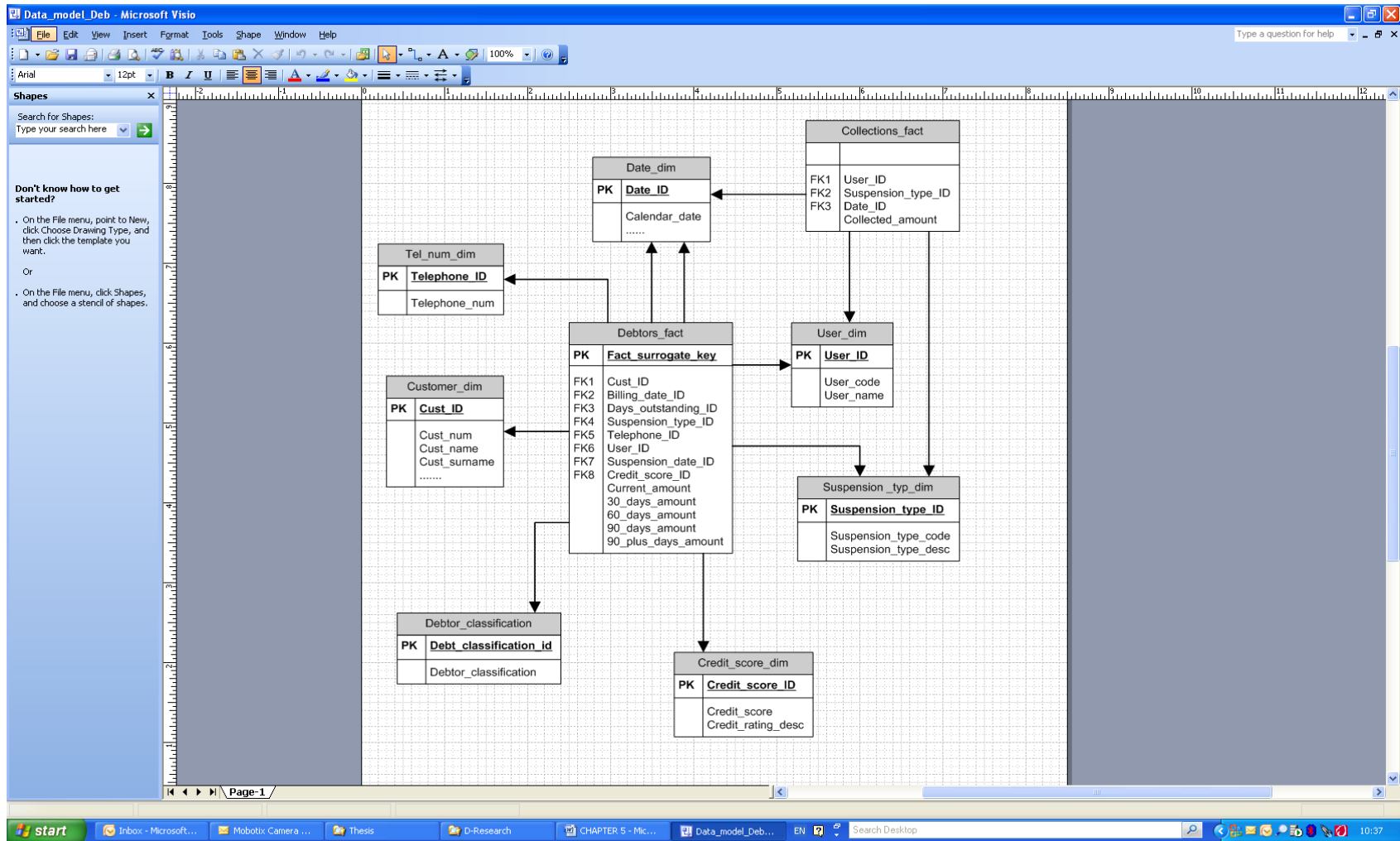


Figure 5.3.2.1.2 UML for credit management BI solution.

5.3.3 DWDW architecture also caters for disaster recovery

The UML file was retained for disaster recovery purposes. Gregory and Rothstein (2007:13) state that the disaster recovery plan (DRP) concentrates on ways to minimise the loss of data to the business, caused by the unavailability of IT systems. In a company with a limited BI budget, it was important that disaster recovery could be implemented at very low cost.

It was not practical to include the BI solution in backup procedures. BI solutions normally grow into terabytes of data, because all history is stored in the same database. With other IT systems, historical data are normally archived in another database running on a different server. When doing backups of terabytes of data, it can become resource intensive on the IT hardware. It, therefore, was more practical, in the event of a disaster, to re-create the database of the BI solution and reload the data from the systems. All the systems are normally included in the DRP. Therefore, the BI solutions can be loaded in the event of a disaster from backups of the source systems or the archive of the loading history data.

By excluding the BI solution from the DRP, space was saved on the DRP hardware and less processing power was needed for daily BI backups. For a company with a limited BI budget, data storing space and processing power were very limited, and therefore, needed to be used wisely.

When the UML tool was not available, the SQL code used to create the database was stored for disaster recovery (see figure 5.3.2 below).

```
/*=====*/
/* DBMS name:   Microsoft SQL Server 2000           */
/* Created on:  2010/01/07 02:11:22 nm             */
/*=====*/

if exists (select 1
  from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
  where  r.fkeyid   =  object_id('COLLECTIONS_FACT')    and  o.name   =
'FK_COLLECTI_REFERENCE_SUSPENSI')
alter table COLLECTIONS_FACT
  drop constraint FK_COLLECTI_REFERENCE_SUSPENSI
go

if exists (select 1
  from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
  where  r.fkeyid   =  object_id('COLLECTIONS_FACT')    and  o.name   =
'FK_COLLECTI_REFERENCE_USER_DIM')
alter table COLLECTIONS_FACT
  drop constraint FK_COLLECTI_REFERENCE_USER_DIM
go

if exists (select 1
  from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
  where  r.fkeyid   =  object_id('COLLECTIONS_FACT')    and  o.name   =
'FK_COLLECTI_REFERENCE_DATE_DIM')
```

```

alter table COLLECTIONS_FACT
    drop constraint FK_COLLECTI_REFERENCE_DATE_DIM
go

if exists (select 1
    from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
    where r.fkeyid = object_id('DEBTORES_FACT') and o.name =
'FK_DEBTORES_REFERENCE_SUSPENSI')
alter table DEBTORES_FACT
    drop constraint FK_DEBTORES_REFERENCE_SUSPENSI
go

if exists (select 1
    from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
    where r.fkeyid = object_id('DEBTORES_FACT') and o.name =
'FK_DEBTORES_REFERENCE_USER_DIM')
alter table DEBTORES_FACT
    drop constraint FK_DEBTORES_REFERENCE_USER_DIM
go

if exists (select 1
    from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
    where r.fkeyid = object_id('DEBTORES_FACT') and o.name =
'FK_DEBTORES_REFERENCE_DATE_DIM')
alter table DEBTORES_FACT
    drop constraint FK_DEBTORES_REFERENCE_DATE_DIM
go

if exists (select 1
    from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')

```

```

where r.fkeyid = object_id('DEBTORES_FACT') and o.name =
'FK_DEBTORES_REFERENCE_TEL_NUM_')
alter table DEBTORES_FACT
drop constraint FK_DEBTORES_REFERENCE_TEL_NUM_
go

if exists (select 1
from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
where r.fkeyid = object_id('DEBTORES_FACT') and o.name =
'FK_DEBTORES_REFERENCE_CUSTOMER')
alter table DEBTORES_FACT
drop constraint FK_DEBTORES_REFERENCE_CUSTOMER
go

if exists (select 1
from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
where r.fkeyid = object_id('DEBTORES_FACT') and o.name =
'FK_DEBTORES_REFERENCE_DEBTORS_')
alter table DEBTORES_FACT
drop constraint FK_DEBTORES_REFERENCE_DEBTORS_
go

if exists (select 1
from dbo.sysreferences r join dbo.sysobjects o on (o.id = r.constid and o.type = 'F')
where r.fkeyid = object_id('DEBTORES_FACT') and o.name =
'FK_DEBTORES_REFERENCE_CREDIT_S')
alter table DEBTORES_FACT
drop constraint FK_DEBTORES_REFERENCE_CREDIT_S
go

```

```
if exists (select 1
           from sysobjects
           where id = object_id('COLLECTIONS_FACT')
           and type = 'U')
drop table COLLECTIONS_FACT
go
```

```
if exists (select 1
           from sysobjects
           where id = object_id('CREDIT_SCORE_DIM')
           and type = 'U')
drop table CREDIT_SCORE_DIM
go
```

```
if exists (select 1
           from sysobjects
           where id = object_id('CUSTOMER_DIM')
           and type = 'U')
drop table CUSTOMER_DIM
go
```

```
if exists (select 1
           from sysobjects
           where id = object_id('DATE_DIM')
           and type = 'U')
drop table DATE_DIM
go
```

```
if exists (select 1
           from sysobjects
```

```

        where id = object_id('DEBTORES_FACT')
        and type = 'U')
drop table DEBTORES_FACT
go

if exists (select 1
        from sysobjects
        where id = object_id('DEBTORS_CLASSIFICATION_DIM')
        and type = 'U')
drop table DEBTORS_CLASSIFICATION_DIM
go

if exists (select 1
        from sysobjects
        where id = object_id('SUSPENSION_TYP_DIM')
        and type = 'U')
drop table SUSPENSION_TYP_DIM
go

if exists (select 1
        from sysobjects
        where id = object_id('TEL_NUM_DIM')
        and type = 'U')
drop table TEL_NUM_DIM
go

if exists (select 1
        from sysobjects
        where id = object_id('USER_DIM')
        and type = 'U')

```

```

drop table USER_DIM

go

/*=====*/
/* Table: COLLECTIONS_FACT */
/*=====*/

create table COLLECTIONS_FACT (
  USER_ID      int      null,
  DATA_ID     int      null,
  SUSPENSION_TYPE_ID int  null,
  COLLECTED_AMOUNT money  null
)
go

/*=====*/
/* Table: CREDIT_SCORE_DIM */
/*=====*/

create table CREDIT_SCORE_DIM (
  CREDIT_SCORE_ID int      not null,
  CREDIT_SCORE     numeric  null,
  CREDIT_SCORE_DESC text    null,
  constraint PK_CREDIT_SCORE_DIM primary key (CREDIT_SCORE_ID)
)
go

/*=====*/
/* Table: CUSTOMER_DIM */
/*=====*/

create table CUSTOMER_DIM (
  CUST_ID      int      not null,

```

```

CUST_NUM      numeric      null,
CUST_SURNAME  text         null,
constraint PK_CUSTOMER_DIM primary key (CUST_ID)
)
go

/*=====*/
/* Table: DATE_DIM */
/*=====*/

create table DATE_DIM (
  DATA_ID      int          not null,
  CALENDAR_DATE datetime     null,
  constraint PK_DATE_DIM primary key (DATA_ID)
)
go

/*=====*/
/* Table: DEBTORES_FACT */
/*=====*/

create table DEBTORES_FACT (
  SUSPENSION_TYPE_ID int          null,
  USER_ID           int          null,
  DATA_ID          int          null,
  TELEPHONE_ID      int          null,
  CUST_ID           int          null,
  DEBT_CLASSIFICATION_ID int      null,
  CREDIT_SCORE_ID  int          null,
  CURRENT_AMOUNT    money        null,
  DAY_OUSTANDING    numeric      null,
  30_DAYS_AMOUNT    money        null,

```

```

60_DAYS_AMOUNT    money    null,
90_DAYS_AMOUNT    money    null,
90_PLUS_DAYS_AMOUNT money    null
)
go

/*=====*/
/* Table: DEBTORS_CLASSIFICATION_DIM          */
/*=====*/

create table DEBTORS_CLASSIFICATION_DIM (
  DEBT_CLASSIFICATION_ID int          not null,
  DEBT_CLASSIFICATION_DESC text        null,
  constraint          PK_DEBTORS_CLASSIFICATION_DIM          primary          key
(DEBT_CLASSIFICATION_ID)
)
go

/*=====*/
/* Table: SUSPENSION_TYP_DIM                  */
/*=====*/

create table SUSPENSION_TYP_DIM (
  SUSPENSION_TYPE_ID int          not null,
  SUSPENSION_TYPE_CODE numeric      null,
  SUSPENSION_TYPE_DESC text        null,
  constraint PK_SUSPENSION_TYP_DIM primary key (SUSPENSION_TYPE_ID)
)
go

/*=====*/
/* Table: TEL_NUM_DIM                         */
/*=====*/

```

```

/*=====*/
create table TEL_NUM_DIM (
    TELEPHONE_ID    int        not null,
    TELEPHONE_NUM    numeric    null,
    constraint PK_TEL_NUM_DIM primary key (TELEPHONE_ID)
)
go

/*=====*/
/* Table: USER_DIM */
/*=====*/

create table USER_DIM (
    USER_ID        int        not null,
    USER_CODE      numeric    null,
    USER_NAME      text       null,
    constraint PK_USER_DIM primary key (USER_ID)
)
go

alter table COLLECTIONS_FACT
add constraint FK_COLLECTI_REFERENCE_SUSPENSI foreign key
(SUSPENSION_TYPE_ID)
references SUSPENSION_TYP_DIM (SUSPENSION_TYPE_ID)
go

alter table COLLECTIONS_FACT
add constraint FK_COLLECTI_REFERENCE_USER_DIM foreign key (USER_ID)
references USER_DIM (USER_ID)
go

```

```

alter table COLLECTIONS_FACT
  add constraint FK_COLLECTI_REFERENCE_DATE_DIM foreign key (DATA_ID)
    references DATE_DIM (DATA_ID)
go

alter table DEBTORES_FACT
  add constraint FK_DEBTORES_REFERENCE_SUSPENSI foreign key
(SUSPENSION_TYPE_ID)
  references SUSPENSION_TYP_DIM (SUSPENSION_TYPE_ID)
go

alter table DEBTORES_FACT
  add constraint FK_DEBTORES_REFERENCE_USER_DIM foreign key (USER_ID)
    references USER_DIM (USER_ID)
go

alter table DEBTORES_FACT
  add constraint FK_DEBTORES_REFERENCE_DATE_DIM foreign key (DATA_ID)
    references DATE_DIM (DATA_ID)
go

alter table DEBTORES_FACT
  add constraint FK_DEBTORES_REFERENCE_TEL_NUM_ foreign key
(TELEPHONE_ID)
  references TEL_NUM_DIM (TELEPHONE_ID)
go

alter table DEBTORES_FACT
  add constraint FK_DEBTORES_REFERENCE_CUSTOMER foreign key (CUST_ID)
    references CUSTOMER_DIM (CUST_ID)

```

```

go

alter table DEBTORES_FACT
  add constraint FK_DEBTORES_REFERENCE_DEBTORS_ foreign key
(DEBT_CLASSIFICATION_ID)
  references DEBTORS_CLASSIFICATION_DIM (DEBT_CLASSIFICATION_ID)
go

alter table DEBTORES_FACT
  add constraint FK_DEBTORES_REFERENCE_CREDIT_S foreign key
(CREDIT_SCORE_ID)
  references CREDIT_SCORE_DIM (CREDIT_SCORE_ID)
go

```

Figure 5.3.3 Code to create a BI solution for debtors' management.

5.4 Evaluation stage (action research)

This new DWDW architecture was evaluated by means of interviews, which are discussed in chapter 8.

5.5 Reflection stage

The reflection stage of this study will be discussed in chapter 9.

5.6 Conclusion

Companies in developing countries are battling to implement satisfactory BI solutions. A SDLC was developed to address the implementation of a BI solution in a company with a limited budget for BI. Because there is also an urgent need to get information to the BI end-user within a day or two, the development of the total solution did not stop as soon as the information was delivered to the BI end-user.

Part of any BI project includes the optimisation of the solution. With the DWDW lifecycle model, in conjunction with the DWDW architecture, users initially use the solution in the development environment, and after final optimisation, they are moved to the production environment.

CHAPTER 6

Maturing a BI solution: The need for the BI literacy and culture maturity model

6.1 Introduction

The earlier part of this thesis (chapters 3, 4 and 5) concentrated on the new DWDW lifecycle model and the associated DWDW architecture, which support the incremental implementation of a BI solution. This new approach facilitates the implementation of BI for companies with a limited BI budget and helps to develop confidence in the users of the BI solution.

This chapter will discuss a case study in a telecommunications company in South Africa and, thereafter, the importance of BI literacy and culture. During this case study, the company urgently needed information on spare telephone infrastructure. Data stored in a data warehouse needed to be accessed in order to use the spare infrastructure for additional services.

The chapter concludes with the introduction of a model that was used to establish confidence in end-users, when they were using the BI solution.

6.2 Developing a BI division

The telecommunications company used for this study already used a data mart (DM) which included information on spare infrastructure. Staff did not know how to use the DM because of a lack of technical knowledge and training on the BI tools available to the business. The requirement to manage the agreement with government (as explained in chapter 1) had, however, forced the company to start using its available data. When staff in the company started asking for information on spare infrastructure, it was realised that the information was available in the DM. However, as mentioned above, it could not be used or accessed because of a lack of technical knowledge. This resulted in sales staff selling telephone services in areas where no infrastructure was available.

The company had arrived at a point where many of the technical challenges and trade-offs were well understood, but where more attention was needed on how BI could be used to deliver business value. The typical BI user is not interested in the technicalities of a BI solution and is more concerned with the value that the solution adds to the business (Williams & Williams, 2007:21). Rasmussen *et al.* (2002:148) state that BI should facilitate required data interrogation and provide the correct answers to the critical business questions to ensure that the company gains and maintains a competitive advantage. In the South African company concerned - which was the sole provider of fixed line telecommunications in the country - it was at that stage important to start using the BI tools that were already available to it.

To be able to use the information and data available in the DM, the company's management had to decide between the following alternatives:

1. Train all the business people on the existing BI tool (Business Objects)
2. Implement a new BI tool which would also need training
3. Create a new division to fill the gap between the technical working of the BI solution and the business itself

The first two options were seen as expensive, risky and very time-consuming, as electronic information was not typically used within the company. Business people did have access to electronic information, but would print it out before using it. When going to a meeting, managers would ask for printed reports from the BI team, and take printed copy with them. Dashboards were printed on paper and placed on the wall. Secretaries printed out new information every day and stuck it on the wall. In other words, training alone would not be sufficient for the company, and, therefore, it made sense to use the third option.

The third option was selected because it was seen as the quickest solution and was also a stepping-stone to get the company to a level where it used electronic information. This is because, as mentioned in chapter 1, there was not a lot of time to get the people to optimally use the BI solution available in the company. A new division was therefore created to fill the gap between the BI solution and the BI end-users.

A decision was taken to use a regional approach to implement the new BI division. The company was divided into six regions to render an appropriate

telecommunications service throughout South Africa. One manager per region was appointed with between three to five knowledge workers, depending on the size of the region. Knowledge workers were staff with a combination of BI and business skills.

6.2.1 The spare infrastructure requirement

With a DM available and BI tools already in place, and together with the BI team, the implementation and management of targets set by government were greatly facilitated. South Africa has a 41% rural population and these rural communities use schools and clinics in the rural areas where there was no infrastructure (South African Institute of Race Relations, 2000:359). Initial implementation in urban schools and clinics would have been easy, but that would not have fulfilled the additional requirement for services to customers in under-serviced (rural) areas. The installation of prepaid phones (coin or card) was also a way of providing a telephone service to the afore-mentioned rural communities.

In order to achieve targets set by government, the first step was, therefore, to determine where the spare infrastructure was, in order to send the sales teams to these areas. This prioritisation was essential, otherwise sales personnel would have sold in areas where there was no infrastructure, resulting in customers waiting for their services while the company had to plan and build that infrastructure. Building the infrastructure could take months or years, depending on the size of the project.

A report on spare and used infrastructure was initiated by the BI division. The telecommunications infrastructure consisted of two network areas:

1. Core network (exchanges and transmission systems)
2. Access network (cables to the different houses, coin phones and businesses)

Figure 6.2.1 (below) is a high-level graphical representation of the network of a fixed line telecommunications company.

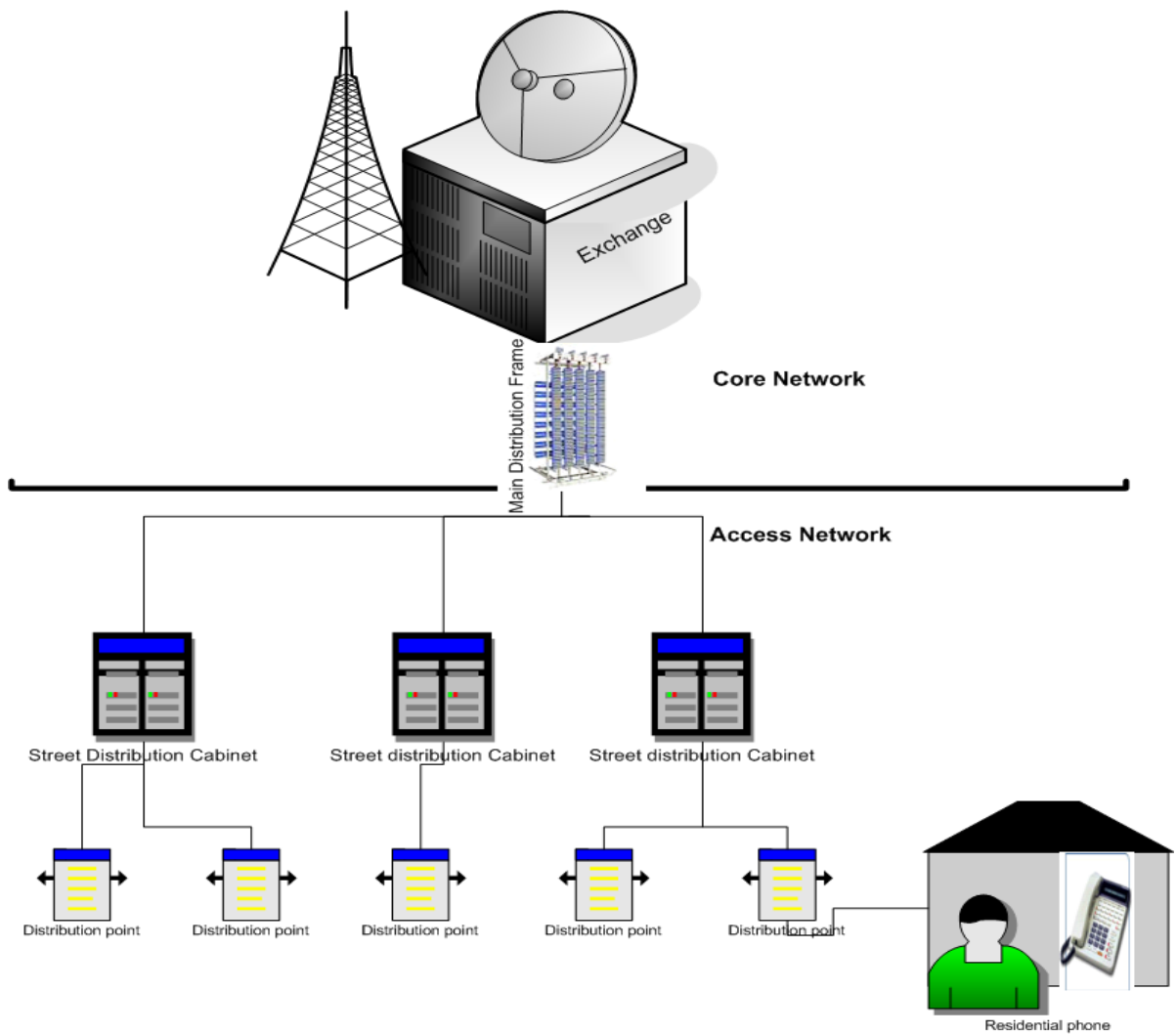


Figure 6.2.1 Fixed line telecommunications network.

On the core network, each connection needed a telephone number and a port on the exchange. On the access network, each connection needed the following positions on the network:

1. Strip and line on the main distribution frame (MDF)
2. An exchange and distribution point position on the street distribution cabinet (SDC)
3. One position on the distribution point (DP)

A sales list detailing the spare infrastructure was provided to the sales division in each of the different regions. Sales lists were made available on a weekly basis.

6.2.2 The spare infrastructure BI solution

The company's BI requirement for spare infrastructure does not include any numeric values (facts) - for example, total sales. From the dimensional model in figure 6.2.2 below, it is clear that there were no facts; the information consisted of a combination of dimensions. A fact table, as explained in chapter 3 is, therefore, not suitable for this solution. A factless fact table is used to join different dimension tables for reporting purposes. Allen and Terry (2005:411) state that a factless fact table consists of a number of foreign keys connecting different dimension tables for reporting purposes.

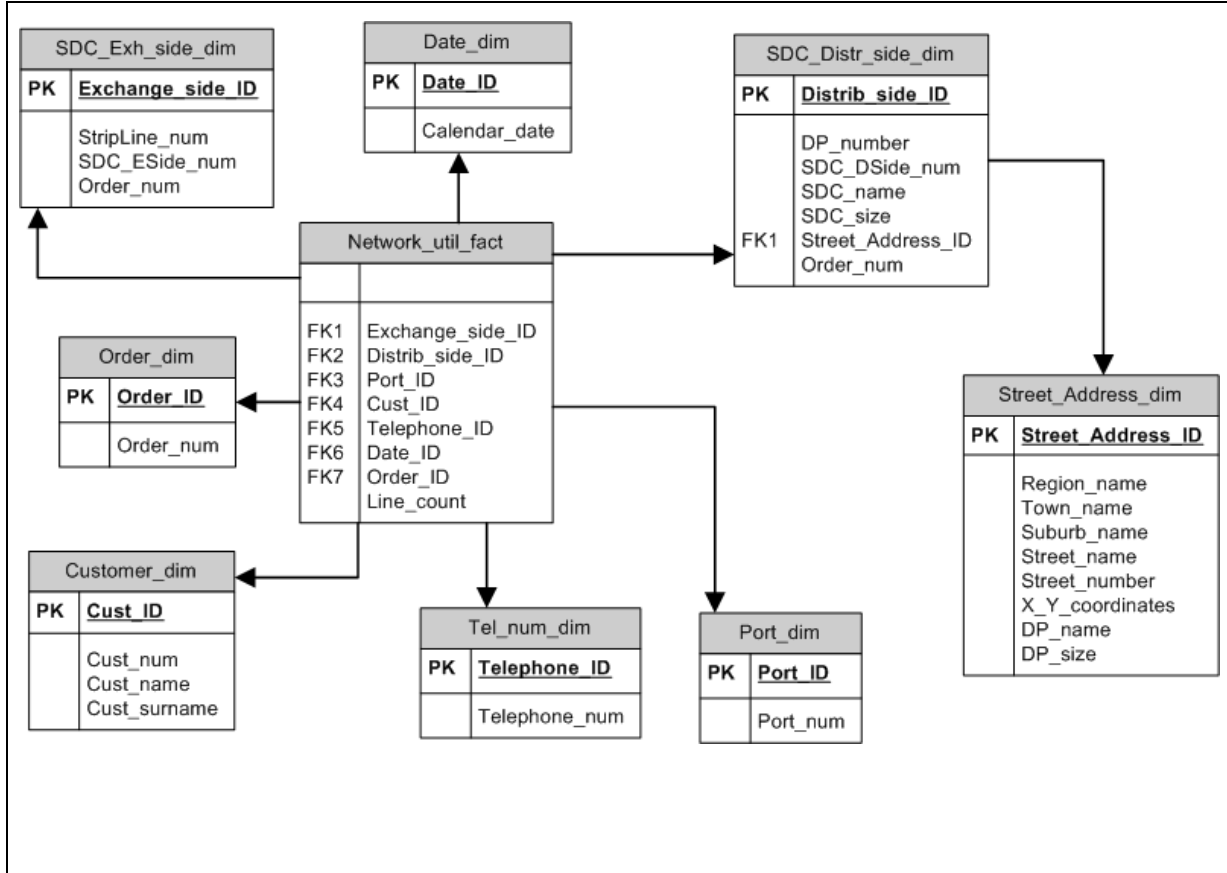


Figure 6.2.2 Infrastructure dimensional model for spare infrastructure.

6.3 BI culture as part of corporate culture

The aim of a BI implementation is to get all the business people to use the BI solution. A culture of understanding for the need and use of BI is required in a company which is implementing BI. Salisbury (1996:162) compares culture with electricity because it is invisible but very powerful. The energy present in electricity can be dangerous, but could be very valuable at the same time. Bodley (1994:1) believes that culture involves at least three components: what people think, what they do and the material products they produce.

BI end-users in the company had very little understanding of the need for a BI solution. These end-users came from companies where BI may have been implemented, and they probably had different experiences with it. These experiences may have been positive or negative. Positive experiences could benefit the new implementation, whilst bad experiences could be fatal to actual project implementation. Ravi (2008:218) states that many BI projects have failed for the following reasons:

- *Misinterpretation of user requirements*

When starting the implementation of a BI project, companies using existing SDLCs normally embark on an extensive project. This would normally be to add as much value as soon as possible. This approach can also be very risky at the same time, because the larger the requirement, the more the possibility for making mistakes. Large requirements can create expectations that cannot be fulfilled immediately. Therefore, the time it takes to design and build this BI solution creates frustration that

negatively influences the implementation of the project. Adelman and Moss (2000:143) call this large scale implementation approach, the Big Bang BI project.

- *Reconciliation of data*

Data profiling can be much more complicated than previously supposed, because of the volume of data that the profiling team needs to attend to. Because of the high volume of data that need to be profiled, some mistakes are always made. The mistakes associated with the new BI solution result in poor appreciation of the BI solution, which discourages its use.

- *Resistance to BI implementation*

BI end-users may have experienced one or more of the above circumstances. Because these circumstances were negative, a negative perception of the entire solution may have developed. This view could create a certain corporate culture that does not encourage the implementation of BI. Burlew (2006:72) states that “culture is created naturally as well as automatically. Culture is automatically created out of the combined thoughts, attitudes, values and views of the people in the group”.

People have different experiences with a BI solution, some positive and others negative. Schein (2009:105) states that as soon as a cultural change is announced in a company, discomfort and anxiety result, because employees realise they may have to give up their beliefs, values, assumptions and attitudes, and probably need to learn some new ones. If the process is left to

happen automatically without any guidance, the culture could still form, but it may not benefit the implementation and usage of the BI solution.

Lowney (2003:209) states that “a culture can blind people to facts that don’t match its assumptions ... and ... an entrenched culture can make implementing new and different strategies very difficult”. When the corporate culture of the company does not benefit the implementation of a BI solution, it does not mean that the corporate culture as a whole is negative. It might just mean that the culture does not benefit the implementation of BI. Du Plessis and McDonald (2007c:229) therefore consider that the culture in a company - directly related to the implementation of a BI culture - can be termed a BI culture. In order to successfully implement a BI solution, there needs to be a positive BI culture.

6.4 The role of BI literacy in a company’s BI culture

Quinn (2003:5) states that in most organisations non-technical business users comprise 80 to 90 percent of all information workers, with analysts and power-users making up the remaining 10 to 20 percent. Pareek (2006:105) states that the successful implementation of BI depends on the overall increase of computer literacy in the company. That means that non-technical users need to be increased in number. Du Plessis and McDonald (2007c:230) refer to the ratio of 10:90 as the BI literacy ratio. They also state that a BI culture can be a direct result of BI literacy (2007c:230).

When implementing BI in a company, the report writing and online analysis is normally done by the information workers, analysts and power users. Quinn

(2003:5) states that when companies select BI tools, they look at sophisticated analytical features and functions, that satisfy the technical users only. That means that the other 80 to 90 percent of business people are not exposed to the new BI tool and just receive the results. This approach might be an easy way to quickly reap the benefits of the newly implemented tool, but it is also creating a culture where a large proportion of non-technical staff believe that writing reports and doing online analysis is not their job.

The literacy of the non-technical users on the BI tool is important for the creation of a BI culture in the company. Tyner (1998:99) states that literacy is the ability to locate, evaluate, use and communicate. It is difficult to split the two realities, literacy and culture, because the one could be the result of the other and leading the business to not using the BI solution. Therefore, in order to create a positive BI culture as part of corporate culture in a company, a model is needed to guide the process of enhancing and measuring the culture within an organisation.

6.5 BI literacy and culture maturity model

A BI literacy and culture maturity model was needed to measure the BI literacy and culture maturity of business users. This model had to suggest ways to increase the BI usage level in an organisation.

DuMoulin (2007:1) states that if data or information cannot be defined, they cannot be controlled. If they cannot be controlled, they cannot be measured and

therefore, cannot be improved. BI literacy and culture needed a model that can define, control, stabilise and measure BI literacy and culture in the company.

Dahinden (2003:105) states that models are powerful tools for representing complex structures and relationships in order to better understand or visualise. The aim of the BI literacy and culture maturity model was - by means of definition and control - to stabilise and improve the literacy and culture of BI in a company.

6.6 Data analysis

Cohen *et al.* (2009:316-329) state that when analysing a case study, care should be taken that the most relevant events are used for examination. The event deemed relevant to this case study was to establish ways to mature the company in terms of BI literacy and culture. Kumar (2005:114-176) states that researchers should avoid concentrating too much on pre-conceived ideas or theories in analysing the data, when using the case study research strategy. Reinard (2007:236-296) asserts that this is the data analysis approach, where the theory emerges from the data, and which means that the theory is grounded in the data.

6.6.1 Data categorisation

Goulding (2002:74-76) states that open coding prescribes that, when labelling data units, the labels should not come from the literature, but from the data. The labels derived from the case study used for this research, were six levels of

BI literacy and culture maturity. Figure 6.6.1.1 (below) shows the six maturity levels for BI literacy and culture that emerged from the case study.

Understanding the maturity model can help to control the company progress through the levels. It is important to know at what level the company is operating, and why it operates at that level. This will help the company to understand how and when to move to the next level.

As mentioned earlier, BI literacy is placing the company at a specific usage level. When a company does not address this literacy issue, it will become a culture. This means that business people believe that they do not need the training, because the BI team is responsible for writing reports. This produces the argument that literacy is placing a company at a BI usage level, but culture keeps it at that usage level. Figure 6.6.1.1, below, represents the BI literacy and culture maturity model.

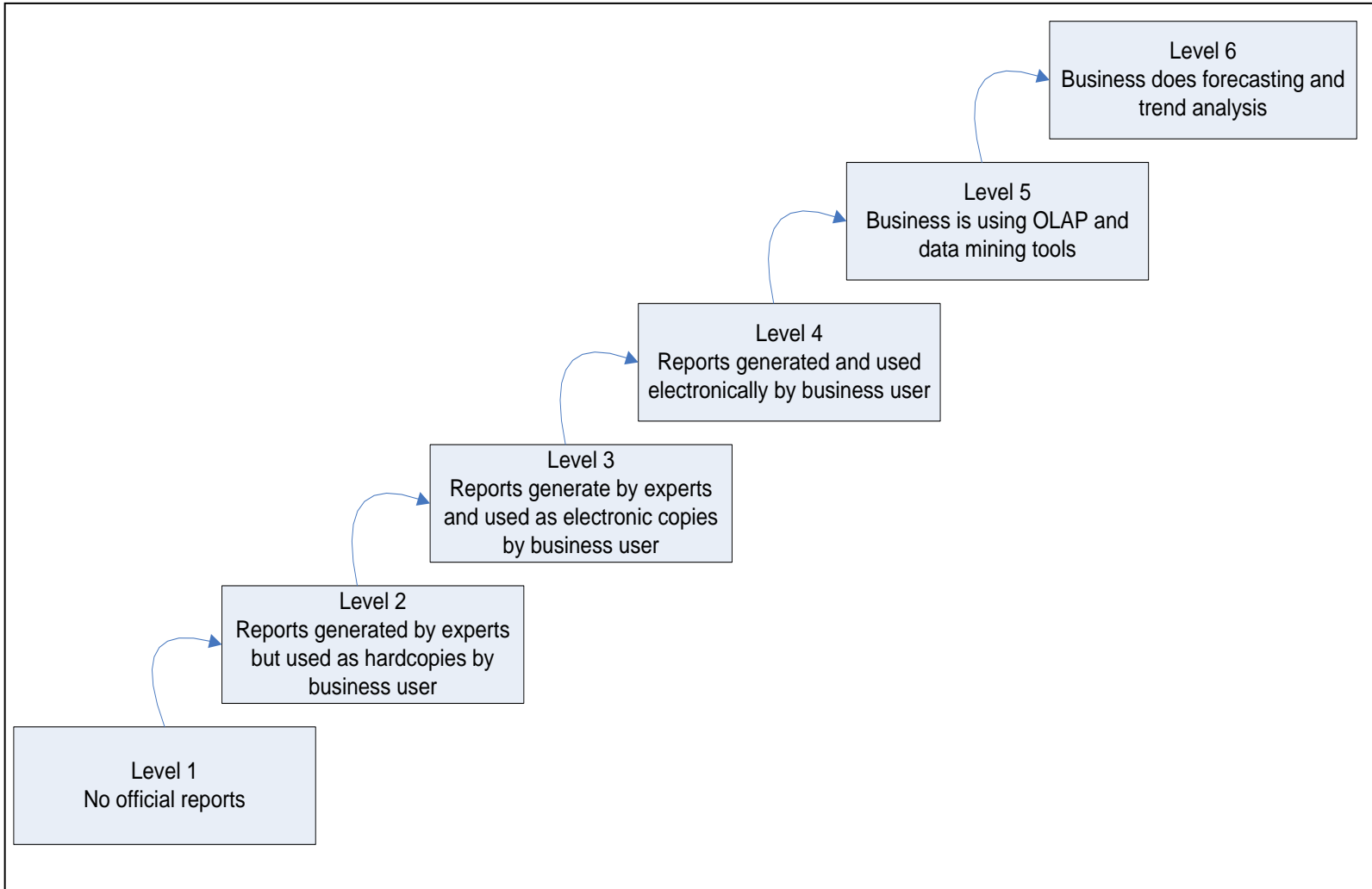


Figure 6.6.1.1 BI literacy and culture maturity model.

The six different levels can be summarised as follows:

Level 1 No formal measurement

Level 1 does not occur in most new companies and is normally not a level that lasts very long in bigger companies, because it is very hard to manage any organisation without any form of measurement. The business management team, therefore, plays a major role in implementing measures to manage the performance of the organisation. Reports are in some instances built into OLTP systems, resulting in a lack of integration, because each OLTP system has its own reporting solution, which is isolated from the other systems.

Level 2 Reports generated by experts, but used as hard-copies by business

Business people at this level know that information is printed out for them. To run and use reports electronically forces people to step into new roles. This is a cultural change which could be initiated and improved by creating a new category of staff with both ICT and business skills (BI division), as mentioned earlier in this chapter. The business people receive training to use the electronic reports and the BI staff are available to assist with any problems that might be experienced.

Level 3 Reports generated by experts and used as electronic copies by business

This level is very prominent in organisations in developing countries, where the BI skills level amongst business people is low. The low BI literacy could be

a result of the culture that exists in the organisation. In companies at this level a culture exists where only BI people are responsible for using BI tools.

To move to the next level, a new culture is needed where business people are responsible for using BI tools and only BI people maintain these tools. This new culture needs to be driven by the business management team which should lead by example. Business management should use electronic reports for management meetings. Generating electronic reports could be difficult at first, but the BI team can play a substantial role in educating relevant staff.

Level 4 Reports generated and used electronically by business users

This level takes the longest to reach, because it needs more than just the PC skills of the business people. Previously, reporting was done by someone else (BI team), but now business people should do it for themselves. One way to get to this stage more rapidly is by only building a BI solution that is needed urgently, or by using this information for performance appraisals. The business people receive training to generate or run the reports with a proper BI tool, and the BI section is available to assist with any problems that might be experienced.

Level 5 Business is using OLAP and data mining tools

Moving from the use of reports to the use of online analytical processing (OLAP) could mean a lot for the business. This is the first step towards a complete business intelligence solution. This level of maturity within the business requires that the data warehousing maturity model (Eckerson, 2004:1)

is at level 5 (Adult) (see figure 6.6.1.2 below). Data mining could result in new information that could be modelled back into the BI solution. This level requires creative BI analysts who are very well equipped with both OLAP and business skills.

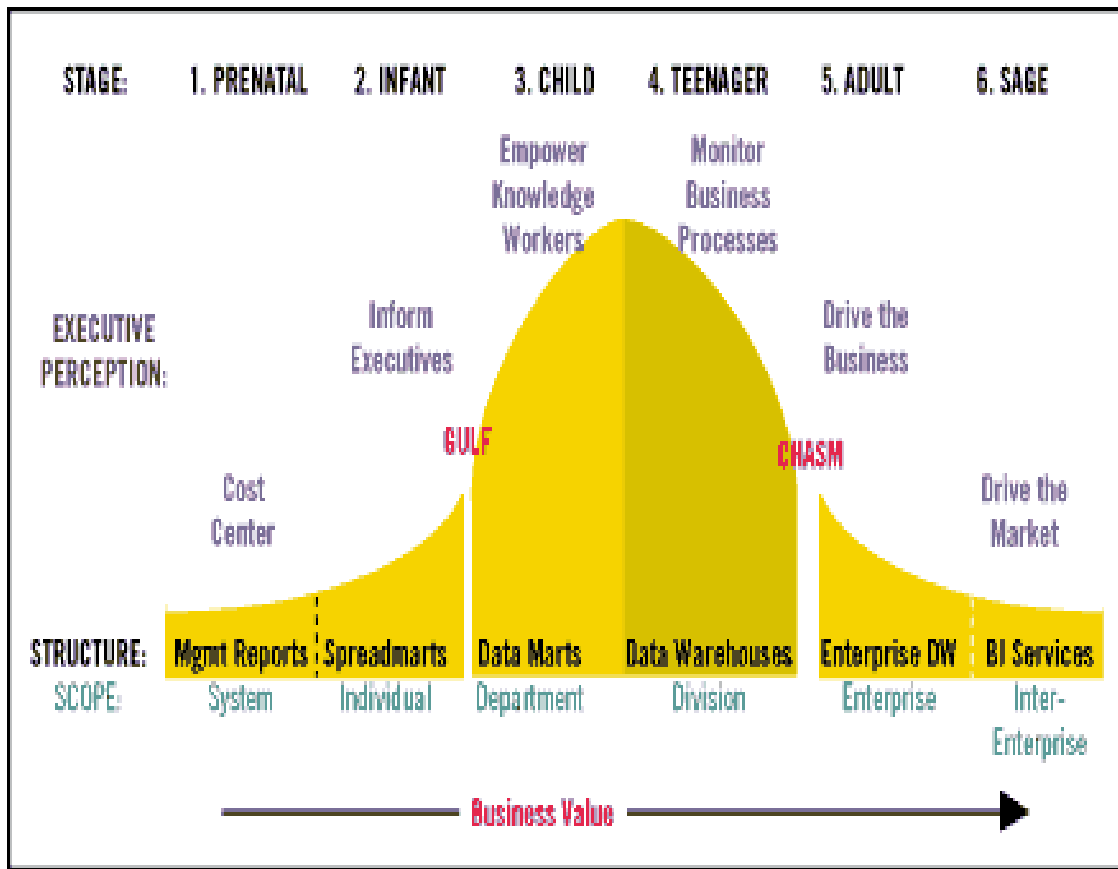


Figure 6.6.1.2 Data warehousing maturity model (Eckerson, 2004:1).

Level 6 Business does forecasting and trend analysis

Level 6 is to a certain degree not dependent on the literacy of business people, but is a strategic decision made by top management and is therefore dependent on the appointment of statisticians in the company. The approval of these statisticians is normally dependent on the value delivered by the BI solution up to level 5.

6.7 Conclusion

This chapter discussed some of the challenges faced by companies in building and maturing a BI solution. The BI literacy and culture maturity model was created to monitor the growth of the BI maturity of the organisation. The model ensures that the technical part of the BI solution grows at the same pace as the human (soft) part of solution. A BI solution that is not used by the business cannot be described as a successful BI solution, even though it may be technically sound. Therefore, BI literacy should be stimulated throughout the BI implementation process. It is also very important that value is added throughout the implementation process.

Utley (2008:6-12) argues that BI has no intrinsic value. Instead, it should be used properly by people in the business to enable the company to reap the benefits. It is, therefore, important that a BI solution is used by business people to reap the benefits that this kind of solution brings to a company.

CHAPTER 7

Business intelligence strategic interactive and incremental framework

7.1 Introduction

In the previous three chapters, two aspects of BI were addressed. Firstly, the focus was on implementing a SDLC to facilitate the implementation of BI in a company with a limited budget. An architecture that supports the DWDW lifecycle model for the incremental implementation of a data warehouse was developed. Secondly, BI literacy and culture were emphasised in the telecommunications company used for this study. This led to the development of the BI literacy and culture maturity model for establishing end-user confidence in the BI solution.

This chapter concentrates on placing the maturity model and the DWDW lifecycle model into one conceptual framework. This framework will be discussed using the case study in the South African telecommunications company, which was used for this thesis.

7.2 Background

The telecommunications company in this study had exclusive rights as the sole landline provider for a limited period. The company had several implementation targets from the government, which needed to be achieved in order to retain exclusive rights for its telecommunication network. One of the targets was to render new telephone services in the under-serviced areas of the country. Complying with these targets had in some instances caused an increase in debtors' days, which would probably have led to bad debts for the company. Whilst complying with the targets, the company also had to prepare itself for competition (Du Plessis & McDonald, 2007c:153). Bad debts and the resulting negative cash flow could have a negative effect on the company's future profits and in attracting new investors. A BI solution was thus needed to provide management with the necessary information for making decisions about maintaining a balance between acquiring new telephone services and maintaining the financial health of the company.

The first priority of the company was to achieve the targets set by government. Therefore, the company had to install new lines with a limited infrastructure budget. The company thus concentrated on the optimal use of existing infrastructure and only provided new infrastructure where it was required. The BI solution has played a vital role in managing the existing infrastructure.

7.3 A new BI strategy

The telecommunications company needed a new business strategy. A strategy needed to be based on issues, such as profits and market share. Moss and Atré (2003:31) consider that a BI solution should always be business-driven rather than technology-driven. For the BI team it is important to align the BI strategy with the business strategy, because BI serves as an enabler for the business strategy in a company (Williams & Williams, 2007:106).

Aligning BI strategy and business strategy was not always easy in the telecommunications company, because the business strategy was formulated by the business management team and the BI strategy formulated by the BI team. The business strategy in the new company was, to a certain extent, not common knowledge in the company. The company was still newly established and was very concerned about company information being leaked to the competition. It was, therefore, important for the BI management team to build a trust relationship with the business management team. This trust wasn't built immediately. It was developed over years by providing timely and relevant information to the business management team. Howson (2007:116) compares the relationship between the BI team and the management team of the business with the Yin Yang symbol. The Yin Yang symbol was used in early Chinese philosophy (see figure 7.3a below).

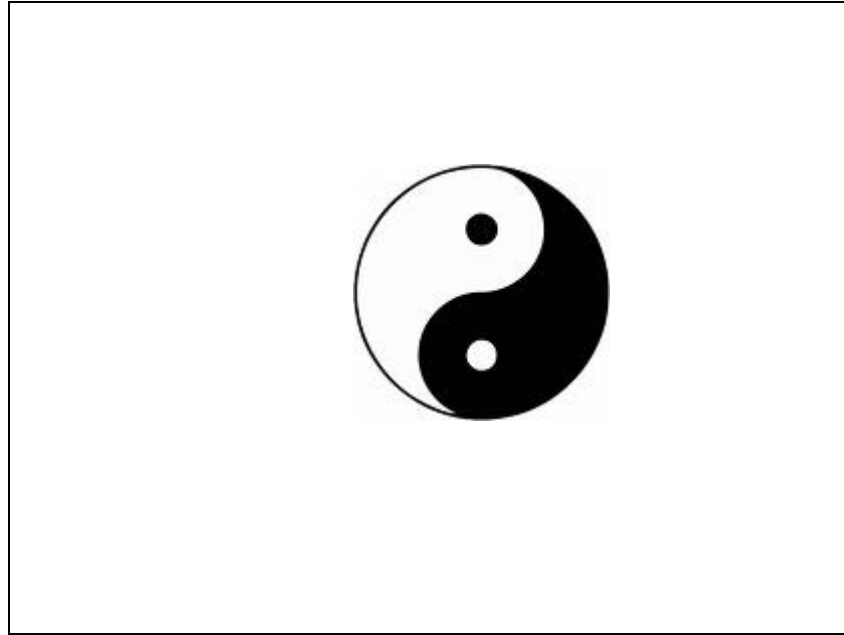


Figure 7.3a Yin Yang symbol (Howson, 2007:116).

Howson (2007:116) states that the white portion is called Yang and the black portion is Yin. Yang represents “movement, initiative, heat and fire”, while the Yin represents “passiveness, cold, and water”. Howson (2007:116) argues that the relationship between BI and management is like the Yin Yang. The Yin Yang, whilst reflecting difference, also conveys a balance that cannot exist without one of the two parts. In the black part of the Yin Yang is a white dot and vice versa - which demonstrates that interaction creates strength.

Ouyang and Wang (2008:587) state that a conceptual framework can be used to present a preferred approach to an information project. Figure 7.3b (below) presents the business intelligence strategic iterative and incremental framework (BISIIF) that was developed to facilitate the alignment of the BI strategy with the business strategy (Du Plessis & McDonald, 2007c:154). Gray (1969:25)

states that a conceptual framework reduces the scope of the field to a single preferred approach. The BISIIF consisted of the steps that were followed to grow BI in the company (Du Plessis & McDonald, 2007c:154). These steps were supported by the previously discussed two models which formed part of the framework - namely the DWDW lifecycle model and the BI literacy and culture maturity model. Each of the components in this framework is now discussed.

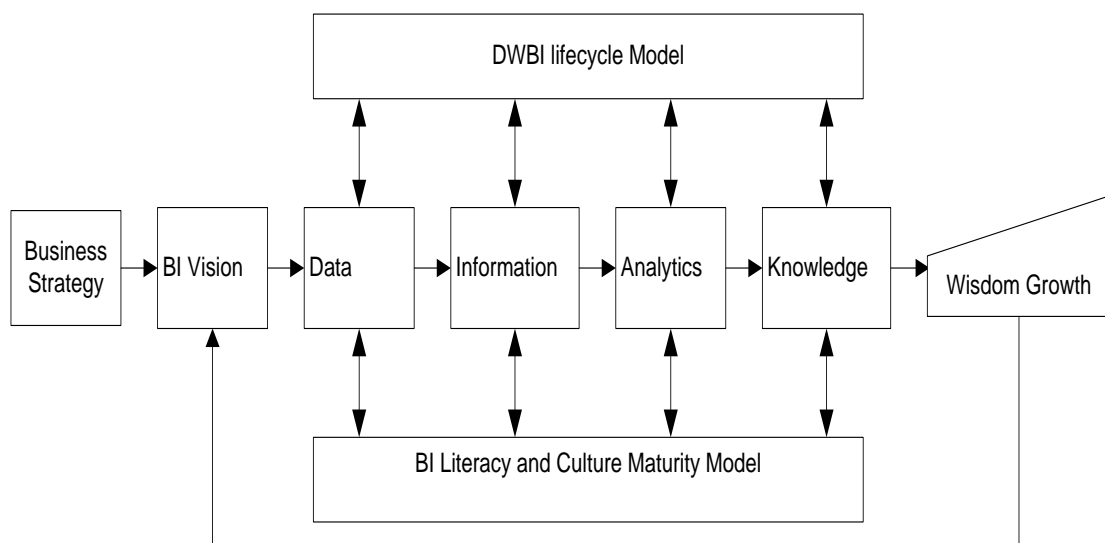


Figure 7.3b Business intelligence strategic iterative and incremental framework.

7.3.1 Business intelligence vision

Business intelligence, as with all other strategic initiatives, should have a clear vision. Dennis *et al.* (2005:21) state that the vision of the company should be to

drive the definition of the BI Framework and to assist in selecting the appropriate methodologies, key phases, steps and roles along the pathway - to deliver decision support information to the business. Monteith (2005:151) believes that BI is responsible for the transition of data into wisdom, through the following steps: Data, Information, Analytics, Knowledge and Wisdom.

These steps were incorporated into the above-mentioned BI strategy framework, and the implementation is explained by means of the following examples from the implementation process in the case study.

7.3.2 Data

Frentzos and Theodoridis (2007:217) state that “with implementation of database management and advances in data storage technology, companies have been collecting, processing, storing and accumulating vast amounts of data about people, locations, transactions, concepts and events - that could be analysed”. The existing OLTP systems formed the basis for the first iteration of the framework. Because the BISIF is iterative in nature, after the wisdom step, new data needs can be identified which could result in changes in the OLTP systems. This could, for example, include the need for an economic centre as discussed later. Courtley *et al.* (2005:211-229) states that wisdom could highlight data needs, which could in turn require a change of the BI strategy, that facilitates the process of collecting the data needed into the OLTP systems.

7.3.3 Information

As the company was doing business, a huge amount of data was collected on the OLTP systems of the company. Each data element is part of a transaction and, therefore, did not produce any meaning by itself. Aspden (2004:129) states that on its own, a data element such as ‘product’ does not make sense unless presented with other elements of data. Only accumulation of data into a meaningful context provides information.

ETL processes were used to provide BI specialists in the company with the capability to extract data from the OLTP databases and transform the data into information. Figure 7.3.3.1 (below) is an example of the data model that was created as part of the process and was used to instil wisdom in the company.

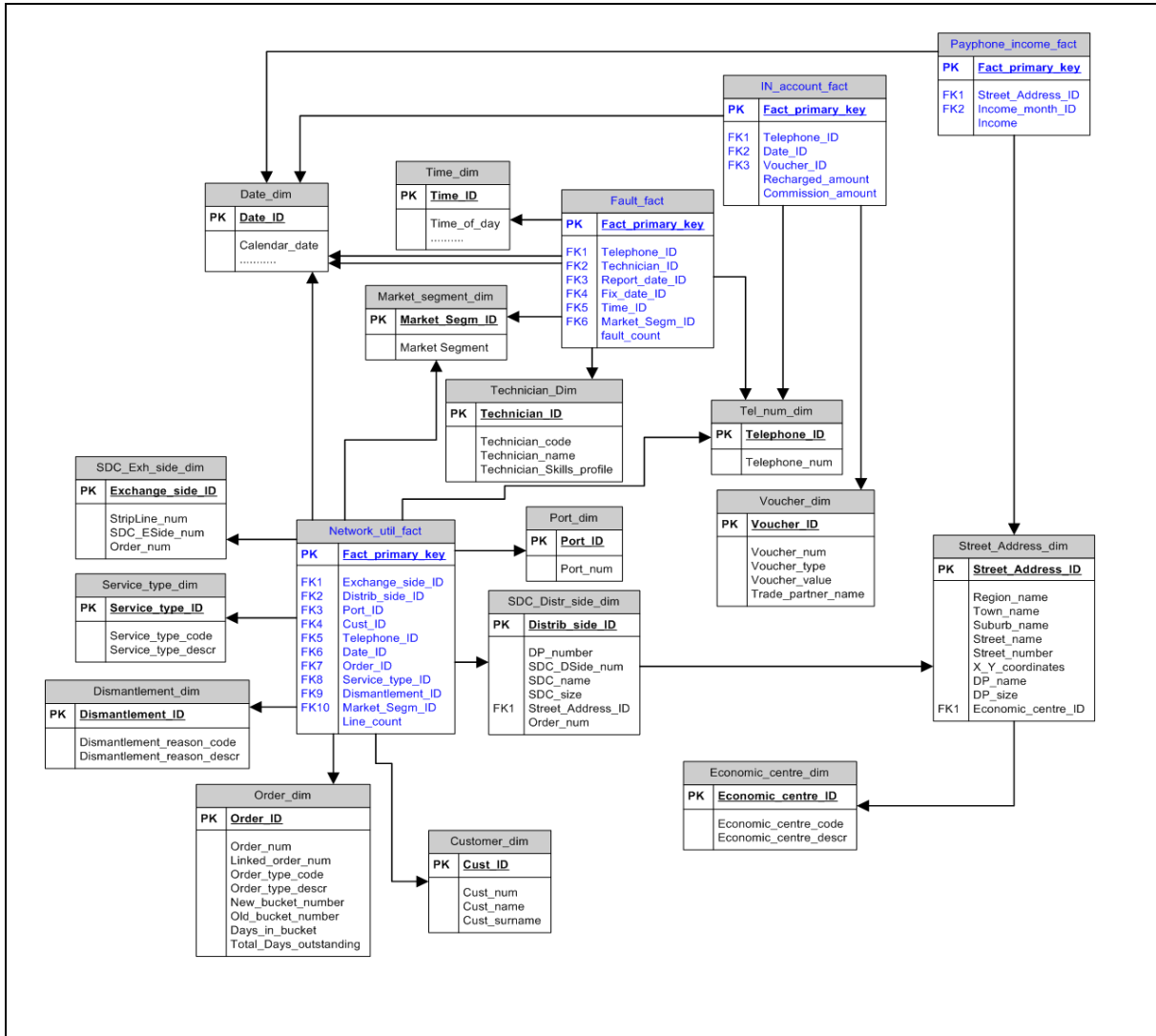


Figure 7.3.3.1 Technical infrastructure management data model.

The overall data model can be divided into two models - the technical infrastructure management model, and the financial management model. This thesis concentrates only on the technical infrastructure management data model. This model was created out of several iterations of the BISIIF, with each iteration delivering new data. The installation and fault management process are the two business processes measured in the telecommunications

company. These two processes will be discussed, in an attempt to explain the BISIIIF.

7.3.3.1 Installation and fault management

Telephone service installation and repair were two core processes in the telecommunications company in the case study. The first iteration of the BI strategic framework delivered the information to enable the business to manage these two core processes through the mean time to install (MTTI) and mean time to repair (MTTR) measures. The “Network_util_fact”, “Order_dim”, “Service_type_dim”, “Tel_num_dim”, “Fault_fact”, “Technician_dim”, “Date_dim”, and “Time_dim” tables in figure 7.3.3.1 were used to measure MTTI and MTTR in the company.

After the first iteration of the BI strategic framework, it became clear that there were opportunities to cut installation and fault reparation cost and time. Personnel responsible for the telephone service were dispatched to the different faults and orders. The sequence of dispatching to these tasks was done just by starting from the top of the list and working down to the last fault or order on the list for the day. The only criteria for priority on the list were the fault reporting time and required installation date. This resulted in technical personnel wasting time and money by driving long distances to finish a task in one area, and then attending to a task in another area, and thereafter returning to the first area for the next task.

The skills of the technician were also of importance when dispatching a technician to a task. Technicians had different skills and therefore sometimes did not have the right skills for a particular task. Larger telephone systems like PABX systems require specialised skills. The “Technician_skills_profile” attribute was, therefore, added to keep record of the skills of the technicians.

There were also priorities between the different market segments, which were kept in “Market_segment_dim”. A technician could for instance repair a residential line prior to a business line. One can argue that the parties were paying the same rental, so why should there be rules in terms of priority customers? However, residential and business services differ in terms of usage. Businesses and corporate customers make more phone calls than residential customers. In addition, one day without a telephone service for a business and corporate customer could result in financial losses.

Another service that could damage the business of the customer was data lines. Corporate customers with large information technology solutions rent data lines from the company, and online business is done using these data lines. This type of service was stored in the “Service_type_dim”.

A system was required to help with the prioritising of tasks for the technical staff. The four factors that were taken into consideration when prioritising the tasks for technical staff, were:

- Position of the technician (where did the technician complete the last task)

- Market segment of the next task
- Service type of the next task
- Skills of the technician

By making use of a proper dispatching system that was built on top of the data warehouse, the company reduced transport costs by about twenty five percent.

7.3.3.2 Billing

The next important process was the billing process. There were two different billing processes, namely the post-paid billing process and the pre-paid billing process. The third iteration of the BI strategic framework delivered the information to manage these two billing processes. The billing method was the only difference between pre-paid and post-paid phone services. With the post-paid service, the customer was billed at the end of the month for the service rendered that month. For the pre-paid service, the customer, in order to be able to make calls, needed to load his/her account on the Intelligent Network (IN) platform. Service fees were deducted at the beginning of every month, and fees for airtime used were deducted per call from the funds available on the specific pre-paid account.

Recharge vouchers were sold on a commission basis by the trade partners of the company. In figure 7.3.3.1 “IN_account_fact” and “Voucher_dim” were used for the reporting of commissions paid to trade partners. The trade partners

bought the vouchers from the company and sold them to the customers. Records of numbers of vouchers sold to a specific trade partner, were recorded on the system. Commission was only paid to trade partners when the customer recharged his/her account on the IN platform.

Based on the account balance of a pre-paid account, a pre-paid customer could be in one of the following states:

- Active (money available and the customer still able to make and receive calls)
- Inactive (not enough money available for service fees for the new month)
- Locked (pre-paid service inactive for one month)
- Dismantled (pre-paid service is locked for one month)

Information was provided to identify the inactive pre-paid services, and that list was loaded into an auto dialler that dialled these customers and reminded them to recharge their prepaid services.

The billing mart provided the business with information to analyse income and to do projections on income trends, which took the BI solution to the wisdom step.

7.3.3.3 Dismantlement

It was important to create dismantlement orders on the order system for those pre-paid customers that were in the dismantlement state in order to free up the infrastructure for new services. The reasons for dismantling telephone services - pre-paid as well as post-paid - could be divided into two main groups. The first was customer-initiated dismantlement, and the second company-initiated dismantlement. Company-initiated dismantlement was relevant in only two instances. The first being non-payment and the second transfer of a telephone service. Non-payment was different for post-paid services. When the customer did not comply to payment agreements made with the company, the services were dismantled and the network infrastructure could be used for a new customer. The “Dismantlement_reason_code” attribute in the “Dismantlement_dim” in figure 7.3.3.1 was used to capture the reasons for dismantlement.

As far as transfers were concerned, all existing customers - pre-paid and post-paid - sometimes moved from one exchange area to another. To free the network infrastructure at the old address, a dismantlement order was created. For the new address, a new order was then created at the same time, and the two orders were linked to ensure that both were done simultaneously (see figure 7.3.3.1 for the Linked_order_num in Order_dim table).

Customer-initiated dismantlement could be for several reasons. Some customers simply could not afford the service any longer and would in most instances arrange to pay the last bill in affordable instalments. Another

customer initiated dismantlement can also be referred to as churn - this being an unsatisfied customer. XWave (2004:1), a company that has undertaken research on BI in the telecommunications industry, believes that the greatest potential market and source of competitive advantage may not be in new products or sales programs, but in current customers. Therefore, dismantlement reason codes were tracked against the area where the service existed. On several occasions it was found that a service delivery problem caused several customers to dismantle their service for the same reason - for example, dissatisfaction with the service.

7.3.4 Analytic

7.3.4.1 Analysis of the telephone market

While combining data and trying to create information was extremely useful, separating or regrouping information extended the value of information further. Applications that had OLAP capabilities (Actuate) provided users with the ability to analyse information and determine relationships, patterns, trends and exceptions (see figure 7.3.4.1 below). The analysis of available information was done as follows.

Corporate, business and residential market segments existed in the company and marketing was done in these segments. The sales sections were also divided accordingly. These market segments worked very well, because there were definitely different needs in these segments and they were divided

according to these differences. The residential segment, however, could be split further, because of the differences in income of households and different behaviour patterns as a result of that. Analysis in this market revealed that the different income levels were influencing the behaviour of the customer when it came to renting of a telephone service. In the corporate and business segment a telephone service was an important business tool, but in the residential segment, in some households, it could be seen as a luxury, as food and shelter were of much higher importance. The analysis revealed that the income of households influenced the need for a telephone service in the households. It was also clear that people were staying in different suburbs based on their income.

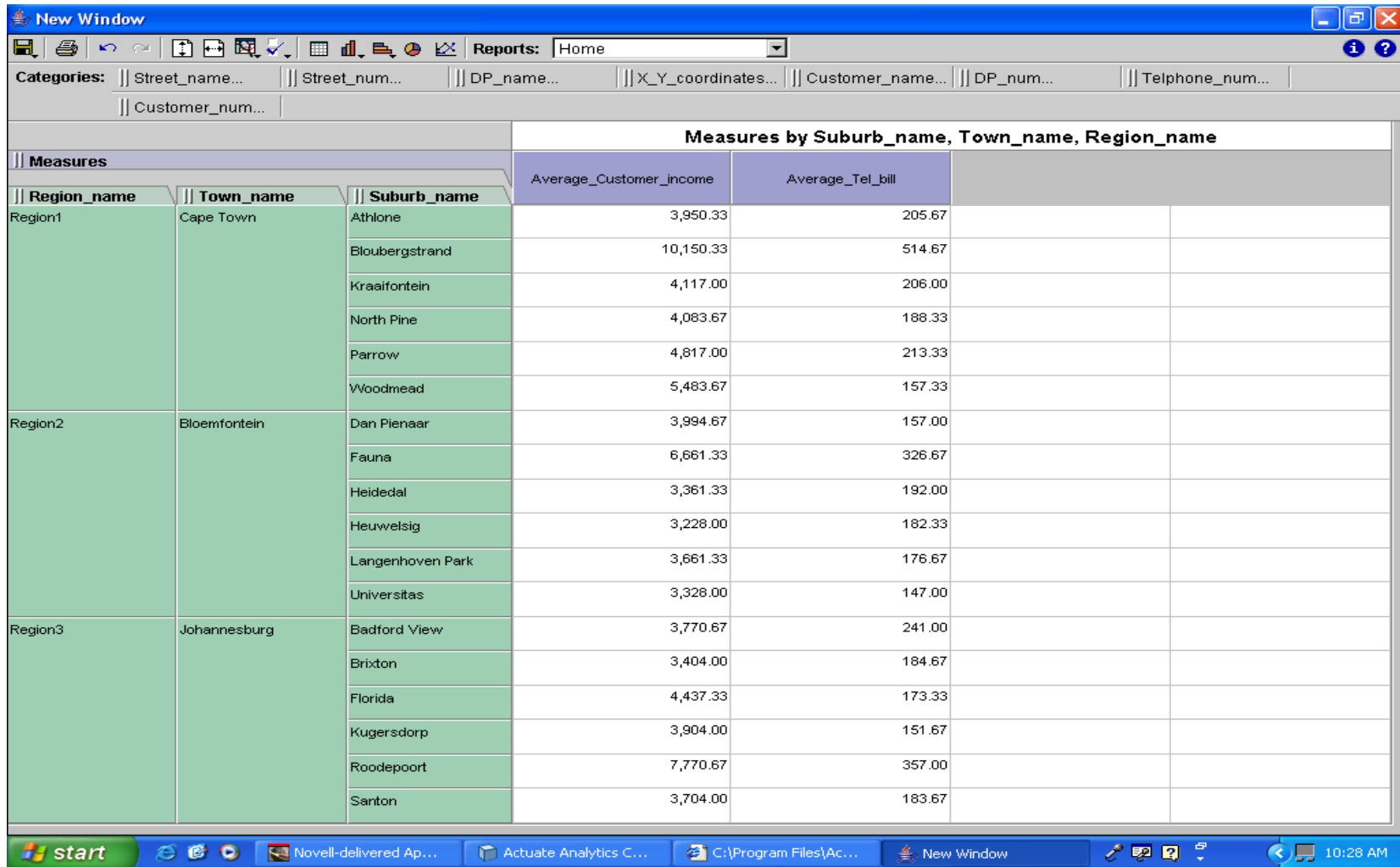


Figure 7.3.4.1 Actuate OLAP data cube for customer billing information.

7.3.4.2 Market segmentation

The need, therefore, arose to split the residential market segment into different economic centres. With a landline telephone company, large amounts of capital were invested into infrastructure and it was, therefore, important to invest in areas where the infrastructure would be used in order to guarantee a return on the investment. The income information was provided by the customers when they applied for the telephone service. This was used, and an average was calculated for a specific area. That area was then allocated to an economic centre based on the average income, dismantlement rate and telephone usage in that area. Three economic centres were decided upon.

- Economic centre A - Top income group
- Economic centre B - Middle income group
- Economic centre C - Lower income group

In the case of a totally new area, the economic centre was initially allocated based on the prices of the stands, rules dictating time periods for building after purchase, and, if applicable, the minimum price allowed for the building of houses in that area. This information was used until more income information became available, which only happened when customers started applying for telephone services in these new areas. The rating of an economic centre was repeated every financial year.

7.3.4.3 Marketing based on the market segmentation

The economic centres were not only used for the building of new infrastructure, but also to sell new products and services in the residential market segment. Previously, a list of addresses in the under-serviced areas with available network infrastructure was available from the BI solution to the sales team. After the exclusivity period expired, a list of addresses of possible customers in the different economic centres, with spare network infrastructure, was also available to the sales team.

7.3.4.4 Planning infrastructure based on market segmentation

The classification criteria used to rate the different areas were flexible and could change based on the inflation rate. A type 1 slowly changing dimension table (Economic_centre_dim) was used for the economic centre rating (Kimball & Ross, 2002:75). Every year the economic centre records were overwritten in this dimension table. There was no need to keep a history of the economic centres, as it was only used for marketing purposes. The records in the dimension table consisted of three columns: the surrogate primary key, the economic centre code and the economic centre description. The description included the minimum and maximum value of the economic centre.

The analysis of the market segments was taken one step further. The trends of the previous installations were analysed each year, and this information was used to

budget for the number of lines that needed to be installed in the following year. The budget established for new sales was according to the needs in the different economic centres. This budget gave network planning staff a clear understanding of the network access need.

7.3.5 Knowledge

The next level of elevated understanding was knowledge. Knowledge can be produced from data or information by applying logical inferences and data analyses. Hidden trends can easily be identified and displayed when using data mining techniques for data analyses. Knowledge management can be explained using a payphone case study.

Payphones represent a market segment on its own. During the exclusivity period, targets were set by government to install payphones in the under-serviced areas. In some instances the main aim with the installation was the achievement of targets and not the profitability of the service. The “Payphone_income_fact”, “Date_dim”, and “Street_address_dim” tables in figure 7.3.3.1 enabled the company to measure the profitability of a specific payphone. X and Y coordinates were gathered for all payphones, and the income of the payphones was measured on a monthly basis. The GIS personnel in the company linked their geographical system to this information in the data warehouse. On the GIS system it was possible to see representation of all payphones in the country, plotted on a map of South Africa. It was also possible to drill into a more detailed map of the

particular city to see the payphone and the surrounding area (buildings and streets). Month-specific income could be displayed for each payphone.

This BI solution simplified the management of payphones. It was easy to see where profitable and unprofitable payphones were. Management could see how many payphones were close to a particular payphone, and could then relocate them to areas where they could be profitable. This was a very good example of a BI solution at a knowledge step. With information it was only possible to see what had happened, and based on that information a decision had to be made. Whether it was the right decision could only be seen in a month or two. With a company's BI solution at the knowledge step, the company can plan ahead at much smaller risk.

7.3.6 Wisdom

Wisdom is the utilisation of accumulated knowledge. By utilising knowledge, a higher level of understanding of the data was created. The BI strategic framework was not a once-off exercise.

There was more than one iteration of the framework. As the company worked through the framework, each interaction increased the wisdom of the company. As the wisdom increased, new requirements developed, and this is how the BI solution was built into the company.

Service Repairs	Area 1	Area 2	Area 3	Area 4
Business	70	20	50	42
Residential post-paid	53	54	300	12
Pre-paid	12	32	254	14
Payphones	5	23	12	12

Table 7.3.6 Telephone service repairs for January 2004.

Table 7.3.6 (above) represents the repairs of the different services in the different demographical areas. The knowledge gained from this table is that residential

post-paid and pre-paid customer services have high numbers of service repairs reported in Area 3.

Wisdom relates to analysing the knowledge and adds human intellect to analysis of the knowledge. Area 3 is part of an urban area with many flats, resulting in a high density of residential post-paid and pre-paid services. That meant that the number of services reported was high, although the percentage of the total number reported was low. This situation led to a new requirement for a data mart that included the total telephone services in these areas. This is the reason why BISIIF has a link from Wisdom back to the BI Vision.

7.4 BI literacy and culture maturity model

A BI literacy and culture maturity model was developed to measure the literacy of the business users (Du Plessis & McDonald, 2007a:134). This was done in tandem with the steps of the BISIIF, because it was important to know the culture and maturity level before the BI team could move to the next step of the BISIIF. It is important that the BI solution grows at the same pace as the culture and maturity of the company. If the BI culture or maturity is not sufficiently developed in an organisation to move to the analysis step of the BI strategic framework, then it does not make sense to move ahead with the technical part of the framework. Ignoring the BI cultural and literacy side of the BISIIF could sink the implementation of the BI strategy.

7.5 Conclusion

This chapter discussed some of the challenges faced in implementing a telecommunications BI solution in a developing country. The business intelligence strategic iterative and incremental framework (BISIIF) was used to implement the BI strategy in the company. This framework is iterative and incremental in nature and, therefore, does not support once-off large enterprise data warehouse projects. The benefit of using this framework is that companies do not need large budgets to build a BI solution.

Companies in developing countries are normally not eager to spend money on BI solutions, because they are sometimes seen as a luxury and other information technology solutions compete with BI for funding. BI in developing countries, therefore, calls for a framework that can deliver smaller chunks of the BI solutions (data marts), which can deliver value earlier in the process to ensure continued funding. This framework is important for BI literacy and culture. BI literacy and culture need to grow with the solutions, otherwise the behaviour of people in the company can cause the BI implementation to fail.

CHAPTER 8

Evaluation of the DWDW lifecycle model

8.1 Introduction

Chapters 3, 4 and 5 discussed the development of the DWDW lifecycle model, for the implementation of a BI solution. Chapter 6 discussed BI maturity in the company. This chapter will discuss the interviews held with BI professionals and business people in the telecommunications company where the model was developed. The main purpose of the interviews with these people was to test the DWDW lifecycle processes that had already been discovered by the evaluation of case studies, and which had been improved upon by action research in the same company.

The interviews took place about ten years after the implementation of the DWDW lifecycle model. The disadvantage of doing these interviews after such a long time, is that the impact that the model has made is no longer being measured. The advantage is, however, that the model has been used for 10 years and, therefore, has been thoroughly tested.

8.2 Interview schedule

Oates (2008: 93) states that surveys can be done using several data generation methods, such as interviews, observations, questionnaires and documentation. This study used interviews as the data generation method. Kumar (2005:126-127) refers to questions used in an interview as the interview schedule.

The first interview was developed measuring the model against the research question and sub-questions (see section 1.7 of Chapter 1 for research questions). The second interview of this study investigated the success of the data warehouse implemented, whilst using the DWDW lifecycle model.

Questions 1 to 6 of the first interview schedule concentrate on the basic demographics of the BI team interviewed, and questions 7.1 to 7.12 in the same schedule concentrate on how the BI team has used the DWDW lifecycle model to implement the BI solution in the company (see Appendix A). The second interview schedule tested the perception of the super-users of the BI solution built using the DWDW lifecycle model (see Appendix B). ‘Super-user’ was the term used for business people that were regularly using the BI solution.

Both interview schedules include closed-ended and open-ended questions. Denscombe (2007:175-195) states that a structured interview consists of identical, pre-determined questions with different pre-coded answers for interviewees, while semi-structured interviews allow the interviewee to elaborate on the issue

introduced. Both interview schedules in this study used a combination of structured and semi-structured interviews.

8.3 Sampling methods

Babbie (2010:192-199) states that there are two kinds of sampling:

- Non-probability sampling
- Probability sampling

Oates (2008:96) states that non-probability sampling provides just a very weak result (high level estimate) with a general baseline to a much wider population, while probability sampling includes the following different sampling techniques:

- Random sampling

Random sampling is used where people or items are selected randomly. This sampling technique is normally used where a larger sampling frame exists, and a certain representative sample is used.

- Systematic sampling

Systematic sampling builds on random sampling and puts more structure to a bigger sampling frame by, for example, selecting every 100th person on an existing list.

- Stratified sampling

Stratified sampling is used where the sample group is divided up in the same ratio as the big sampling frame. For example, a retail company with 10,000 employees, where the stress level of employees has to be investigated - consists of 70% females and 30% male employees. With stratified sampling, only 50 employees are selected for the study, but this sample needs to have the same ratio of females and males (70%:30%)

- Cluster sampling

Cluster sampling is used where people naturally occur together, for example when a group that lives or worked together, is used for a study.

The sampling technique used for both interview groups in this study was cluster sampling. Cluster sampling was selected because there is only one BI team (supporting different business areas) in the company, and they were based in the same building. The BI super-users are also a very small group that are extensively using the BI solution. Statistics measuring the number of times the BI solution was used by a particular user, each month, revealed that the selected super-users were accessing the BI solution at an average of 82% of the total usage of the BI solution per day. This average was calculated over a two year period.

8.4 Compiling the interview schedules

Reinard (2007:342) warns that each question on an interview schedule should be:

- Brief and to the point
- Relevant to the interview schedule
- Unambiguous, by using words without double meaning
- Specific, by not asking vague questions
- Objective and should not lead the interviewee to an answer

All of the above points were considered when the interview schedules were compiled. More than 90% of the two respective interview schedules consisted of statements that require from the interviewee to state the degree he or she agreed or disagreed with the statement. The interviewee could select one of the following options:

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
-------------------	-------	--------	----------	----------------------

The 'unsure' option was included to ensure that people who do not have knowledge of a particular part of the DWDW lifecycle model, do in fact have the option to give a neutral opinion.

Each statement on the two interview schedules gave the interviewee the option to motivate his or her statement. In most cases, the interviewees that used the ‘unsure’ option, used the motivation space to indicate that they did not have experience with that particular BI area. Figure 8.4 represents the organogram of the BI team interviewed.

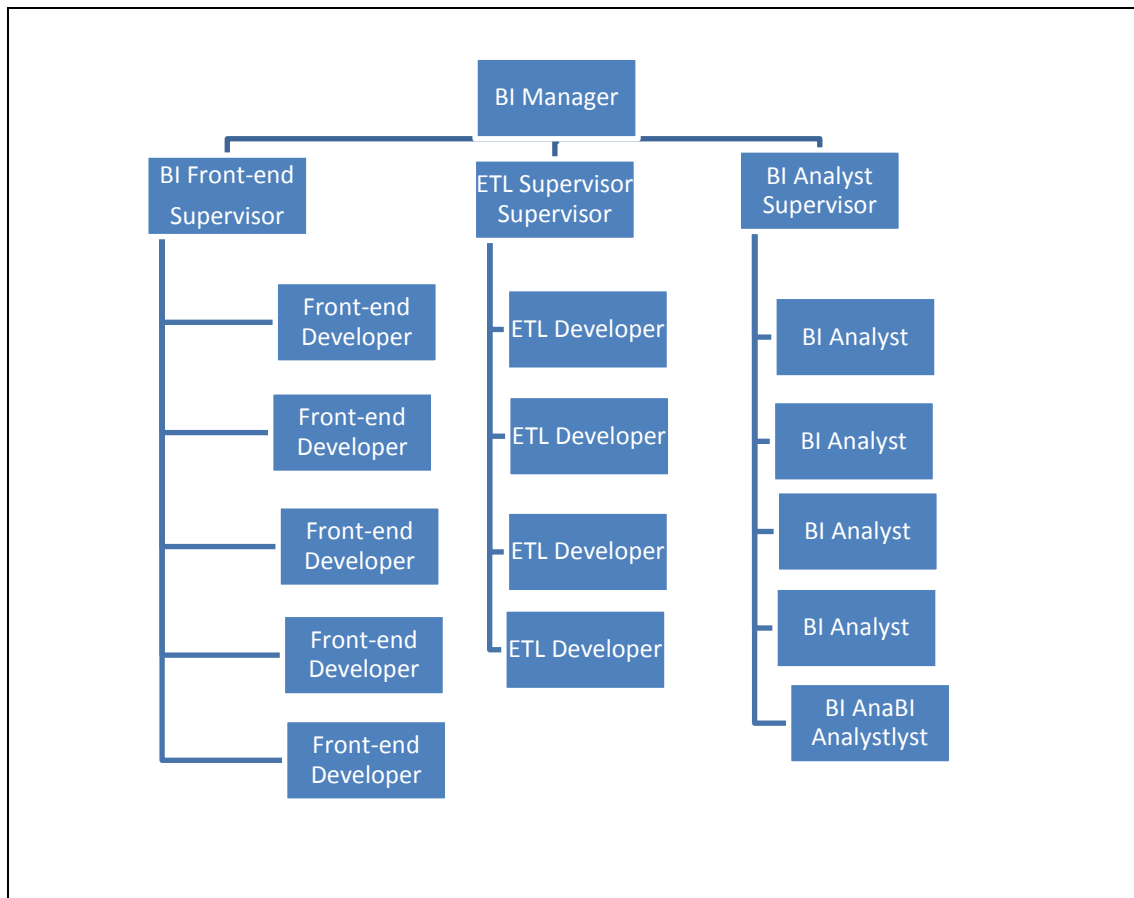


Figure 8.4 Organogram of the people interviewed.

Although the members interviewed physically worked in groups, the interviews were done individually to prevent bias - a situation where people's ideas are influenced by other people's ideas. Oates (2008:195) states that the disadvantages of group interviews are that people in higher positions may take over the talking in the group and that other people's ideas are not heard. Six super-users were also interviewed separately to test their perception of the BI solution from a business perspective.

8.5 Pilot interview schedules

Evans (2007:190) emphasises the importance of conducting a pilot study to test the interview schedule before the actual interviews are conducted. Two pilot interviews were done to ensure that the interview schedules contained the expected elements for this study. The pilot for the interviews with the BI team was done with the BI manager, and the pilot for the interviews with super-users was done with one of the super-users of the BI solution.

Kumar (2005:126-127) suggests that the interviewees used in the pilot should be informed that the interviews are only a test. Therefore, the two people interviewed were informed that their interviews were there to test the respective interview schedules.

8.6 Transcription of interview data

Transcription of audio recordings was not necessary because written notes were made of interviews. Breakwell (2006:242) states that when an audio recording of an interview is taken, it is important to transcribe the audio data, because it is much easier to analyse the relevant data once they are in a written format. Oates (2008:194) also suggests that the interviewee's words be captured as they are spoken and that the lines be numbered for easy reference back to the different statements. The different lines on the interview notes were thus numbered, and all notes were read to the interviewee at the end of the interview to eliminate possible misinterpretations. This was done because Oates (2008:194) states that - if possible - the interview notes or transcripts should be taken back to the interviewee to confirm the correctness of the facts.

8.7 Data analysis

Babbie (2010:51) states that when analysing qualitative research data, it is important to identify key themes, by segmenting the data. To segment the data one needs to categorise each segment first. The categories can be found by:

- Looking for existing theories in literature (also known as the deductive approach)
- Observing categories in the data (also known as the inductive approach)

- Ensuring that people in other roles share the same satisfactory rating
- Comparing different criteria used by different researchers

All of the above categorisation approaches were investigated, and a hybrid of the “inductive approach” and the “deductive approach” (Babbie, 2010:51) was used. The data for this study were categorised based on categories found in literature, as well as categories developed by the researcher, which is consistent with Oates (2008:296).

Categories developed by Bielkowicz *et al.* (2002:311-321) were used as a guideline to categorise the data collected - to analyse the data collected during the above-mentioned two sets of interviews. The results are reported on in the next two sections.

8.8 Interview with BI team

Table 8.8 summarises the qualification level of the BI team interviewed.

Qualification	Number
Senior certificate	11
1 year diploma	0
2 year diploma	0
3 diploma year diploma or degree	4

4 year degree or honours degree or higher	2
---	---

Table 8.8 Qualifications of the BI team interviewed.

8.8.1 Experience of BI professionals

A total of 13 interviewees (76.4 %) out of 17 BI professionals interviewed, have more than four years BI experience (see Table 8.8.1.1 below).

Years' Experience	Number of Employees
1-2 Years	0
2-3 Years	1
3-4 Years	3
4+ Years	13

Table 8.8.1.1 Years' experience in this company.

The group of BI professionals interviewed had an average of five years and more experience in their respective BI positions (see Table 8.8.1.2 below).

Experience	Average Years' Experience
ETL Developer	6
Front-end Developer	5
Business Analyst	11
Supervisor	6

Table 8.8.1.2 Distribution of BI experience.

From the group interviewed, only two BI professionals did not have experience of any other BI SDLC. Some 12 BI professionals had experience in both Kimball's BDLM and Inmon's CLDS (see table 8.8.1.3 below), and not one of the 12 interviewees had less than a year's experience in these two SDLCs.

Inmon's CLDS	12
Kimball's BDLM	15
No other experience	2

Table 8.8.1.3 Experience BI team had of other SDLCs.

8.8.2 Interview data analysis

Table 8.8.2 (below) represents the data used to evaluate the DWDW lifecycle model which was used to implement the BI solution in the telecommunications company.

Interview Questions	Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
<i>1. General test on the benefits of the DWDW lifecycle model</i>					
There is another BI SDLC that can deliver answers to new business questions in a shorter period than the DWDW lifecycle model.	0	0	0	5	12
DWDW architecture is saving the company money.	13	4	0	0	0
It is difficult to deliver projects on time when using the DWDW lifecycle model.	0	0	0	5	12
Wave 1 of the DWDW lifecycle model ensures a quick response to business questions.	17	0	0	0	0

The incremental built of the BI solution has a negative influence on the BI maturity growth.	0	0	2	5	10
<i>2. Test on processes prescribed by the DWDW lifecycle model on how to document business requirement specification</i>					
Using a BI tool like MS Excel to capture the requirements, minimises misunderstanding about the requirements.	16	1	0	0	0
It is easier to use the requirement documentation prescribed by Kimball's BDLM to do the requirements, when following the DWDW lifecycle model.	15	1	1	0	0

<i>3. Test on database design techniques prescribed by the DWDW lifecycle model</i>					
Wave 1 of the DWDW lifecycle model saves time, because very little time is spent on database design.	17	0	0	0	0
The flat file structure used in wave 1 of the DWDW lifecycle model, complicates the database design in wave 1.	0	0	0	6	11
Processes prescribed by wave 2 of the DWDW lifecycle model optimise the BI solution, which ensures good performance of the final solution.	14	2	1	0	0
<i>4. Test on the ETL processes prescribed by the DWDW lifecycle model</i>					
Conformed dimensions between the development environment and production environment	0	0	2	0	15

complicate the movement of data marts between the two environments.					
I think there is an advantage in the performance of the ETL, when using the OLTP servers for ETL processes.	11	5	0	1	0

Table 8.8.2 Evaluation of the DWDW lifecycle model, used by the BI team.

8.8.3 Benefits of the DWDW lifecycle model

Based on the interviews with the BI team, 15 out of 17 interviewees had knowledge of other BI SDLCs, and they considered that none of the other BI SDLCs can be used to attend to business questions faster than the DWDW lifecycle model. All 17 interviewees agreed that the DWDW lifecycle model helped the BI team to deliver projects much more quickly than the other known SDLCs.

In chapter 3 of this thesis, the DWDW lifecycle model prescribes an incremental build approach. Based on the interviews with the 17 members of the BI team, 15 interviewees (88%) believe that the incremental build approach benefited their BI implementation by improving the usage of the BI solution. They explained that

business people are using 100% of the BI solution, because it was built based on immediate business needs. The business could use the BI solution to manage a business threat or to simply implement a business opportunity.

8.8.4 DWDW lifecycle model business requirement specification

Statements grouped in section 2 of Table 8.8.2 tested the techniques prescribed by the DWDW lifecycle model to speed up the process of collecting business requirements and to minimise misunderstanding or misinterpretation. Some 100% of the interviewees agreed that Excel - because it is the reporting tool known to the users - was a good way of representing the requirements. The use of Excel as the template for capturing the requirements requires that Excel also be used as the reporting tool (BI Tool).

8.8.5 DWDW lifecycle model database design techniques

Wave one of the DWDW lifecycle model requires that the database is designed in a flat file structure to shorten the time it takes from the time the business requirement is received until the information is delivered to the business (see chapter 3, section 3.3.3).

The DWDW lifecycle model prescribes that the star schema design of the BI solution is only done in wave 2 of the model. The interviews done with the BI

professionals revealed that - as part of the optimisation of the BI solutions - the star schema design of the BI database (data warehouse) definitely belongs in wave 2 of the DWDW lifecycle model. The interviewees explained that the flat file structure prescribed by wave 1 of the DWDW lifecycle model saves database design time. Therefore, less design time is spent, and information is delivered much earlier to the business.

8.8.6 ETL processes of the DWDW lifecycle model

The DWDW lifecycle model recommends that the hardware of the OLTP systems be used to do the extraction and transformations of the ETL processes (see chapter 2, section 2.7.2). A total of 16 interviewees confirmed that there is an advantage in the performance of the ETL, when using the OLTP hardware.

For the movement of the database from the development environment to the production environment, the DWDW lifecycle Model requires that conformed dimensions be used. The results of the interviews show that there is an enormous benefit when using conformed dimensions for this purpose. Some 9 of the 16 interviewees favoured the use of conformed dimensions and also stated that they cannot imagine moving a data warehouse database between two environments without using conformed dimensions.

The one interviewee that marked “unsure”, is one of the BI analysts, and he had had no previous experience on the database side of a BI solution.

8.9 Analysis of BI team interview data

From the analysis of the interviews held with the above-mentioned BI team, it is confirmed that the DWDW lifecycle model has:

- Improved the time from business requirement to required information
- Reduced the cost involved with building the BI solution

Summary of interview with super-users:

The six BI super-users interviewed were working for the following departments:

- Finance
- Engineering
- Billing
- Credit Management
- Sales
- Admin

Table 8.9 below summarises the qualifications of the BI super-users interviewed.

Qualification	Number
Senior certificate	4
1 year diploma	0
2 year diploma	0
3 diploma year diploma or degree	1
4 year degree or honours degree or higher	1

Table 8.9 Qualifications of the super-users interviewed.

The interviews were done separately to ensure that one interviewee was not influenced by one of the others. All six of the users interviewed had two and more years' experience in the BI solutions built using the DWDW lifecycle model. The experiences of the different interviewees were as follows (see Table 8.9.1 below):

Department	Years
Finance	4
Engineering	3
Credit Management	5
Sales	3
Admin	2

Table 8.9.1 BI solution's utilisation in years.

Table 8.9.2 (below) summarises the usage experiences of the super-users interviewed. The results have been divided into two areas of discussion:

- Quality of information received from the BI solution
- Performance of the BI solution

Interview Questions	Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
<i>1. Quality of information received from the BI solution</i>					
The information received from the BI solution is sufficient to meet my daily needs.	2	4	0	0	0
The level of detail of the information received from the BI solution is sufficient for my daily needs.	5	1	0	0	0
Information received from the BI solution is always accurate and reliable.	2	4	0	0	0
<i>2. Performance of the BI solution</i>					
I am experiencing slow response with newly supplied information on the BI solution.	5	1	0	0	0
BI projects are taking too long to deliver the required information.	0	0	0	3	3
Latency of information on the BI solution is sufficient for the business that we are running.	6	0	0	0	0
I am prepared to wait longer for new information to appear on the BI solution, provided the performance is good once it is available.	0	0	0	1	5

Table 8.9.2 Evaluation of the performance of the BI solution built using the DWDW lifecycle model.

8.9.1 Quality of information received from the BI solution

Based on the six super-users, six agreed in the interviews held with them that the BI solution is providing them with enough information to do their daily work. Some two interviewees strongly agreed with this statement, while four interviewees just selected the “agree” option. From the analyses done on the reasons for the answers of the interviewees, it came to light that the interviewees selected the “agree” option because the statement is true for requirements already handed to the BI team, but that there are new BI requirements coming up regularly that need to be placed in the BI solution.

Five the interviewees that strongly agreed that they are happy with the level of detail of information in the BI solution explained that information in the BI solution was always available at the same level of detail to the level it was captured on the source systems. That means that it cannot be presented at a more detailed level, except when it is captured at a lower level in the source system. From the interviews, it was also realised that the information can be aggregated to any level required by the business.

All the interviewees agreed that that accuracy of information on the BI solution is good when compared with the data in the source systems. The interviewees stated that it should be remembered that the BI solution is dependent on the source systems for the accuracy of the data and information.

8.9.2 Performance of the BI solution

The interviewees agreed that there are slow responses when new information is captured on the BI solution. The interviewees stated this is acceptable for them because the BI team has explained to them the reason for this initial slow response, and they are happy that there is no long waiting period before new information is loaded on the BI solution.

The latency of the information on the BI solution is acceptable to the super-users because the BI team was sticking to the required latencies, but a new need is developing to get information with a latency of seconds (almost real-time).

8.9.3 Analysis of super-user interview data

From the analysis done with the super-users, it was revealed that they are happy with the quality of information received from the BI solution. The detailed level of information in the BI solution is acceptable for the super-users. However, there is a request that the BI solution is updated, with a latency of less than one minute for information required at this frequency by the business.

8.10 Conclusion

The interviews held with the above-mentioned two groups confirmed that when using the DWDW lifecycle model, money can be saved on hardware for BI, by using OLTP hardware to do the ETL processes. It was also confirmed by the interviews that the DWDW lifecycle model speeds up the implementation of a BI solution for an urgent business requirement by delivering the required information using wave one of the DWDW lifecycle model. The interviews tested the views of the BI team on the difference between the existing SDLCs and the new DWDW lifecycle model. According to these interviews, the main benefits of using the DWDW lifecycle model, were as follows:

- Using a BI tool that is already known to the BI end-users gives the BI end-users a demonstration of exactly how the final report would look.
- No time is spent on database design in wave one of the DWDW lifecycle model. That means that the BI solution reaches the BI user much quicker.
- The ETL process prescribed by the DWDW lifecycle model is optimised to use minimum hardware resources. Some of the processes are run on OLTP hardware during night-time. OLTP hardware is normally under-utilised during this time.

Chapter 9 will conclude this thesis, by looking at the research question and sub-questions and reflecting on the extent to which the research questions were answered.

CHAPTER 9

Conclusion

9.1 Introduction

This thesis concludes by referring back to the main research question: How can BI be implemented successfully in a South African telecommunications company, considering the high cost, limited time of implementing these kinds of solutions and the negative culture for BI that may be in existence?

When implementing a BI solution in any company, the implementation team has to deal with different challenges. The company used for this study has experienced several challenges. However, this study only covers the three main challenges experienced with the implementation of a BI solution. The first main challenge identified was the high cost of implementing a BI solution, the second was the low BI literacy rate of the company, and the third was the limited time to implement the BI solution.

The scope of this empirical study consisted of a historical case study, an action research cycle, as well as interviews which were all done in the same telecommunications company in South Africa.

The main research question was investigated by using five secondary research questions:

- 1) Where did data warehousing originate from?
- 2) Why is a data warehouse and BI solution needed?
- 3) How can a BI and data warehouse solution be implemented successfully, in a company with a limited budget for it?
- 4) How can a BI culture that supports the optimal utilisation of a BI solution, be instilled?
- 5) How can a complete enterprise data warehouse be implemented and grown following an incremental approach?

Questions 1 and 2 were answered by a literature study, which is presented in chapter 2 of this thesis.

Question 3 was investigated by using a case study research strategy (discussed in chapter 3) to develop a DWDW lifecycle model for the implementation of BI in the company. The DWDW lifecycle model was tested and improved in chapter 4, following an action research strategy while addressing a real life BI requirement in the company used for this study.

The DWDW lifecycle model includes administrative processes, like the documentation of user requirements and all technical documentation like database design and front-end design. It also includes technical processes, like building the database, ETL process and BI architecture.

After the company used for this study in the previous chapters had used the DWDW lifecycle model for 10 years, the researcher went back to do interviews with the BI technical team and the BI users. The results from this follow-up study were discussed in chapter 8.

The BI architecture historically used in a data warehouse was unsuitable for use with the DWDW lifecycle model. A new BI architecture was developed. Chapter 5 explains all the elements of this architecture and how it is used when implementing a BI solution.

Question 4 is answered in chapter 6. This chapter concentrates on the human factor when implementing a BI solution. Chapter 6 explains the concept of a BI culture, and a model is introduced that helps with creating a BI culture when implementing a BI solution. The literacy and culture maturity model was developed to manage and monitor BI literacy and culture.

Secondary question 5 is answered in chapter 7. The DWDW lifecycle model and the BI literacy and culture maturity model form a framework that was used to implement a BI solution in the above-mentioned company. This chapter explains the integration of the two above-mentioned models into a conceptual framework.

9.2 Summary of the thesis

The researcher was part of the project team implementing a BI solution with a telecommunication company. The implementation of a BI solution arose as a result of a business license requirement from the South African government. The telecommunications company used is a South African company that was previously a state department and had been privatised as part of a privatisation process of some state departments after the 1994 national elections in South Africa. The Department of Posts and Telecommunications was one of those departments. During this transition period, BI played a critical role in measuring the performance of the new telecommunications company against targets set by government. This new telecommunications company had to meet the targets set whilst also preparing for competition.

In order to manage the agreement with government, the new company was forced to start using available data on an existing network to identify under-utilised network elements. The data were available in disparate systems and had to be put into a central repository. A BI solution was needed to facilitate the correct business decisions and, therefore, the BI implementation team was looking for ways to hasten the process of implementing a BI solution. The license agreements with government could not wait several years for a BI solution to be completed. A solution was needed immediately. After several iterations of refinement of the SDLC, the DWDW lifecycle model was developed. This DWDW lifecycle model could not run on the ordinary BI architecture because of its different hardware and software requirements, as was explained in chapter 5.

After the information was made available in the BI solution, it was realised that business people did not know how to access and/or use this information. The lack of knowledge on how to use a BI solution resulted in people mis-interpreting the data. It was therefore concluded that BI literacy plays an important role with the implementation of a BI solution. The BI literacy rate needed to grow with the BI solution to ensure the optimal utilisation of the solution. The BI implementation team therefore started to build a model that determined the BI literacy level of the BI people and facilitated the literacy growth in the company. This BI literacy and culture maturity model is discussed in detail in chapter 6.

Chapter 7 concludes with a framework that includes both models developed during this study. The DWDW lifecycle model guided the implementation team on the technical processes of implementing a BI solution, while the literacy and culture maturity model guided the BI users through a process of growing BI literacy and culture in the company.

9.3 Future research

Unfortunately, all the challenges experienced with the implementation of a BI solution could not be resolved in one research project. More research is needed to address problems like dealing with volumes of data in the telecommunications industry. In the telecommunications industry every phone call made creates at least one record. If the specific company has 15 million customers, and each customer makes one call per day, 15 million records are generated each day. That

means that huge volumes of data are accumulating in telecommunication companies.

On the other hand, matters are complicated even more because common telecommunications applications, such as fraud detection systems, corporate base management systems and trade partner base management systems, require real-time and detail data feeds for optimal effectiveness. As a result, business intelligence solutions for the telecommunications industry must be able to handle tremendous extraction and insertion rates on databases, that are likely to grow into the multi-terabyte range. That means that the main challenge is the extraction from source system, transformation of data, and the loading into data warehouse in minute latencies (close to real-time).

High volumes and real-time extractions from a well-designed source database are difficult but possible. However, the dynamic nature of the telecommunications industry results in on-going changes to the database design and, therefore, these databases do not always reflect the best possible design. Unique keys in some of the instances do not exist and, therefore, it becomes very difficult to use database techniques such as database triggers. If one cannot use triggers to detect changes to existing records, full table comparisons to detect changes that appear in records must be used. Full table comparison is time- and resource-intensive on hardware and therefore does not support the requirement of real-time information. From the interviews done in the company used for this study (in 2011), it became clear that

more research is needed on how to implement real-time BI in a telecommunications company.

REFERENCES

ABELLO, J.M., PARDALOS, P.M., RESENDE, M.G.C. 2002. Handbook of massive data sets. Norwell: Kluwer Academic Publishers. 1223 p.

ADELMAN, S. & MOSS, L.T. 2000. Data warehouse project management. Roanoke, VA: Addison-Wesley. 404 p.

AGOSTA, L. 2000. The essential guide to data warehousing.[Online]. Available: <http://www.pearsonhighered.com/academic/product/0,3110,013085087X,00.html>. Date of access: 23 Nov. 2008.

ALLEN, S. & TERRY, E. 2005. Beginning relational data modeling. 2nd ed. Berkeley: Apress. 604 p.

ARMSTRONG, M. 2006. A handbook of management techniques: A comprehensive guide to achieving management excellence and improved decision making. London: Kogan Page Limited. 605 p.

ANDERSON, K. & KERR, C. 2002. Customer relationship management. New York, NY: McGraw Hill. 164 p.

ANON. 1996. White paper on telecommunications policy. [Online]. Available: http://www.polity.org.za/html/govdocs/white_papers/telewp.html?rebookmark=1. Date of access: 22 Oct. 2007.

ASPDEN, P. 2004. Patient safety: Achieving a new standard for care. Washington, DC: National Academy of Sciences. 528 p.

BABBIE, E.R. 2010. Basic research methods for librarians. Belmont: Wadsworth. 530 p.

BARAGOIN, C., MARINI, M., MORGAN, C., MUELLER, O., PERKINS, A. & YIM, K.H. 2001. IBM Redbooks: 123 Building the operational data store on DB2 UDB using IBM data replication, WebSphere MQ family, and DB2 warehouse manager. 3334 p. [Online]. Available: <http://www.redbooks.ibm.com/redbooks/pdfs/sg246513.pdf>. Date of access: 14 Dec. 2009.

BASKERVILLE, R.L. 1999. Investigating information systems with action research. *Communications of the AIS*, 2(19): 1-133.

BENNETT, S., SKELTON, J. & LUNN, K. 2005. Schaum's outlines of UML. Washington, DC: McGraw Hill. 400 p.

BIELKOWICZ, P., PATEL, P. & TUN, T. 2002. Object-oriented information systems: 8th International Conference, OOIS. Evaluating Information System Development Method: A new framework. London: Sage. 550 p.

BIFFL, S. & BOEHM, B. 2006. Value-based software engineering. New York, NY: Springer. 438 p.

BIJAOUI, P. & HASSLAUER, J. 2008. Designing storage for Microsoft. Exchange 2007 SP1. Oxford: Elsevier. 539 p.

BLUM, F.H. 1955. Action research: A scientific approach? *Philosophy of Science*, 22(1): 1-7.

BODLEY, J. 1994. An anthropological perspective. [Online]. Available: <http://www.wsu.edu:8001/vcwsu/commons/topics/culture/culture-definitions/bodley-text.html>. Date of access: 17 Dec. 2008.

BOEHM, B.W. & PAPACCIO, P.N. 1988. Understanding and controlling software cost. *IEEE Transactions on Software Engineering*, 14(10): 1462-1477.

BRACKETT, M.H. 1996. The data warehouse challenge: Taming data chaos. New York, NY: Wiley. 579 p.

BREAKWELL, B.R. 2006. Research methods in psychology. 3rd ed. London. Sage. 524 p.

BURLEW, L.D. 2006. Successful work adjustment: A life span approach. New York, NY: Nova Science Publishers, Inc. 241 p.

BURTON, B., GEISHECKER, L., SCHLEGEL, K., HOSTMANN, B., AUSTIN, T., HERSCHEL, G., SOEJARTO, A. & RAYNER, N. 2006. Business intelligence focus shifts from tactical to strategic. Gartner Research, ID Number: G00139352: 1-5.

CHEN, Z. 2002. Intelligent data warehousing: From data preparation to data mining. New York, NY: Wiley. 243 p.

CHIASSEON, M., GERMONPREL, M. & MALHIASSEN, L. 2009. Pluralist action research: A review of the information systems literature. *Information System Journal*, 19(1): 31-54.

COGHLAN, D. & BRANNICK, T. 2009. Doing action research in your own organization. 3rd ed. London: Sage. 170 p.

COHEN, L., MANION, L. & MORRISON, K. 2005. Research methods in education. 5th ed. New York, NY: Routledge Falmer. 523 p.

COLLIER, K.W. 2011. Agile Analytics: A value-driven approach to business intelligence and data warehousing. Boston, MA: Pearson Educations. 524 p.

COLOGON, R. & COHEN, D.R. 2008. FileMaker Pro 9 Bible. Indianapolis, IN: Wiley. 768 p.

COOPER, A. & REIMANN, R. 2003. About Face 2.0: The essentials of interaction design. Indianapolis, IN: Wiley. 540 p.

COURTNEY, J.F., HAYNES, J.D. & PARADICE, D.B. 2005. Inquiring organizations: Moving from knowledge management to wisdom. Hershey, PA: Idea Group Publishing. 373 p.

DAHINDEN, U., QUEROL, C., FAGER, F. & NILSSON, M. 2003. Citizen interaction with computer models. (In Kasemir, B. ed., Public participation in sustainability science: A handbook. New York, NY: Cambridge University Press. p. 105-125.)

DEMAREST, M. 2002. Two-tiered enterprise decision support systems architecture. [Online]. Available: <http://www.noumenal.com/marc/waremart.pdf>. Date of access: 14 Oct. 2009. 18 p.

DENNIS, A., WIXOM, B.H. & TEGARDEN, D.P. 2005. Systems analysis and design with UML version 2.0: An object-oriented approach. New York, NY: Wiley. 525 p.

DENSCOMBE, M. 2007. The good research guide: For small-scale social research projects. New York, NY: McGraw-Hill. 349 p.

DUMOULIN, T. 2007. ITIL implementation roadmap - part 2 [Online]. Available: http://blogs.pinkelephant.com/index.php?/troy/comments/itil_implementation_roadmap_part_2/. Date of access: 22 Jan. 2008.

DU PLESSIS, D. & MCDONALD, T. 2007a. Challenges in building and maturing of a telecommunications business intelligence solution in a developing country. (In Khosrow-Pour, M. (ed.), Managing worldwide operations and communications with information technology. Proceedings of 2007 Information Resources

Management Association International Conference, Vancouver, B.C., Canada, May 19-23. p. 217-221.)

DU PLESSIS, D. & MC DONALD, T. 2007b. The double wave data warehouse lifecycle model. Proceedings of the Eighth International Conference on Data Mining. Published in book format: Data mining VIII: Data, text and web mining and their business applications, edited by Zanasi, A., Brebbia C.A. & Ebecken, N.F.F.. New Forest: WIT Press. p. 105-114.

DU PLESSIS, D. & MCDONALD, T. 2007c. Strategic framework to implement a telecommunications business intelligence solution in a developing country. Proceedings of the Ninth International Conference on Enterprise Information Systems, Funchal, Portugal, June 12-16. p. 227-232.

DYCHE, J. 2000. eData turning data into information with data warehousing. 3rd ed. Heidelberg: Springer. 269 p.

ECKERSON, W. 2004. Gauge your data warehouse maturity. [Online]. Available: http://www.dmreview.com/article_sub.cfm?articleId=1012391. Date of access: 18 Aug. 2008.

EVANS, J. 2007. Your psychology project: the essential guide. San Francisco, CA: Sage. 271 p.

FRENTZOS, E. & THEODORIDIS, Y. 2007. On the effect of trajectory compression in spatiotemporal query. (In Ioannidis, Y., Novikov, B. & Rachev, B. Advances in databases and information systems: 11th East European Conference. Heidelberg: Springer. 217-233 p.)

GORDON, R.S. 2008. What's the alternative? Career options for librarians and info pros. New Jersey: Information Today, Inc. 272 p.

GOULDING, C. 2002. Grounded theory: A practical guide for management of business. London: Sage. 186 p.

GRAHAM, A. & COYLE, B. 2000. Framework for credit risk management. Chicago, IL: The Chartered Institute of Bankers. 140 p.

GRAY, D.A. 2004. Start and run a profitable consulting business. Vancouver: Counsel Press Ltd. 248 p.

GRAY, R.B. 1969. International security systems: Concepts and models of world order. New York, NY: Peacock Publishers. 227 p.

GREGORY, P. & ROTHSTEIN, P.J. 2007. IT disaster recovery planning for dummies. Indianapolis, IN: Wiley. 360 p.

HAISTEN, M. 1999. Real-time data warehouse: The next stage in data warehouse evolution, part 1. [Online]. Available:

http://www.dmreview.com/article_sub.cfm?articleId=946. Date of access: 23 Sept. 2008.

HAN, J. & KAMBER, K. 2006. Data mining: Concepts and techniques. 2nd ed. San Francisco, CA: Elsevier. 770 p.

HART, M. 2006. Progress of organisational data mining in South Africa. [Online]. Available: <http://www-direction.inria.fr/international/arima/006/pdf/arima00601.pdf> . Date of access: 21 June 2010.

HARTS, D. 2007. Microsoft Office 2007 Business intelligence: Reporting, analysis, and measurement from the desktop. Osborne: McGraw-Hill. 444 p.

HASSAN, R. 2008. The information society: Digital media and society series. Cambridge, MA: Polity Press. 266 p.

HOFFMAN, D.R. 2003. Effective database design for geoscience professionals. Tulsa, OK: Penn Well Corporation. 263 p.

HOPE, J. 2006. Reinventing the CFO. Boston, MA: Harvard Business Press. 258 p.

HOWSON, C. 2007. Successful business intelligence: Secrets to making BI a killer app. New York, NY: McGraw-Hill. 244 p.

HWANG, M.I. & XU, H. 2007. The effect of implementation factors on data warehousing success: An exploratory study. *Journal of Information, Technology and Organizations*, 41(3): 32-45.

IMHOFF, C., GALEMMO, N. & GEIGER, J.G. 2003. Mastering data warehouse design: Relational and dimensional techniques. New York, NY: Wiley. 438 p.

INMON, W.H. 1995. Data warehouse performance. New York, NY: Wiley. 444 p.

INMON, W.H. 2003. The DSS environment - data warehouse, data marts, and data mining: a glimpse at the past, a peek at the future [Online]. Available: <http://www.billinmon.com>. Date of access: Oct. 2007.

INMON, W.H. 2005. Building the data warehouse. 4th ed. New York, NY: Wiley. 543 p.

JONES, P. 2008. SAP business information warehouse reporting: Building better BI with SAP BI 7.0. Washington, DC: McGraw Hill. 256 p.

KELLY, S. 2007. Data warehousing in action. Indianapolis, IN: Wiley. 332 p.

KERREN, A., EBERT, A. & MEYER, J. 2007. Human-centered visualization environments: GI-Dagstuhl Research Seminar. Vaxjo: Springer. 403 p.

KIMBALL, R. 2006. The data warehouse lifecycle toolkit: With SQL Server 2005 and the Microsoft Business Intelligence Toolset. Indianapolis, IN: Wiley. 792 p.

KIMBALL, R. & CASERTA, J. 2004. The data warehouse ETL toolkit: Practical techniques for extracting, cleaning, conforming, and delivering data. New York, NY: Wiley. 491 p.

KIMBALL, R. & ROSS, M. 2002. The data warehouse toolkit: The complete guide to dimensional modelling. 2nd ed. New York, NY: Wiley. 436 p.

KUMAR, R. 2005. Research methodology: A step-by-step guide for beginners. London: Sage. 269 p.

LAVIN, M.R. 1992. Business information: How to find it, how to use it. 2nd ed. Phoenix, MD: Oryx Press. 499 p.

LIAUTAUD, B. & HAMMOND, M. 2001. E-business intelligence: Turning information into knowledge into profit. New York, NY: McGraw-Hill. 306 p.

LOVELAND, S., MILLER, G., SHANNON, M. & PREWITT, R. 2005. Software testing techniques: Finding the defects that matter. Hingham: Charles River Media, Inc. 362 p.

LOWNEY, C. 2003. Heroic leadership: Leadership principles for lasting success. Chicago: Loyola. 280 p.

MADSEN, M. 2010. Lowering the cost of business intelligence with open source. A comparison of open source and traditional vendor costs. [Online]. Available:

http://www.cognus.cl/media/users/2/100886/files/10961/lower_costs_with_osbi.pdf.

Date of access: 21 Sept 2010.

MALINOWSKI, E. & ZIMANYI, E. 2009. Advanced data warehouse design: From conventional to spatial to temporal applications. Berlin: Springer. 435 p.

MARTIN, R. 2008. Data warehouse 100 success secrets - 100 most asked questions on data warehouse design, projects, business intelligence, architecture, software and models. London: Lulu Com. 164 p.

MONTEITH, M. 2005. Teaching secondary school literacies with ICT. New York, NY: McGraw-Hill. 172 p.

MOSS, L.T. & ATRE, S. 2003. Business intelligence roadmap: The complete project lifecycle for decision-support applications. Boston, MA: Pearson Education Inc. 543 p.

NEELY, T.Y. 2006. Information literacy assessment: Standards-based tools and assignments. New York, NY: ALA Editions. 216 p.

OATES, B.J. 2008. Researching information systems and computing. London: Sage. 341 p.

OLSON, E. 2003. Data quality: The accuracy dimension. San Francisco, CA: Elsevier. 293 p.

OPPEL, A.J. 2009. Databases: A beginner's guide. Washington, DC: McGraw-Hill. 496 p.

OUYANG, H. & WANG, J. 2008. Data warehouse software. (*In* Information technology curriculum integration encyclopedia, 1: 179-184.)

PAREEK, D. 2006. Business intelligence for telecommunications. Baco Raton: Taylor & Francis Group LLC. 294 p.

PEKOWSKY, L. 2004. JavaServer pages. Boston, MA: Pearson Education. 347 p.

PFLEEGER., S.L. & ATLEE, J.M. 2009. Software engineering: Theory and practice. 4th ed. New Jersey: Pearson Higher Education. 765p.

PISELLO, T. 2005. The value of business intelligence. [Online]. Available: http://searchcio.techtarget.com/tip/0,289483,sid182_gci1129774,00.html. Date of access: 8 Mar. 2007.

PONNIAH, P. 2001. Data warehouse fundamentals. New York, NY: Wiley. 516 p.

QUINN, K.R. 2003. Establishing a culture of measurement: A practical guide to BI. New York, NY: Information Builders, Inc. 20 p.

RAINARDI, V. 2008. Building a data warehouse: With examples in SQL server. New York, NY: Springer. 523 p.

RAMESH, G. & BHATTIPROLU, R. 2006. Software maintenance: Effective practices for geographically distributed environments. New Delhi: McGraw-Hill. 465 p.

RANKINS, R., JENSEN, P. & BERTUCCI, P. 2003. Microsoft SQL server 2000 unleashed 2003. New York, NY: Sams Publishing. 342 p.

RASMUSSEN, N., GOLDY, S. & SOLLI, P.O. 2002. Financial business intelligence: Trends, technology, software selection, and implementation. New York, NY: Wiley. 283 p.

RAVI, V. 2008. Advances in banking technology and management: Impacts of ICT and CRM. London: IGI Global. 362 p.

REINARD, C.R. 2007. Introduction to communication research. London: McGraw-Hill. 588 p.

SALEHI, M. & ABEDINI, B. 2009. Financial distress prediction in emerging markets: Empirical evidences from Iran. *Business Intelligence Journal*, 2(2): 398-409.

SALISBURY, D.F. 1996. Five technologies for educational change: Systems thinking, systems design, quality science, change management, instructional technology. Cape Town: Educational Technology. 247 p.

SCHACH, S. 2005. Object-oriented and classical software engineering. 6th ed. New York, NY: McCraw-Hill. 581p.

SCHEIN, E.H. 2009. The corporate culture survival guide. San Francisco, CA: Wiley. 256 p.

SCHEPS, S. 2008. Business intelligence for dummies. Indianapolis, IN: John Wiley. 384 p.

SCHRAGE, M. 2000. Serious play: How the world's best companies simulate to innovate. Boston, MA: Harvard Business Press. 244 p.

SEARS, A. & JACKO, A. 2009. Human-computer interaction: Development process. Manchester: CRC Press. 356 p.

SILVERS, F. 2008. Building and maintaining a data warehouse. Boca Raton: CRC Press. 304 p.

SIMSION, G.C. & WITT, G.C. 2005. Data modeling essentials. 3rd ed. San Francisco, CA: Elsevier. 532 p.

SIU, B. 1997. Data mining, data warehousing & client/server databases. Proceedings of the 8th International Database Workshop, Hong Kong, 29-31 July. New York, NY: Springer. 303 p.

SONG, I. & LEVAN-SHULTZ, K. 1999. Data warehouse design for e-commerce environments. New York, NY: Springer. p. 374-388.

SOUTH AFRICAN INSTITUTE OF RACE RELATIONS. 2000. Survey. Pretoria: Institute Publisher. 700p.

SPENIK, M. & SLEDGE, O. 2003. Microsoft SQL Server 2000 DBA survival guide. 2nd ed. New York, NY: Sams Publishing. 847 p.

STACKOWLIAK, R., RAYMAN, J. & GREENWARD, R. 2007. Oracle data warehousing and business intelligence solutions. Indianapolis, IN: Wiley. 386 p.

SUMATHI, S. & SIVANANDAM, S. 2006. Introduction to data mining and its applications. Heidelberg: Springer. 828 p.

TYNER, K.R. 1998. Literacy in a digital world: Teaching and learning in the age of information. New Jersey: Lawrence Erlbaum Associates Inc. 291 p.

UNHELKAR, B. 2008. Practical object oriented analysis. Tampa, FL: Thomson. 236 p.

UTLEY, C. 2008. Business intelligence with Microsoft Office Performance Point Server 2007. Osborne: McGraw-Hill. 358 p.

VIVIERS, W., SAAYMAN, A. & MULLER, M. 2005. Enhancing a competitive intelligence culture in South Africa. [Online]. Available: <http://www.emeraldinsight.com/researchregister>. Date of access: 17 June 2007.

WALLY, B. 2003. The digital age story teller. [Online]. Available: <http://www.bockinfo.com/docs/bifaq.htm>. Date of access: 25 Jul. 2005.

WESTERMAN, P. 2001. Data warehousing: Using the Wal-Mart model. San Diego, CA: Academic Press. 297 p.

WHITE, C. 2007. Survey of BI professionals indicates fewer than 40 percent have BI deployed. [Online]. Available: <http://www.ebizq.net/news/8224.html>. Date of access: 13 Apr. 2007.

WIGHTMAN, J. 2008. Pro SQL Server 2005 Integration Services. Berkeley: Apress. 548 p.

WILLIAMS, S. & WILLIAMS, N. 2004. The business value of business intelligence. *Business Intelligence Journal*, 1(2): 205-223.

WILLIAMS, S. & WILLIAMS, N. 2007. The profit impact of business intelligence. San Francisco, CA: Elsevier. 218 p.

XWAVE. 2004. The value of business intelligence. [Online]. Available: http://www.xwave.com/industries/telecom/solutions/busintell/bi_value.asp. Date of access: 12 Oct. 2009.

YIN, R.K. 2003. Case study research: Design and methods. 3rd ed. Thousand Oaks, CA: Sage. 439 p.

APPENDIX A

Background

The DWDW Lifecycle Model has been developed and implemented in your company.

Purpose interview

The purpose of this interview is to evaluate the DWDW Lifecycle Model in terms of the influence it may or may not have on implementing a Business Intelligence (BI) solution in a company with a limited BI budget, and the influence it may or may not have on BI literacy in that company. The output of this questionnaire will be used to highlight strong and/or weak points of this model.

Which one of the following is applicable to you?

1. Gender

Male	
Female	

2. Educational level

Senior Certificate	
1 year diploma	
2 year diploma	
3 diploma year diploma or degree	
4 year degree or Honours degree or higher	

3. How many years' experience do you have in BI?

1-2 Years	2-3 Years	3-4 Years	4+ Years

4. Please select from the list your number of years' experience in the different BI positions.

BI Position	No experience	1-2 Years	2-3 Years	3-4 Years	4+ Years
ETL Developer					
Data Base Administrator					
Front-end Developer					
Business Analyst					
BI supervisor					

5. How long have you been working for this company?

Less than 1 Year	1-2 Years	2-3 Years	3-4 Years	4+ Years

6. Which other BI SDLC have you used before?

Inmon's CLDS	
Kimball's BDLM	
Other	

If other please name the SDLC:.....

7. Please name the SDLC that you are currently using in your company.

.....

8. If you have used the SDLCs in the table below, please mark the number of years you have been using it.

	1-2 Years	2-3 Years
Inmon's CLDS		
Kimball's BDLM		
DWDW Lifecycle Model		

9.To what degree do you agree or disagree with the statements below?

9.1 There is another BI SDLC that can deliver answers to new business questions in a shorter period than the DWDW Lifecycle Model.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
----------------	-------	--------	----------	-------------------

Please motivate your answer

.....
.....

9.2 DWDW architecture is saving the company money.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
----------------	-------	--------	----------	-------------------

Please motivate your answer

.....
.....

9.3 It is difficult to deliver projects on time when using the DWDW Lifecycle Model.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
----------------	-------	--------	----------	-------------------

Please motivate your answer

.....
.....

9.4 Wave 1 of the DWDW Lifecycle Model ensures a quick response to business questions.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
----------------	-------	--------	----------	-------------------

Please motivate your answer

.....

9.5 The incremental built of the BI solution has a negative influence on the BI maturity growth.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
----------------	-------	--------	----------	-------------------

Please motivate your answer

.....

9.6 Using a BI tool like MS Excel to capture the requirements, minimises misunderstanding about the requirements.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
----------------	-------	--------	----------	-------------------

Please motivate your

answer.....

9.7 It is easier to use the requirement documentation prescribed by Kimball's BDLM to do the requirements, when following the DWDW Lifecycle Model.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
----------------	-------	--------	----------	-------------------

Please motivate your answer

.....

9.8 Wave 1 of the DWDW Lifecycle Model saves time, because very little

time is spent on database design.				
Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
Please motivate your answer				
9.9 The flat file structure used in wave 1 of the DWDW Lifecycle Model, complicates the database design in wave 1.				
Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
Please motivate your answer				
9.10 Wave 1 of the DWDW Lifecycle Model saves time because very little time is spent on database design.				
Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
Please motivate your answer				
9.11 Processes prescribed by wave 2 of the DWDW Lifecycle Model optimise the BI solution, which ensures good performance of the final solution.				
Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree

Please motivate your answer.....				
9.12 Conformed dimensions between the development environment and production environment, complicate the movement of data marts between the two environments when using DWDW architecture.				
Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
Please motivate your answer.....				
9.13 I think there is an advantage in the performance of the ETL, when using the OLTP servers for ETL processes.				
Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
Please motivate your answer.....				
...				

APPENDIX B

Background

The DWDW Lifecycle Model was used to develop the BI Solution in your company.

Purpose interview

The purpose of this interview is to evaluate the BI Solution implemented in your company using the DWDW lifecycle model.

Which one of the following is applicable to you?

1. Gender

Male	
Female	

2. Educational level

Senior Certificate	
1 year diploma	
2 year diploma	
3 diploma year diploma or degree	
4 year degree or Honours degree or higher	

3. Which one of the following departments are you from:

Finance	Engineering	Billing	Credit Management	Sales	Admin	Other
---------	-------------	---------	----------------------	-------	-------	-------

Please specify other.....

4. How long have you used the BI Solution?

1 Year	2 Years	3 Years	4 Years	5 Years	More Than 5 Years
--------	---------	---------	---------	---------	-------------------

To what degree do you agree or disagree with the statements below?

5. The information received from the BI solution is enough to meet my daily needs.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
-------------------	-------	--------	----------	----------------------

Please motivate your answer.....

6. The level of detail of the information received from the BI solution, is sufficient for my daily needs.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
-------------------	-------	--------	----------	----------------------

Please motivate your answer

7. Information received from the BI solution is always accurate and reliable.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
-------------------	-------	--------	----------	----------------------

Please motivate your answer

1. I am experiencing slow response with newly supplied information on the BI Solution.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
-------------------	-------	--------	----------	----------------------

Please motivate your answer

9. BI projects are taking too long to deliver the required information

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
-------------------	-------	--------	----------	----------------------

Please motivate your answer

10. Latency of information on the BI Solution is sufficient for the business that we are running.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
-------------------	-------	--------	----------	----------------------

Please motivate your answer

11. I am prepared to wait longer for new information to appear on the BI solution, provided the performance is good once it is available.

Strongly Agree	Agree	Unsure	Disagree	Strongly Disagree
----------------	-------	--------	----------	-------------------

Please motivate your answer