ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

DEPARTMENT OF INFORMATION SCIENCE

A FRAMEWORK DEVELOPMENT FOR APPLICATION OF DATA WAREHOUSE IN THE ETHIOPIAN BANKING INDUSTRY

THE CASE OF DASHEN BANK

BY: ZELALEM TESFAYE

JAN, 2022

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

DEPARTMENT OF INFORMATION SCIENCE

A FRAMEWORK DEVELOPMENT FOR APPLICATION OF DATA WAREHOUSE IN THE ETHIOPIAN BANKING INDUSTRY THE CASE OF DASHEN BANK

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information systems.

BY

ZELALEM TESFAYE

JAN, 2022

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

DEPARTMENT OF INFORMATION SCIENCE

A FRAMEWORK DEVELOPMENT FOR APPLICATION OF DATA WAREHOUSE IN THE ETHIOPIAN BANKING INDUSTRY THE CASE OF DASHEN BANK

BY

ZELALEM TESFAYE

Name and signature of Members of the Examining Board

Name	**Title**	**Signature**	**Date**
_____	Chairperson	_____	_____
_____	Advisor(s),	_____	_____
_____	Internal Examiner	_____	_____
_____	External Examiner,	_____	_____

JAN, 2022

ACKNOWLEDGEMENTS

First and foremost, praises and thanks to the God, the Almighty, for His Help and blessings throughout my research work to complete the research successfully.

I would like to express my sincere gratitude to My colleagues at Dashen Bank for their unreserved support as well as participating in the group discussion without hesitation, I would like to thank my family specially my brother Anteneh Tesfaye, for giving me continues moral support throughout my study.

My sincere thanks also goes to my advisor Dr. Wondwossen Mulugeta for his continuous support, for his patience and guidance.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ATM	Automated Teller Machine
CMIS	Customer Management information system
CRM	Customer relationship management
DWH	Data warehouse
DSS	Decision support system
ETL	Extraction, Transformation and Loading
NBE	National Bank of Ethiopia
OCRM	Operational Customer relationship management
OLAP	Online Analytic processing
OLTP	Online transaction processing
POS	Point of sell

ABSTRACT

Data warehouse is the fastest growing technology integrated with Business intelligence to make a better decision using analytics on the current as well as historical data of the organization. The purpose of this research is to develop a framework for the application of Data warehousing technology in the Ethiopian banking industry, the case of Dashen bank. The research followed the Kimball dimensional data modeling technique to show the conceptual and logical data model. For this particular Applied research, a qualitative research methods are believed to be more suitable in describing the process of developing the data warehouse and Data collection is mainly based on Focus Group discussion between the staff members of the IT department, Observation and document assessment to evaluate the current status of data integration and readiness of the bank for the application data warehouse. With this the readiness of the bank is evaluated, the data concentration area in the bank's current system is also identified, source databases and the type of data with these systems is also identified and these findings analyzed and represented using the UML class diagram on the conceptual level.

Based on the conceptual level, conformed dimensional tables and Fact tables with their attributes and measures are identified. The logical design section of the study shows the relationship of facts and dimensions using star schema of Kimball dimensional modeling. The finding of this research tried to shows the benefits of the application of data warehousing technology in the Ethiopian banking industry using the conceptual and logical design of the data warehouse in the case study organization.

CHAPTER ONE

INTRODUCTION

1.1. BACKGROUND

Internet and associated information technologies are said to be behind an information revolution that is transforming the way people live and work (James D. Torr 2003), Since the introduction of internet and the web the amount of data moving around the web and stored in different storage devises is increased exponentially, that the conventional databases are not able to store, process and analyze to produces information that can support the decision making process in an organization. The data warehouse exists to facilitate decision support in the organization. (Breslin 2004).

Organizations become interested in data warehousing when William H. Inmon (Bill Inmon) introduces the idea of data warehouse and elaborate the opportunity that this huge amount of date, brings to their advantage in the 1990's. Bill Inmon introduces the idea of data warehousing in his seminal work of building data warehouse in 1990 (Breslin 2004).

Data warehouse has been defined by different scholars, of these the well-known definition of Data warehouse is Bill Inmon's which states that "data warehouse is a subject-oriented, integrated, stable, time-varying data sets to support management decision-making process." (Chen Wenzhe 2015) but Jacek Maślankowski argued that this definition is not about the data warehouse system but it concerns the characteristics of the data in the warehouse not the data warehouse system and accepts rather Mattison's definition of data warehouse. Data warehouse, whose primary purpose is integrating information from different legacy systems, and make it usable to business people, in the support of their efforts to reduce cost and improve revenue" (Maślankowski2013).

Arvind Kumar tries to compromise by combining the ideas of both Inmon and Mattison's definitions and come up with an alternative definition as, data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and that usually resides at a single site and is subject-oriented, integrated, nonvolatile, and time-variant. (Kumar 2016)

According to Maślankowski data warehouse can be defined in three ways, concerning data based on Inmon's definition, concerning objectives based on Mattison definition and based on its functionality.

- Concerning data, usually based on Inmon's definition

  "Data warehouse is a subject-oriented, integrated, stable (Non-volatile), time-varying data sets to support management decision-making process."

- Concerning the objectives of the system, based on Mattison definition

  "Data warehouse is a system whose primary purpose is to extract information out of legacy systems, and make it usable to business people, in the support of their efforts to reduce cost and improve revenue" (Mattison, Cited by Maślankowski)

- Based on its functionality.

A data warehouse is an information system that provides an integrated and total view of the enterprise, make the enterprise's current and historical information easily available for strategic decision making, makes decision support transaction possible without hindering operational systems, renders the organization's information consistent, presents a flexible and interactive source of strategic information" [Ponniah Cited by Maślankowski].

The most accepted and flexible definition is based on W. H. Inmon According to Bill Inmon Data warehouse has four basic characteristics which are Subject-oriented, Integrated, Non-volatile, and Time-variant.

Subject Oriented: A data warehouse is subject oriented because it provides information around a subject rather than the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue and online transaction records.

Integrated: A data warehouse is constructed by integrating data from heterogeneous sources such as relational databases, flat files, etc. This integration enhances the effective analysis of data. Moreover, it keeps consistent naming conventions, format, and coding.

Time Variant: The data collected in a data warehouse is recognized with a particular period and offers information from the historical point of view.

Non-volatile: Data warehouse is also non-volatile means the previous data is not erased when new data is entered, once data is inserted in the warehouse, it can't be updated or changed it is a read-only data which is refreshed periodically.

1.2. BANKING HISTORY IN ETHIOPIA

According to NBE, based on the agreement between Emperor Minilik and Mr. Magilivray, representative of the British owned National Bank of Egypt in 1905; the first modern banking system is introduced in Ethiopia by the name Bank of Abyssinia with a capital of 500,000-pound sterling, on Feb 16, 1906. Nwanne in his study of Ethiopian banking system discussed the Ethiopian banking history started when the first private Bank of Abyssinia, whose shares were sold in Addis Ababa, New York, Paris, London and Vienna, inaugurated by Emperor Minilik II on February 15, 1906. Until it is liquidated by Emperor Haile Selassie, who is not interested to give the issuing of Bank Notes to a foreign owned Bank, in 1931with a new Banking reform. (T. F. I. Nwanne2015) bank of Abyssinia financed many projects including the famous Franco – Ethiopian Railway, which reached Addis Ababa in 1917.

In 1931 the Ethiopian government purchased the Abyssinian Bank, which was the dominant bank, and renamed it the Bank of Ethiopia the first nationally owned bank on the African continent (Gedey 1990; Degefe 1995; Sited by Alemayehu Geda 2006). on August 29, 1931 Bank of Abyssinia was legally transformed to Bank of Ethiopia by Emperor Haile Selassie and become the first state owned indigenous bank in Africa with a capital of £750,000. After the Smooth Transition Bank of Ethiopia operated until the Italian invasion with five branches around the country.

During the five-years of Italian occupation (1936-41) banking activity expanded. The Italian banks were particularly active. (Alemayehu Geda 2006). During the Italian invasion the Italians established and start to operate branches of their main banks called Banca d'Italia, Banco di Roma, Banco di Napoli and Banca Nazionale Del Lavoro. Until it is withdrawn in 1943 another foreign bank called Barclays Bank come to Ethiopia with British troops in 1941.Barclays Bank was established and it remained in business in Ethiopia between 1941 and 1943 (Gedey 1990; Degefe 1995; Sited by Alemayehu Geda 2006). Following the withdrawal of Barclays Bank in 1943 the Ethiopian government established the State Bank of Ethiopia. Bank of Ethiopia was operating as both a commercial and a central bank until 1963 when it was remodeled into today'

s National Bank of Ethiopia (the Central Bank, re-established in 1976) and the Commercial Bank of Ethiopia (CBE). After this many other private banks were established; were in operation until the 1974 revolution. (Alemayehu Geda 2006)

Following the 1974 revolution the government takes control of the whole economy and nationalized all large corporations and merged the three private banks Addis Ababa Bank, Banco di Roma and Banco di Napoli in 1976 to form the second largest Bank called Addis Bank. The two largest banks Addis Bank and Commercial bank of Ethiopia are also merged in 1980 to form the sole commercial bank in the country until the establishment of private commercial banks in 1994 when the communist administration is removed.

1.2.1. ABOUT DASHEN BANK S.C

Dashen Bank is one of the biggest private banks established in September 1995, when free market economy is re-introduced, with a vision of being the highest in the banking services and start operation in 1996 with eleven branches. Currently the Bank operates through a network of more than 400+ Branches, ten Forex Bureaus, 400+ ATMs and 2000+ Point-of-Sale (POS) terminals stretched across the country. The bank also established correspondent banking relationship with 462 banks covering 70 countries and 170 cities across the world.

Dashen Bank also works in partnership with international leading brands in the electronic payments industry like AMEX, VISA, MasterCard & China Union-Pay and prominent money transfer operators like Western Union, Money-Gram, Express Money & Dahabshiil, Ezremit, Trans-fast, World-Remit and Ria. With the aim of creating cashless society Dashen bank also introduced “Amole” Digital payment platform that offers subscribers digital payment capacity and access to aggregated digital product and service from Retailers, Entertainment Industries, Airtime Dealers, Bill Payment Points, Airlines, Social Media Players and Third-Party Service providers.

The bank structure was modified many times through time to accommodate the ever changing need of the financial and technological development. Giving more emphasis to the development of digital banking, currently the bank structure is organized in a way that the implementation of ICT to deliver the banks products and services mate the customer’s needs. There are four departments under the supervision of the chief of information officer which are led by directors

each (Application development and support department, system security department, infrastructure Department, IT service delivery and IT projects and standards).

1.3. RESEARCH BACKGROUND AND STATEMENT OF THE PROBLEM

With the introduction of advanced core Banking system (CBS), multi-channel online transaction processing system, ATMs, agent Banking, Internet banking and Mobile Banking services in the Ethiopian banking industry a huge amount of data is being collected to the banks data storage. Although this large amount of data holds much information which benefits the banks when analyzed, there is no publicly available information which shows the banks are using their operational as well as historical data analysis to study their customer's behavior, operational performance and market predictions, except some banks are using their operational data from their core banking system for reporting purpose

Chen Wenzhe from the Management School of Ocean University of China presented a study under a title The Application of Data Warehouse Technology in Modern Finance. Chen says the main objective the study is to establish a bank CRM data warehouse system. The overall objective is to establish banking enterprise database systems, integrate customer information resources to achieve comprehensive customer information management functions. (Chen 2015).

The study discussed the advantage of the implementation of a CRM data warehousing technology for the banking industry as a big leap in the development of the industry which makes a big difference in terms of customer handling, and states that the system of CRM data warehouse management will implement features including customer relationship management, customer behavior, customer groups and market activities analysis and management.

The general Model of Chen's design passes through five different stages making the Banks CRM the main focus point of the study to develop his own model of the DWH as shown below. Although knowing the customer's behavior is a necessity and is very important in the current Banking business to attain the competitive advantage, tracking the internal transactional data, which shows the day to day performance of the bank of different products provided by the bank like deposit, lone and other data are equally important in the decision making process for the management.

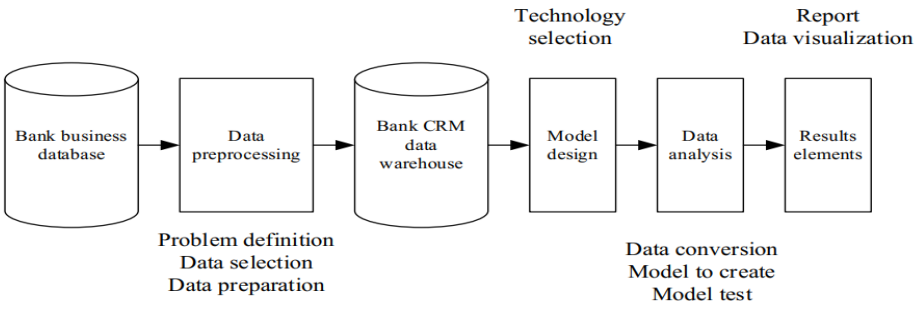
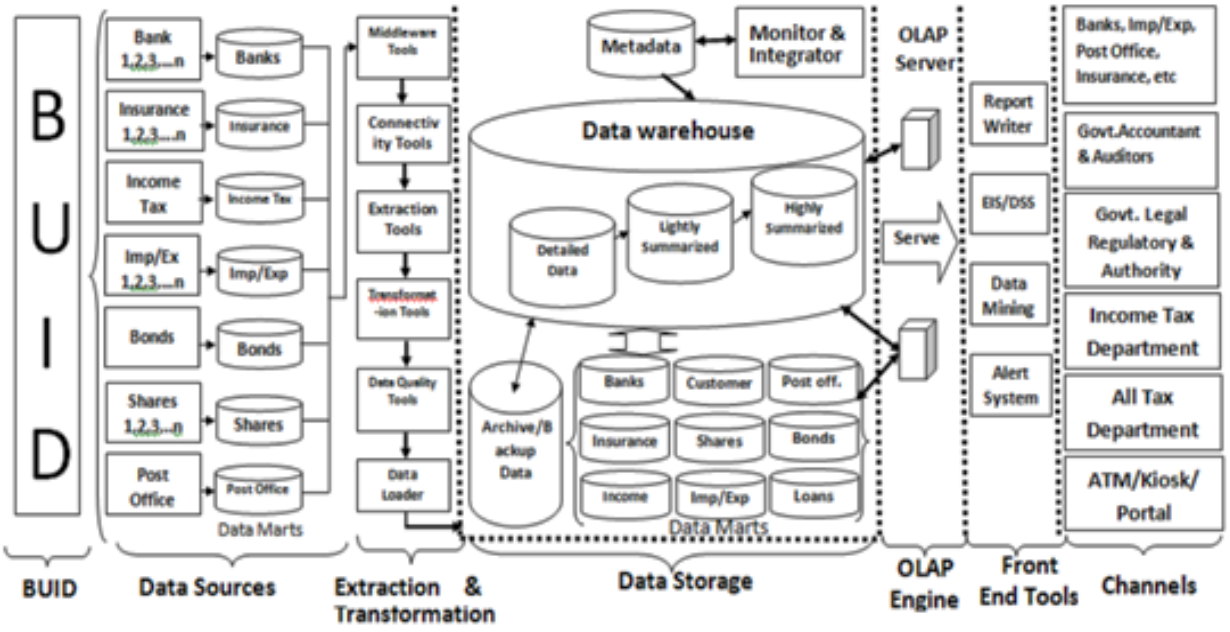


Figure 1. Chen's design of CRM data warehouse diagram

As we can see from Chen's CRM data warehouse diagram, the design aimed to enhance the efficiency of the bank to take advantage of the customer information stored in the bank's databases and better understanding their customer to become best competitor in the current banking industry. Therefore, Chen's design is customer centric and not transaction centric, which gives basic information on the day to day change on the banks deposit, lone and physical cash availability. Moreover, the study ignores the fact that in addition to the detailed customer behavioral information high level management and decision makers' CEOs and CIOs needs aggregate data, which shows the daily performance of the Bank's, on deposit, lone, physical cash availability and foreign currency availability for decision making purpose.

The other study which focuses on the application of data warehousing technologies for financial institutions (Vaibhav et al. 2014) under the title Data Warehouse Architecture for Financial Institutes to Become Robust Integrated Core Financial System using BUID, the study mainly focuses on the gap of inability of the current core banking system to monitor transactions of customers with multiple account number from different branches of a Bank, and introduces a seven layer data warehouse architecture which is designed to connect all financial institutes in a network to store customer's data in a Data Warehouse through data mart using a Bank universal Unique identification(BUID) so that the transaction made by the customer can be monitored easily all over the bank's system and government authorities.

Figure 2. Data Warehouse Architecture for Financial Institutes by (Vaibhav et al. 2014)



The current core banking system of the Bank doesn't provide the bank universal unique identification code and cannot identify the customer's transactions of different accounts at one place or under one roof (Vaibhav et al. 2014) The study suggested a new architecture design of seven vertical layers named as BUID, Data Sources, Extraction and Transformation, Data Storage, OLAP Engine, Front End Tools.

The study has a much wider aim of not only introducing the application of data warehouse for commercial bank but also connecting the financial institutions with other financial institutions and governmental authorities to track the transaction information of customers by introducing Bank unique universal identification number (BUID). The study elaborates how the current gap on the core banking system obstructs the different governmental authorities form tracking the transaction information. Such transactions are not sense by other financial system or other financial institutes and therefore government authorities like Income Tax department, Financial Industry Regulatory Authority, Financial Services Authority, etc… cannot collect customer's transactions information from one core financial system. (Vaibhav et al. 2014)

Although it is a valuable study which not only monitor the transactions flow of customers but also to integrate financial institutions with government authorities, however the current core banking system of Dashen bank do not enforce the implementation of the use of a Bank unique universal

Identification number and the readiness of the other financial institution and governmental authorities to implement such a wide scale data warehouse architecture in our country financial system is not assessed. Moreover, the main focus of the study is more on how to track the transactions of customers with multiple accounts both by the bank and external regulatory authorities using BUID rather than preparing the data collected like deposit, lone, physical cash availability and foreign currency availability by the different systems in the data warehouse.

Although Dashen bank has integrating few of the operational database to satisfy the increasing report need of the Branches and departments the bank is not able to produce analysis to measure the performance of products and Merchants in terms of transactions performed. Since the ATM transaction data collected from the terminals has never been analyzed, the ATM terminals, which are purchased with high foreign currency and installed at the various locations in the city, have never been checked for their location profit performance, Hence the Bank is not getting the expected benefit from the terminals performance since some of the ATMs are under performing.

The different products (Cards) issued and acquired by the bank are used on the point of sell (POS) terminals distributed to its merchants. The product brand is purchased, the physical card is purchased, the card printing machine is purchased with foreign currency and there is also a yearly charge for the brand. Since the transaction data of all products issued and acquired has never been processed and analyzed, the charge collected from each product has never been checked, the bank is not be able to identify which product is performing well and which is not, hence the bank is subject to pay and hold on the underperforming products each year.

The customers of the Bank who are registered as a merchant and received the POS machine from the bank are expected to avail the terminals to customers and use them at all times. Since the merchant's transaction performance on the banks POS terminal is not checked, processed and analyzed, the Banks is not able to measure the underperforming and well performing merchants with regard to the usage of the POS terminals and hence the transactional charge income generated from each merchant is not known.

Hence Dashen Bank as one of the pioneers in the Ethiopian private banking industry, to achieve its vision of being the biggest bank in Africa should start to use the data in its storage to its advantage. Hence this study focuses on the following research questions

- What are the main data of the bank that should be integrated?
- What is the suitable data warehousing architecture for the bank?
- What is the suitable Conceptual Data Warehousing model for the bank's data set?
- What is the suitable Logical Data Warehousing model for the bank's data set?

1.4. OBJECTIVE OF THE STUDY

1.4.1. GENERAL OBJECTIVES

The general objective of the study is to develop a framework for the application of data warehouse technology in the Ethiopian banking industry specifically in Dashen Bank Share Company.

1.4.2. SPECIFIC OBJECTIVE

To achieve the above general objectives, the study will have the following specific objectives

- Assess how information is extracted, integrated and presented to in the Bank.
- Preparing the conceptual, logical and physical data model of the data set within the bank which are relevant for decision making process.

1.5. SIGNIFICANCE OF THE STUDY

Banks in Ethiopia use different systems to perform their day to day operation, from Core Banking system to Online Transaction processing systems and more. Hence all these systems collect a huge amount of date each day about the different aspects of the banks operation.

This research will provide new insights into the banks data collection by showing the framework for the application of data warehouse to use the historical data that can be retrieved and analyzed to provide useful insight into the organization's operations.

Through this research, the bank will further realize the benefit of data warehouse, data marts and online analytical processing, Specifically, this research will benefit the Bank in deciding which terminals, merchants and Branches are performing to the expected standard of the bank and make strategic decisions to get the best out of its infrastructure.

1.6. SCOPE AND LIMITATION OF THE STUDY

The scope of the study is limited to defining the business need for data warehousing technology in Dashen Bank S.c and illustration the application and design of a data warehouse which provide accurate and relevant information on conceptual and logical level by incorporating the data stored in different operational databases of the bank.

Although the study aims at meeting its goals of presenting a model for the application of data warehouse in Dashen bank S.c, it has potential limitation. The lack of practical experience on the area and the shortage of local studies related with data warehouse are some of the major limitation and the readiness of the banks to internalize the business needs of the application of data warehouse are among the many.

1.7. DEFINITION OF TERMS

Database: A database is a systematic collection of data that is organized to be easily accessed, managed and updated.

Relational databases: A relational database is a database in which data is organized into rows, columns and tables which are indexed for easier data access.

Data warehouse: Data warehouse is a repository of information collected from multiple sources, stored under a unified schema, and that usually resides at a single site and is subject-oriented, integrated, nonvolatile and time-variant. (Arvind Kumar 2016).

Data Mart: A Data Mart is a subset of the data warehouse which is highly structured subject areas repositories where data is stored and managed until it is needed for decision support based on the needs of a given department of an organization.

Data Mining: Data mining is the process of finding certain kinds of patterns in the data using a data mining tools like online transaction analytical processing (OLAP).

Core banking system (CBS): Core banking system is software that is used to provide a centralized online banking service for users and manage accounts, balances, transactions, journal entries along with the storage of client data, receipts, and other reporting tools.

Transaction processing systems (TPS): A transaction process system is an information processing system for business transactions involving the collection, modification and retrieval of transaction data. TPS is also known as transaction processing or real-time processing and known for its Characteristics of performance, reliability and consistency.

BI (Business Intelligence): The term Business Intelligence refers to technologies, applications and practices for the collection, integration, analysis, and presentation of business information. Business Intelligence systems are data-driven Decision Support Systems (DSS).

1.8. ORGANIZATION OF THE THESIS

The study is divided into six chapters. Chapter one of the study consists of the general introduction which includes; the background of the study, the statement of the problem, the objective of study, the research questions, significance of the study, the scope of study, limitations of the study, and the organization of the study. Chapter two is the literature review which first describes the basic concept of data warehouse and the different approaches of a data warehouse design and evaluates the works of other researchers on the subject, their approaches, and the researcher's criticisms.

Chapter three provides the brief description of the methodology which shows the general research approaches the data collection and analysis methods followed in this study. Chapter four describes the analysis phase of the study used to understand the current system status and readiness of the organization. Chapter five and six includes the design of the proposed system and conclusion and recommendation respectively.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. DATA WAREHOUSE CONCEPTS

Data warehouse is considered as a decision support database that is maintained separately from the organization's operational databases. A data warehouse stores a large collection of subject-oriented, integrated, time variant and nonvolatile data in support of management's decision-making process. It presents a multidimensional, logical view of the data, and is hence called a multidimensional database or data cube (Han, 1997).

Ralph Kimball is one of the scholars who is frequently referred for his accepted alternative definitions of data warehouse; According to Kimball a data warehouse development should follow a bottom-up approach that uses dimensional modeling, a data modeling approach unique to data warehousing developed by Kimball, rather than building a single enterprise wide database, Kimball suggests creating one database (or data mart) per major business process. whereas Bill Inmon advocates a top-down development approach that adapts traditional relational database tools to the development needs of an enterprise wide data warehouse and from this enterprise wide data store, individual departmental databases are developed to serve most decision support needs. (Breslin, 2004)

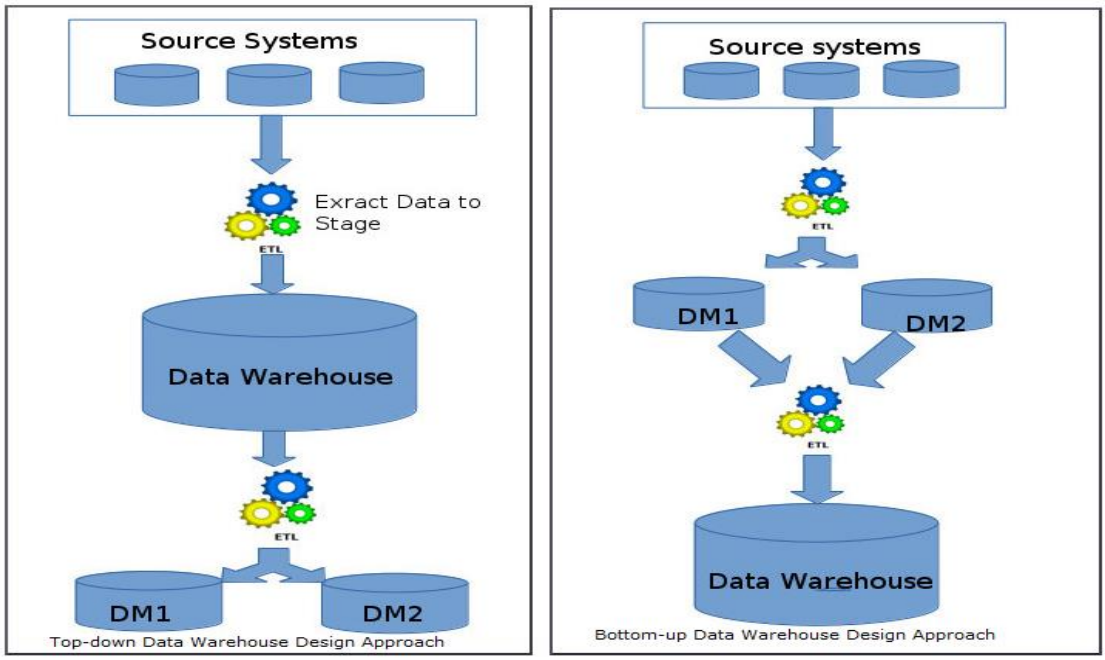


Figure 3. C. Suba Top down and bottom up Data warehouse approaches diagram

2.2. COMPONENTS OF DATA WAREHOUSE

2.2.1. ELEMENTS OF DATA WAREHOUSE

Different scholars describe the number data warehousing system elements as three, four and some even as five. As shown in the below figure Arvind Kumar describes the number as a four separate and distinct components named Operational source, Data staging area, Data presentation area and Data access tools

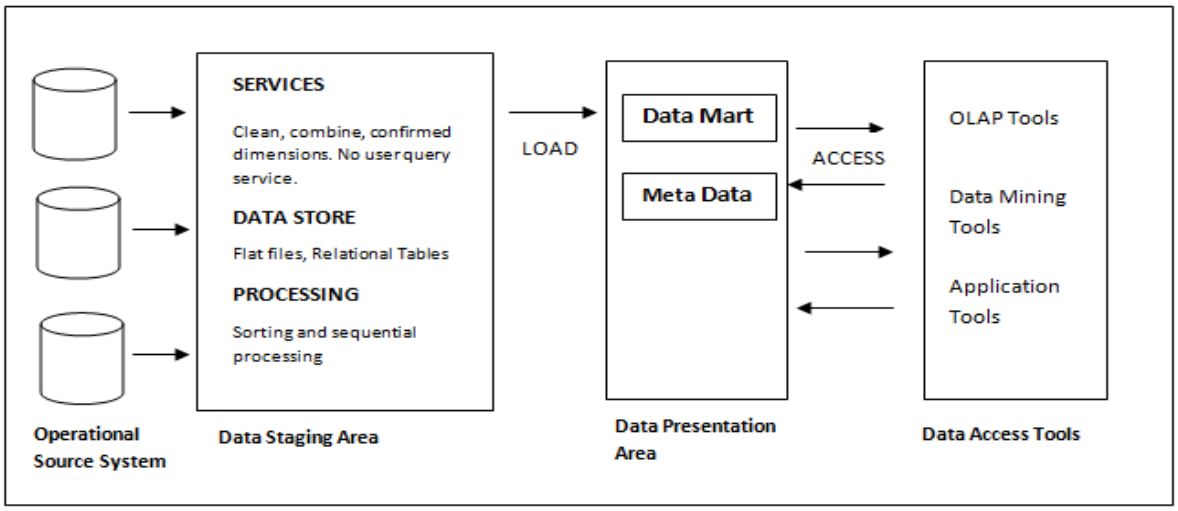


Figure 4. Basic elements of Data warehouse by Arvind Kumar (2014)

A study by Muhammad and his colleagues also states that a typical data warehousing system, which is acceptable across all the applications of data warehouse in real life, has four main elements each with their own functionality (Mohammed et al. 2016) Suba also agreed with the four elements of data warehousing system which grouped as Data capture, Transformation and clearing, Aggregation and Analysis and presentation.

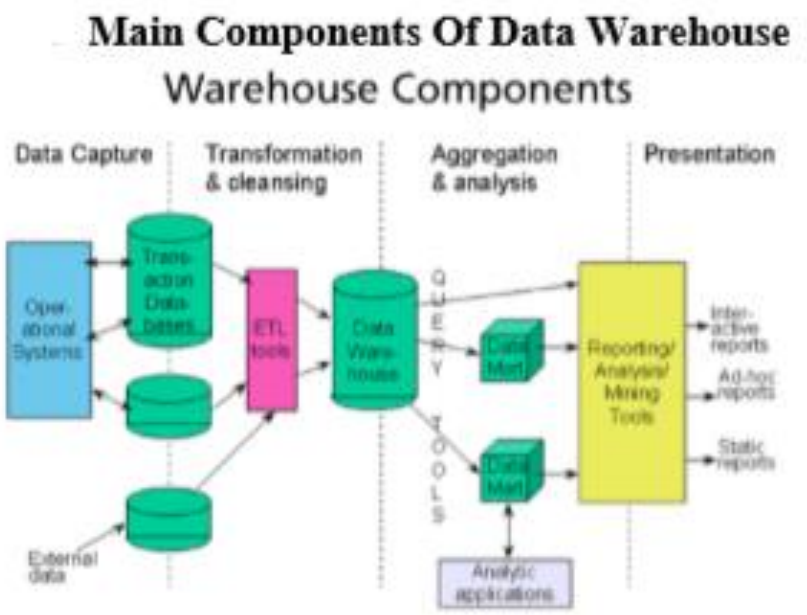


Figure 5. Main components of data warehouse by (Suba, 2018)

Chaudhuri and Dayal describes the data warehousing system with three components. A complete data warehousing system is composed of three primary parts the source databases in the backend, a data warehouse and several data marts in the core, and analysis tools in the front-end. (Yang et al. 2004)

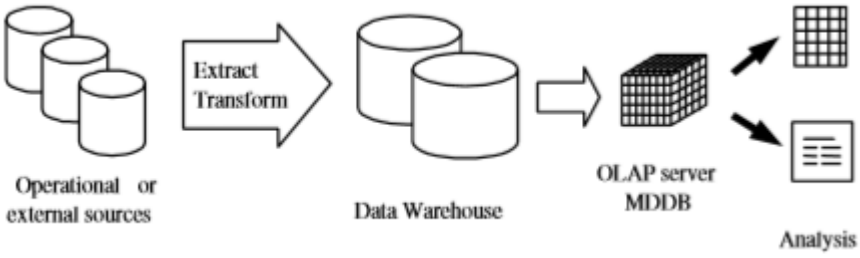


Figure 6. General structure of the Data warehouse system with OLAP service (Yang et al. 2004)

According to Yang Lin and Chung Kuo, since the data is fetched from multiple sources Data staging component serves as the data preparation point to integrate the data. The main Data Storage component which handles the data warehouses data also includes Data marts. This part not only stores and manages the data but also keeps track of data using the metadata repository. The access tools or information delivery part consists of all the different ways of getting the information from the data warehouses available to the users, in this case OLAP.

2.2.2. ONLINE ANALYTICAL PROCESSING (OLAP)

OLAP stands for an Online Analytical Processing and is software that allows users to extract and analyze large volumes of data from a data warehouse, data mart, or some other centralized data store. OLAP refers to the general activity of querying and presenting text and number data from data warehouses and/or data marts for analytical purpose. (Jukic et al. 2008). OLAP is a technology that is design for ad-hoc data access and analysis through querying and presenting text and number data from data warehouses and is characterized as Fast Analysis of shared Multidimensional Information (FASMI). OLAP refers to a set of data analysis techniques developed for analyzing data in data warehouses (Jiawei 1997)

Creating an OLAP system consists in taking a photograph of information at a given moment and turning it into multidimensional data. An OLAP system allows studying a large amount of data; seeing data from different perspectives and supporting decision making. (Galloet al. 2010).Since the analysis is to give deep insight into the stored data using different operations OLAP uses both historical as well as operational transactional data in the analysis. OLAP refers to a set of data analysis techniques developed for analyzing data in data warehouses since 1990s. (Han 1997).

2.2.3. OLAP OPERATIONS

This section gives an overview of the common functionalities of all OLAP tools. The four OLAP operations which are to be performed on multidimensional data that are used regularly by analysts are commonly referred as Roll-up, Drill-down, Slice and dice and Pivot (rotate). According to Jukic the purpose of these operations is to increase (in the case of Drill Down) or decrease (in the case of Drill Up) the granularity of the data shown in the query result. (Jukic et al. 2008).

Chaudhuri also elaborates the functionality as OLAP operations include rollup (increasing the level of aggregation) and drill-down (decreasing the level of aggregation or increasing detail) along one or more dimension hierarchies, slice and dice (selection and projection), and pivot (re-orienting the multidimensional view of data). (Chaudhuri et al 1997).

Roll-up creates a subtotal at any level of aggregation needed from the most detailed up to the grant total. This function is also called consolidation. Drill-down breaks the subtotal at any level of granularity to lower level of granularity in hierarchy, means it gives details of relationship at the lower level. Slice and dice is selecting a subsection of data cube based on the constants in one or few dimensions. If one dimension is fixed, the operation is called slice and if more than one

dimension is fixed, the operation is called dice. Pivoting is swapping of columns and rows. This allow user to look at data from different view. This is also commonly known as rotation. (Lovin 2004)

2.2.4. OLAP TYPES

Although Arvind Kumar describe it to be three types, Radu Lovin states that basically there are only two technologies for the implementation of OLAP servers, namely Relational OLAP (ROLAP) and Multidimensional OLAP (MOLAP) the third is only the hybrid of the two and called Hybrid OLAP (HOLAP). (Lovin 2004).

According to Lovin OLAP might be implemented on standard or extended relational DBMS, called Relational OLAP server. This server support extension to SQL and special access and implementation method to efficiently implement the multidimensional data model and operations. Multidimensional OLAP on the other hand directly store multidimensional data in some special data structures (such as arrays) and implement the OLAP operations over these specials data structures.

The third type which is Hybrid OLAP is a combination of ROLAP and MOLAP and was developed to combine the greater data capacity of ROLAP with the superior processing capability of MOLAP. Typically, it stores data in both relational database and a multidimensional database and uses whichever one is best suited to the type of processing desired.

2.3. WHY DATA WAREHOUSE

The fundamental business driver behind data warehouse is the desire to improve decision making and organizational performance (Watson et al. 2002).

Although Some of the general benefits of data warehouse are considered to be Improved data quality, Better business analytics, Faster queries and Historical insight, Scholars divide the benefits of a data warehouse application into tangible and intangible, moreover they also aggregate the benefits into several levels that have varying degree of impact on the organization locally as well as globally. Data warehouse application can have a wide variety of tangible and intangible benefits (Watson et al. 2002).

According to their study time saving and more and better information are the most tangible from the rest down the list, which shows a clear time improvement to finish a task and an increased in quantity and quality of information.

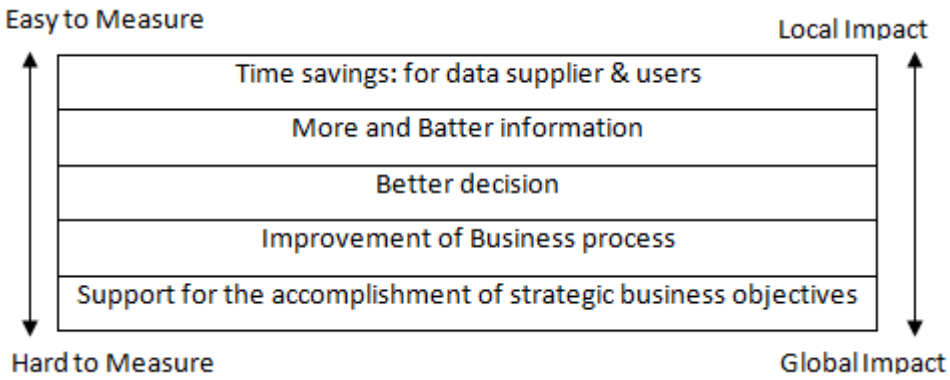


Figure 7. Diagrammatic representation Benefits of Data warehousing (Watson et. al 2002).

Apart from the above list Mohammed Kemal (2019) names some general benefits of data warehousing as improved data quality, process automation, better reporting, better decision making, instant data access, and historic data integrations.

Since the introduction, data warehouse has been widely adapted in several sectors and successful implementations have been reported in healthcare, agriculture, geo-information systems, banking sector and business environments in order to enhance the analytical capabilities of enterprises. (Daniel Fasel 2012).

According to Fasel some industry specific data warehousing benefits are listed in the table below

Banking and finance
• Identify the potential risk of default and manage and control collections • Performance analysis of each product, service and exchange rates • Provide feedback to bankers regarding customer relationships and profitability • Evaluation of customer expenses trends
Government
• Maintain and analyze tax records and their respective providers • Prediction of activities from patterns and trends • Threat assessment and fraud detection

• Resource allocation and management
Manufacturing
• Predict market changes and analyze current business trends • Analyze previous and current market data • Identify profitable product lines and required product features
Business
• Maintain records of producers and consumers • Track items and consumer buying trends (trend analysis)

Table 1. Industry specific data warehousing benefits list by (Kumar 2016)

2.4. DATA WAREHOUSES IN BANKING INDUSTRY

Banking industry is one of the businesses which is hugely depends on the collection, processing, storing and use of customer and transaction data and information. This stored date is an expense in terms of storage space if it is stored without any use, but it is an opportunity to improve the performance of the Bank if it is used. Since the introduction of data warehousing technology the banking industry was one of the most suitable industries considered for the implementation of the technology. In banking, there are many different systems in use, such as card system, reconciliation system, accounting system, ATM management/ switching system etc. Those systems contain different data, and these data can provide insight about the business when they are gathered and processed. Besides providing Enhanced Business Intelligence data warehouse can deliver a lot more benefits to the banking industry like Integrate data from multiple source systems, enabling a central view across the enterprise, saving time Enhances Data Quality and Consistency and provide access to historical data of the Bank. (Ugochukwu et al.2019) While implementing data warehousing technology the scope can include almost all areas of the banks operation like (CRM) customer relationship management, Enterprise Risk Management, Performance Management and Asset and liability Management (Bogdan et al. 2011)

2.5. ETL PROCESS

ETL is short for extract, transform and load, the three database functions that are combined into one pipeline to pull data out of one database and place it into another database. Pipeline infrastructure always implements a set of ETL operations, extracting data from source databases, Extraction is a process that concerns what data that shall be stored in the data warehouse. Only data that can be used for decision making and other typical data warehouse operations is of interest. (Kimball et al. 1998)

Transforming data to match a unified format for specific business purposes there are different operations related to the transformation, the Data cleaning is the correction of misspellings and broken integrity constraints. Aggregation of data means that data is stored at different levels of detail and this improve the performance of the data warehouse when the user runs common queries. Integration means that data on different formats is viewed in one ordinary format that users apply when accessing the data warehouse, and Loading reformatted data to the data warehouse. Some of the functionality performed by ETL tools are

- Eliminating unwanted data in operational databases and loading into Data warehouse.
- Search and replace common names and definitions for data arriving from different sources.
- In case of missing data, populate them with defaults.
- Remove duplicated data arriving from multiple data sources.

2.6. DATA CUBE

A data cube is a pictorial representation of a multidimensional data warehouse model which shows how the data in a multidimensional data warehouse is presented. A data warehouse stores a large amount of data from different sources thus it presents a multidimensional, logical view of the data and hence is called a multidimensional database or data cube. OLAP operations are the ways used to interact with the data cube for multidimensional analysis (Jiawei 1997).

Multidimensional data cubes consist summary tables for all possible decision marking activities in an organization. In an effort to help understand the OLAP functions on the possible data cube a pictorial representation is presented to show the number of product sold in a store located at different locations and a more illustrative pictorial representation is also provided to show the revenue collected form stores which are located in different parts of the world (Jiawei 1997).

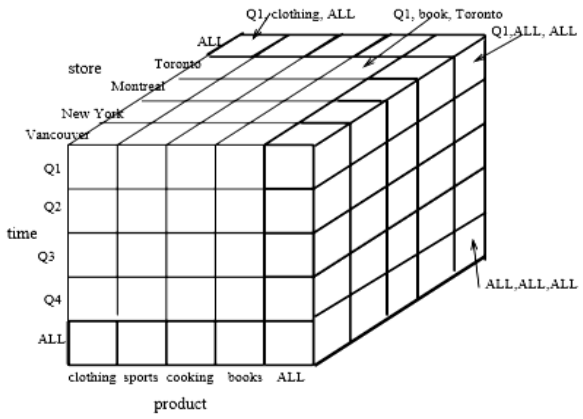


Figure 8. 3D data Cube of a market data warehouse by (Jiawei 1997).

In his example Jiawei provided a three dimensional data cube of a market data warehouse which shows the number of product sold in a store located at different locations as a function of time.

OLAP applications are based on multidimensional modeling that automatically represents data under the metaphor of a cube whose cells correspond to events that occurred in the business domain. Each event is quantified by a set of measures; each edge of the cube corresponds to a relevant dimension for analysis, typically associated to hierarchy of attributes that further describe it. (Rizzi 2007).

The cube metaphor for multidimensional modeling

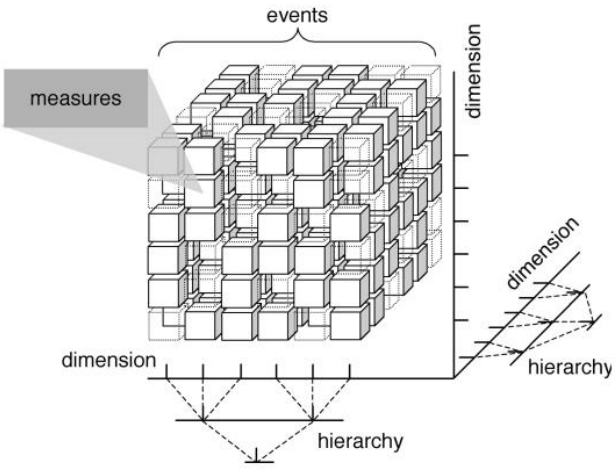


Figure 9. Stefano Rizzi depiction of OLAP Application using CUBE

2.7. SIGNIFICANT PRIOR RESEARCH

Organizations common goal is to make a better business decision. Once a data warehouse is implemented in an organization beside saving time and supporting decision making process it can benefit the organization in numerous ways and brings a competitive advantage for the organization. Data warehouse systems offer efficient access to integrated and historic data from heterogeneous sources to support managers in their planning and decision-making (Beate et al. 2002).

Although there are no publicly available local studies on the application of data warehouse, the study has been addressed previously by foreign scholars. Earl Von F. Lapura and his colleagues presented a study under the title Development of a University Financial Data Warehouse and its Visualization Tool (Lapura et al.2018).In their study they have showed the financial data warehouse developed Based on the Kimball data warehouse development modeling Techniques of multidimensional model that integrate time, finance-unit, account and fund dimensions with the Fund-amount fact table using start schema, which is updated periodically with the accumulated transactional data sourced from a financial database of a university. The study followed a general methodology consisting of four phases namely planning, data warehouse development, visualization application development and system evaluation to produce the conceptual, logical and physical design of the data warehouse with a visualization tool and successfully integrate it with the university web portal for end-user's test. Since the study of Lapura's mainly focuses on the educational system there are many aspects of the of Banking industry are not considered in the study.

Ferreira and his colleagues also presented a study under the title Building an Effective Data Warehousing for Financial Sector (Ferreira et al 2017) and shows the implementation process of a Data Warehouse using multidimensional analysis of business data for a holding company in the financial sector. The goal was to create a business intelligence system that, in a simple, quick but also versatile way, allows the access to updated, aggregated, real information, regarding bank account balances. The established system extracts and processes the operational database information which supports cash management information by using Integration Services and Analysis Services tools from Microsoft SQL Server. The results have shown that the adoption of online analytical processing cubes offers better performance and provides a more automated and

robust process to analyze current and provisional aggregated financial data balances compared to the current process based on static reports built from transactional databases. Although Ferreira's study is on the financial organization the study mainly focuses on the Account balance of customers and cash management information ignoring the performance of terminals who are the basic for the transactions performed.

The study of (Gallo, 2010) under the title Data Warehouse Design and Management Theory and Practice has been significant input to understand and analyze Data Warehouses with their theoretical models, and illustrate a practical implementation in a specific case study on a pharmaceutical distribution company. The case study used Kimball's data warehouse design modeling to show the design and implementation of a Data warehouse populated with data relating to sales of a pharmaceutical product distribution company, which integrate Invoice fact table with Local health district (LHD), category, supplier, city, customer, time and distribution, dimensional tables. Although the study of Gallo is significant to understand analysis and design of a data warehouse the focuses is on the sales of a pharmaceutical product distribution company many of the banking industry aspects that should be considered are missed.

Although there are some Case studies on the application of Data mining techniques by Belachew Reganie (2013) from AAU under the title "application of data mining techniques for customers segmentation and prediction", Mengistu Tesfaye (2013) from AAU "application of data mining in credit risk assessment the case of united bank s.c" and Luel Berhe (2011) from AAU under the title "The role of data mining technology in electronic transaction expansion at Dashen bank s.c" , there is no locally available studies found on the application of data warehouse on the banking industry.

Since data warehousing is a platform for data mining, to perform data cleaning, integration and transformation and prepare the data in a format suitable for data mining tools, all of the above local studies recommended the application of Data warehouse in the banking and finance industry. More importantly since the study by Luel Berhe focuses on Dashen Bank S.C electronic transactional data, which is considered to be one of the source data for this study, has become a springboard for this study. In his recommendation Leul states that the quality of data is one of the factors that shade on the performance of the data mining techniques during extraction of huge amount of data, Hence, the data must be cleansed, integrated and transformed in the format suitable for data mining tools

and thus the application of data warehouse is an important issue that should be done in the future studies.

2.8. CHALLENGES OF DATA WAREHOUSING IN BANKING INDUSTRY

The accumulation of data from different systems created an enormous amount of data to be stored in the banks storage and the use of this huge amount of data is also becoming more and more challenging for the Banks. DW and BI applications have been built to consolidate enterprise business data, allowing taking timely and informed decisions based on up-to-date data. With the rapid growth of accumulated data in different formats the design and implementation of data warehousing and extraction tools is also changing rapidly. The availability of enormous amounts of data from different domains is calling for a shift in the way DW and BI practices are being carried out (Vaisman et al.2012).

Data warehouses provide banks with the ability to integrate data from many sources to create a single source of truth. From this single source of truth, banks are able to generate reporting and analytics tools that leverage data to make the most informed business decisions possible. According to Markus and Dirk Data warehousing for Banks is different. They stated that it is different because unlike many of the software projects, data warehousing projects are not developed keeping a front-end application in mind. For the most part of it, these projects are heavily dependent on the backend infrastructure in order to support the front-end client reporting. Moreover, number of different stake holders involved in data warehousing projects is usually more than any typical IT project.

As per Markus and Dirk, analysis of the existing data warehousing systems revealed the following:

- Data warehouse development has become expensive and slow because of the high complexity of the historically evolved environments.
- The integration of off-the-shelf software was very difficult because interfaces were either not compliant with standards or were not clearly defined and documented.
- Enterprise-wide data quality process could hardly be established, since each warehouse implemented its own data semantics. (Markuset al.2011)

2.9. MULTIDIMENSIONAL DATAWAREHOUSE MODELS

Data warehouse models are basically different from the traditional transactional database systems, which typically dependent on much normalized and related data. According to Gallo and colleagues study the design of a typical relational database is based on ER Model which depends on much related data with different characteristics and contains different quantity, such type of design is therefore do not allow a multi-dimensional view on the data. Hence in order to be able to use the maximum available information we can get from these collection of data it is necessary to adopt a model which allow us a multi-dimensional view of the data.

Multidimensional Data Model can be defined as a method for arranging the data in the database, with enhanced structuring and organization of the contents in the database. The notion of a dimension is an essential and distinguishing concept for multidimensional databases, Dimensions are used for two purposes the selection of data and the grouping of data at a desired level of detail. (Torben,2016). In multidimensional model dimensions are entities in which an organization wants to keep records and views data in the form of a data-cube. A data cube enables data to be modeled and viewed in multiple dimensions and is defined by dimension tables and fact tables. A multidimensional data model is organized around a central theme, for example, sales. This theme is represented by a fact table. The fact table contains the names of the facts or measures of the related dimensional tables.

Ralph Kimball stated in his book, The Data Warehouse Toolkit that a multidimensional data model is important because it enforces simplicity. The central attraction of the dimensional model of a business is its simplicity.... that simplicity is the fundamental key that allows users to understand databases, and allows software to navigate databases efficiently.

To allow a multi-dimensional view of data, techniques have been developed known as “schema”, schema is a collection of database objects, including tables, views, indexes, and synonyms. As per (Gallo et al. 2010) There are three types of schemas for the implementation of multidimensional data models known as star schema, snowflake schema and mixed / constellation/ multi-star model schema.

2.9.1. STAR SCHEMA

The simplest data warehouse schema is star schema because its structure resembles a star. Star schema consists of data in the form of fact tables and dimension tables. The star schema is the basic structure for a dimensional model. It typically has one large central table called the fact table and a set of smaller tables called the dimension tables arranged in a radial pattern around the fact table (IBM technical support1998). In star schema fact table contain a large amount of data, with no redundancy. Each dimension tables are joined with the fact table using a primary or foreign key. As an example the IBM technical support group depicts sales as a fact table in the center of star schema. Arranged around the fact table are the dimension tables of time, customer, seller, manufacturing location, and product.

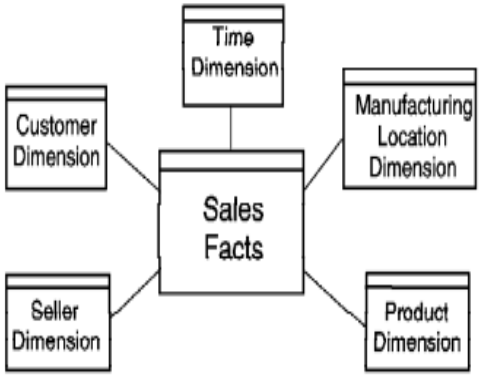


Figure 10. Star schema depiction

The traditional ER model has an even and balanced style of entities and complex relationships among entities, the dimensional model is very asymmetric. Even though the fact table in the dimensional model is joined with all the other dimension tables, there is only a single join line connecting the fact table to the dimension tables. (IBM technical support1998)

2.9.2. SNOWFLAKE SCHEMA

The snowflake schema is represented by centralized fact table which is connected to multiple dimension tables and this dimension table can be normalized into additional dimension tables. The snowflake schema is a more complex than star schema because dimension tables of the snowflake are normalized. The model is the result of decomposing one or more of the dimensions, which sometimes have hierarchies themselves. The above given example diagram of star schema can also

be more normalized and expressed in snowflake model with increased complexity as shown in the figure below. (IBM technical support1998)

Advantages of snowflake compared to the star technique are a query performance improvement through better storage occupation due to elimination of redundant data and use of small normalized tables instead of large not normalized tables. The disadvantages are due to the fact that it presents a more complex structure. There is a greater number of tables and therefore greater difficulty in deciding which table to use in a query. (Gallo et al. 2010)

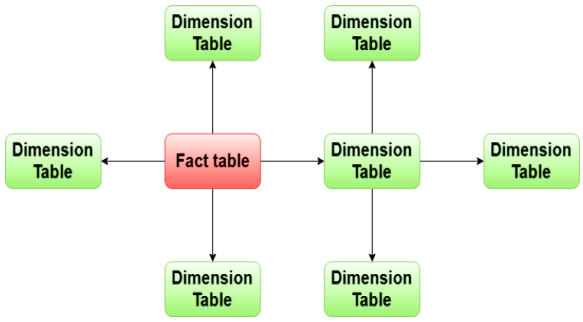


Figure 11. Snow flake schema

Star Schema versus Snowflake Schema	
Star Schema	**Snowflake Schema**
Hierarchies for the dimensions are stored in the dimensional table.	Hierarchies are divided into separate tables.
It contains one fact table surrounded by dimension tables.	One fact table surrounded by dimension table which are in turn surrounded by dimension table
Only single join creates the relationship between the fact table and any dimension tables.	Requires many joins to fetch the data.
Simple DB Design.	Very Complex DB Design.
De-normalized Data structure	Normalized Data Structure.
High level of Data redundancy	Very low-level data redundancy
Single Dimension table contains aggregated data.	Data Split into different Dimension Tables.

Table 2. Comparison of star and Snowflake schema (Kumar 2016)

2.10. MIXED / CONSTELLATION/ MULTI-STAR SCHEMA.

In the situations where some dimension tables with substantial differences in the number of attributes and volume it is not possible to use the star or snowflake model for the whole structure, therefore use of combination of both, called mixed model will be necessary. The decision on which data model to use depends on data characteristics and requirements of the organization that will use the DW. This type of schema can be viewed as a collection of stars, Snowflake and hence is called a galaxy schema or a fact constellation.

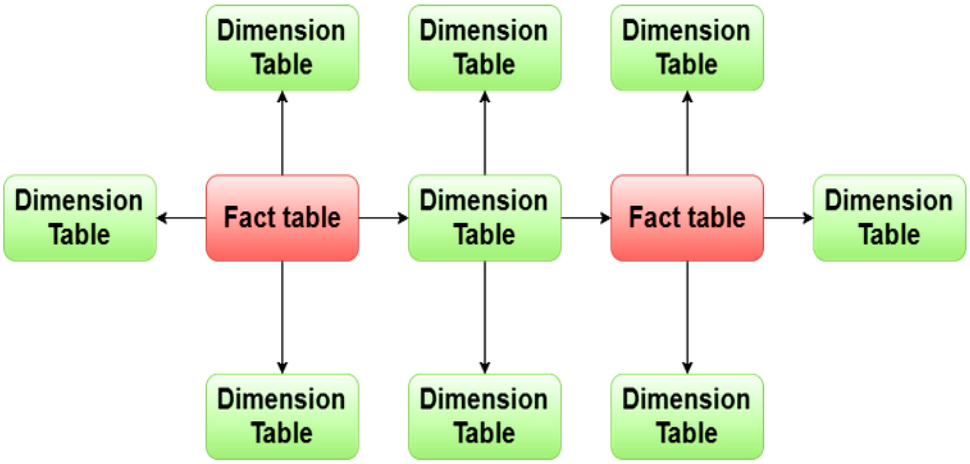


Figure 12. Constellation schema

CHAPTER THREE

3. METHODOLOGY

3.1.INTRODUCTION

In this chapter the research methodology adopted by this study will be explained, the research approaches and design, techniques applied in conducting data collection and sampling techniques are discussed.

3.2. GENERAL RESEARCH METHOD

For this particular applied research, a qualitative research methods are believed to be more suitable in data collection and describing the process of developing the data warehouse, analyzing of the current system and designing of the conceptual and logical design of the data warehouse. This approach is appropriate in this research where the participants were involved in the focus grope discussion of various discussion points and knowledge of data warehousing technology questions are discussed in depth. The qualitative research approach demands that the world be examined with the assumption that nothing is trivial, that everything has the potential of being a clue that might unlock a more comprehensive understanding of what is being studied. (Bogdan et. al 1998)

3.3. POPULATION, SAMPLE

As the aim of the study is to demonstrate the application of data warehousing technology in the Ethiopian banking industry the wider population of the study are all commercial banks currently running in the Ethiopian banking industry, out of these banks Dashen Bank S.c is selected as a sample for the study The study applied convenience sampling technique considering the easy accessibility of data and willingness of the IT as well as operational staff member to participant in the study are the main reasons to select these sampling technique.

Convenience sampling is a type of non-probability or nonrandom sampling where members of the target population that meet certain practical criteria, such as easy accessibility, geographical proximity, availability at a given time, or the willingness to participate are included for the purpose of the study. (Etikan et al. 2016).

3.4. DATA SOURCES

To better understand the readiness of the bank for the application of data warehousing technology, to analyze the current data integration status in the Bank and as a qualitative research the data used

in this study will be mainly gathered through focus group discussion, observation, interview, and document evaluation. Since the main data collection methods are Observation and focus group discussion the main data sources are the professional staff members who are working on the current system, direct observation and available documents.

3.5. DATA COLLECTION PROCEDURE

The focus group formed for the study consists of six members from the different sections of the IT department two member are Data analyst and MIS engineers from the database& MIS department two members are Alternate channel engineers from Online channel department and two members are system administrators from infrastructure department and one manger from database section are willing to participate in the discussion and interview.

A focus group discussion (FGD) is a good way to gather together people from similar backgrounds or experiences to discuss a specific topic of interest. The group of participants is guided by a moderator (or group facilitator) who introduces topics for discussion and helps the group to participate in a lively and natural discussion amongst them. (Mishra 2016)

The focus group discussion questions are designed in a way that the same questions forwarded to all the participants of the focus group to know the level of understanding of the participants on the concepts of data warehouse and its application and to understand the readiness of the bank to apply data warehouse. For the quality of information and time limitation interview is conducted only with the manager who is currently working as the Data analyst and MIS manager.

3.6. DATA ANALYSIS

As Indicated in section 3.4 of this study, to understand the current data integration in the bank and reports generated for the different internal and external parties, observation, document analysis, interview and focus groups discussion techniques are applied. The data collected through the above techniques are analyzed from the following points of views.

- The data source or the number of operational systems involved to feed the warehouse.
- Report generated for the different internal and external parties. (Relevance wise)
- End user accessibility.
- Historic data incorporation.
- Varity and size of data

The group discussion and observations conducted focuses on the systems involved in the data integration and the different tables and views involved from the different systems, moreover reports generated by the current system and the process in which how the diverse user request are entertained. The information input collected from the focus group which includes senior IT professionals currently working on the database and MIT department is used as the major input to identify the different dimensions and fact tables needed to design the data warehouse and the different systems needs to be included in the data design.

Besides identifying the current report deliverables by the system as compared to the request provided to the section, the document investigation mainly focuses on understanding the content of the diverse type of reports generated and know the current end user requirements which are not able to be fulfilled by the current data integration in the bank. The information collected through the interview made with the senior manager is used to triangulate the data collected from the IT professionals through focus group discussion and the information gathered through document assessment.

3.7. RESEARCH APPROACHES

The study aims to show the conceptual, logical and physical design of data warehousing technology in the Ethiopian Banking industry, which solves the data analysis and accessibility problem of the banks historical as well as operational data by increasing the availability and quality of information, through this improve the quality of decision making ability of the management, Hence the research will follow a qualitative research methods based on the previous studies in the area.

This study is designed based on Ralph Kimball's data warehouse development approach. According to Kimball data warehouse development model data warehouse development passes through four key steps of selecting the business process, declaring the grain, identifying the dimensions and identifying the facts.

During selection of Business Process and Analysis, the required tables which contain the data are chosen and transformed into the dimension tables. In Choose the Granularity, the relationship between data and information within transaction table and fact table is identified. The next is Creating Dimension Tables. In this step, the dimension tables that are related to the fact tables are created and during the Creation of Fact Tables, tables which contain measurable data are created.

3.8. WHY KIMBAL'S APPROACH

Data warehouses offers a chance into an organization's historical performance and operations, providing data analysts and business users with information on things such as customer behavior, business trends, operational efficiency and sales, Hence The choice of data warehouse design approach is one of the first issues that should be considered. And the question of why Kimball approach is the appropriate for this study should be answered.

3.8.1. INMON'S APPROACH

Bill Inmon's Atomic Data Warehouse approach is strategic in nature and seeks to capture all of the enterprise data in 3rd Normal Form and store all of this data in the data warehouse. The architecture of the Inmon data warehouse includes all the company's information systems with their databases instead of considering only information fragments (Yessad 2016). Inmon then creates data marts, subject or department focused subset of the data warehouse, which is designed to address the reporting needs of the targeted subset of business users. This process provides the organization with a complete view of their processes, products/services, customers, vendors, etc.

Yessad states that Inmon approach is based on the Entity-relationship diagrams of operational systems. The company's data are loaded without knowing a prior user requirement; this is why it is called "data-driven" approach (Yessad 2016).

According to yessad, Inmon divides the environment of the company's databases into four levels as operational, Atomic (Data Warehouse), departmental (data marts) and individual levels and considers that a data warehouse and a data mart are physically separated.

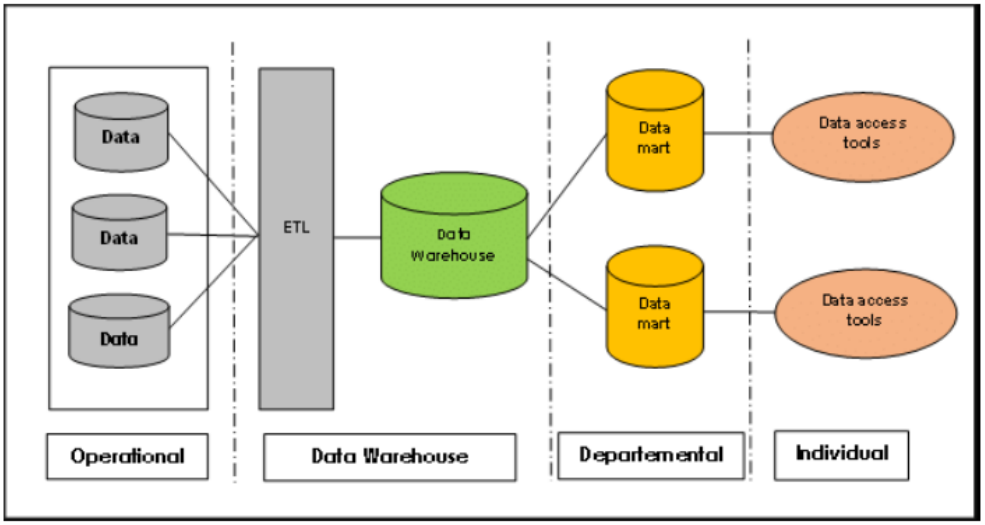


Figure 13. Inmon's data warehouse architecture

Inmon's design approach follows a spiral methodology of development for the implementation of the warehouse which consists three stages. To establish the model, Inmon offers three levels of data modeling namely Entity-Relationship Diagrams, Mid-Level model/DIS (Data Item Set) and the physical model (Breslin 2004)

Entity-Relationship Diagrams (ERD): For modeling a high-level abstraction of data. This model starts by identifying business subjects (entities) of the company and then relations between these entities. For each department of the company that wants to use the data warehouse, an ERD is created and all the ERDs will constitute the ERD of the company.

Mid-Level model/*DIS (Data Item Set):* The second level of modeling is where we find the most information about the company's data model. For each entity identified in the company's ERD, a DIS is created. This level of modeling contains keys, attributes, groups of attributes and connectors. The midlevel data model includes four constructs a primary data grouping, a secondary data grouping, a connector and type of data.

The physical model: This is the last level of modeling. It is created based on the second level of the data model. For each part of the DIS, there will be a unique and separate physical data model. This model looks like the relational tables.

The key advantages of the Inmon approach are:

- The data warehouse serves as the single source of truth for the enterprise, as it is the only source for the data marts and all the data in the data warehouse is integrated.
- Data update anomalies are avoided because of very low redundancy. This makes ETL process easier and less prone to failure.
- The business processes can be understood easily, as the logical model represents the detailed business entities.
- Very flexible – As the business requirements change or source data changes, it is easy to update the data warehouse as one thing is in only one place.

Here are some of the disadvantages of Inmon method:

- The model and implementation can become complex over time as it involves more tables.
- Need resources who are experts in data modeling and of the business itself. These types of resources can be hard to find and are often expensive.
- The initial set-up and delivery will take more time.
- More ETL work is needed as the data marts are built from the data warehouse.
- A large team of specialists need to be around to successfully manage the environment.

3.8.2. KIMBALL'S APPROACH

Kimball approach is based on the concept of dimensional modeling. This approach of building the data warehouse starts with identifying the key business processes and the key business questions that the data warehouse needs to answer whereas Inmon's approach begins with the corporate data model which identifies the key subject areas, and most importantly, the key entities the business operates with.

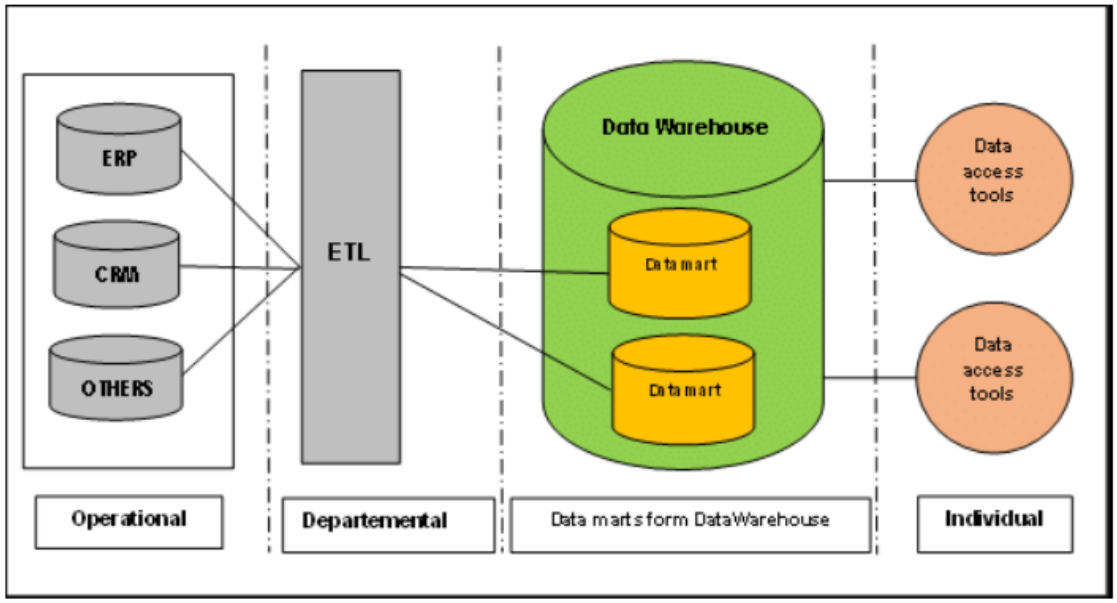


Figure 14. Kimball's data warehouse architecture

In Kimball's architecture, data is copied from operational source systems to a staging area. In the staging area, the data is scrubbed, that is, made consistent and suitable for end-user queries. From the staging area, data is loaded into data marts. The data marts are the source of data for user queries. Each data mart is based on a single business process. More than one department may be

interested in a given business process; therefore, no one department is perceived as the sole owner of a given data mart (Baslen, 2004)

A multidimensional view of data model contains basic concepts like fact table, dimensions, conformed dimensions and bus matrix.

Fact table includes observable data (facts) about the subject we want to study, using various analytical axes (dimensions). This may be the amount of sales, the number of units sold of a product, etc…, Dimension contains the analysis axis (dimensions) wherein we want to study facts. Subjected to a multidimensional analysis, these data give users the information needed for decision-making. This may be customers or products of a company, etc. Data warehouse bus is the part of Kimball's architecture that allows the sum of the data marts to truly be an integrated whole data warehouse. (Yessad 2016) The bus architecture is another way of saying that all data marts must use standardized conformed dimensions. (Baslen, 2004) Data Bus is the creation of the enterprise data warehouse bus matrix, and is the result of a requirements gathering process. The matrix can be defined so that each row represents a business process and each column corresponds to a dimension of the business. (Bogzan 2015)

Conformed dimensions are a dimension that has the same meaning to every fact with which it relates. Conformed dimensions allow facts and measures to be categorized and described in the same way across multiple facts and/or data marts, ensuring consistent reporting across the enterprise. Date is a common conformed dimension because its attributes (day, week, month, quarter, year, etc.) have the same meaning when joined to any fact table. The basic requirements of conformed dimensions are that keys, column names, attribute definitions, and attribute values are consistent across business processes. (Baslen, 2004)

Kimball's Philosophy is summarized in short as

- Make data easily accessible
- Present the organization's information consistently
- Be adaptive and resilient to change
- Protect information
- Service as the foundation for improved decision making.

The fundamental concept of dimensional modeling is the star schema. In the star schema, there is typically a fact table surrounded by many dimensions. The fact table has all the measures that are relevant to the subject area, and it also has the foreign keys from the different dimensions that surround the fact. The dimensions are de-normalized completely so that the user can drill up and drill down without joining to another table. Kimball proposes the concept of 'conformed dimensions'. The key dimensions, like customer and product, that are shared across the different facts will be built once and be used by all the facts to keep the integrity of the data.

Some of the Kimball design method advantages (Breslin, 2004).

- Quick to set-up and build.
- Star schema can be easily understood by business users and is easy to use for reporting.
- Most Business Intelligence tools work well with star schema.
- Needs less database space and makes the management of the system fairly easier.
- The performance of the star schema model is very good.
- Small team of professionals is enough to keep the data warehouse performing effectively
- Works well for department-wise metrics and KPI tracking, as the data marts are geared towards department-wise or business process-wise reporting.

Here are some of the disadvantages of the Kimball method:

- The essence of the 'one source of truth' is lost, as data is not fully integrated before serving reporting needs.
- Redundant data can cause data update anomalies over time.
- Adding columns to the fact table can cause performance issues. If new columns are to be added, the size of the fact table becomes much larger and will not perform well.
- This makes the dimensional model hard to change as the business requirements change.
- Cannot handle all the enterprise reporting needs because the model is oriented towards business processes rather than the enterprise as a whole.
- Integration of legacy data into the data warehouse can be a complex process.

Mary Breslin in her 2004 article of Data Warehousing Battle of the Giants comparing the Basics of the Kimball and Inmon summarizes the above data warehouse design methodology features and

characteristics of Bill Inmon and Ralph Kimball in two tables categorizing the features into three as Methodology and architecture, Data orientation and Philosophy as shown below. (Breslin 2004)

	Inmon	**Kimball**
Methodology and Architecture		
Overall approach	Top-Down	Bottom-Up
Architectural Structure	Enterprise wise (Atomic) data warehouse feeds departmental Databases	Data marts Model a single business process, enterprise consistency achieved through data Bus and conformed dimension.
Complexity of the method	Quite complex	Fairly Simple
Comparison with established development methodology	Derived from Spiral methodology	A four step process, departure from RDBMS.
Discussion of physical design	Fairly thorough	Fairly light
Data Modeling		
Data Orientation	Subject or Data driven	Process oriented
Tools	Traditional (ERDs, DISs)	Dimensional Modeling, departure from relational Modeling.
End User Accessibility	Low	High
Philosophy		
Primary Audience	IT professional	End Users
Place in the Organization	Integral part of the corporate information factory (CIF)	Transformer and retainer of operational data.
Objective	Deliver a sound technical solution based on proven database methods and technologies.	Deliver a solution that makes it easy for end users to directly query the data and still gets reasonable response time.

Table3 Comparison of Essential Features of Inmon's and Kimball's Models by Mary Breslin 2004

Characteristic	**Favors Kimball**	**Favors Inmon**
Nature of the organization's decision support requirements	Tactical	Strategic
Data integration requirements	Individual business areas	Enterprise-wide integration
Structure of data	Business metrics, performance measures, and scorecards	Non-metric data and for data that will be applied to meet multiple and varied information needs.
Scalability	Need to adapt to highly volatile needs within a limited scope	Growing scope and changing requirements are critical

Persistency of data	Source systems are relatively stable	High rate of change from source systems
Staffing and skills requirements	Small teams of generalists	Larger team(s) of specialists
Time to delivery	Need for the first data warehouse application is urgent	Organization's requirements allow for longer start-up time
Cost to deploy	Lower start-up costs, with each subsequent project costing about the same	Higher start-up costs, with lower subsequent project development cost.

Table (4) Specific Characteristics Favoring Inmon's or Kimball's Model by Mary Breslin 2004

3.8.3. DATA VAULT'S APPROACH

Dan Linstedt defines the Data Vault as "a detail-oriented, historical tracking and uniquely linked set of normalized tables that support one or more functional areas of business"(Bogza et al. 2015). The data vault modeling is a hybrid approach based on third normal form and dimensional modeling aimed at the logical enterprise data warehouse. According to Linstedt, the 3NF of Inmon and the dimensional modeling of Kimball have weaknesses as the data volume increases. Thus, the "Data Vault" is interesting for changes in processes and data structures rather than changes in business functions. (Bogza et al. 2015).

The main features of the Data Vault listed by Bogza. (Bogza et al. 2015).

- The structural information is separated from descriptive information (attributes) for reasons of flexibility and avoidance of re-engineering in the case of a change.
- The Data Vault allows parallel loading of data.
- The data are not processed or filtered (tracing the data source).
- Data are never changed (they remain intact).
- The Data Vault structure does not allow a final exploitation of data.

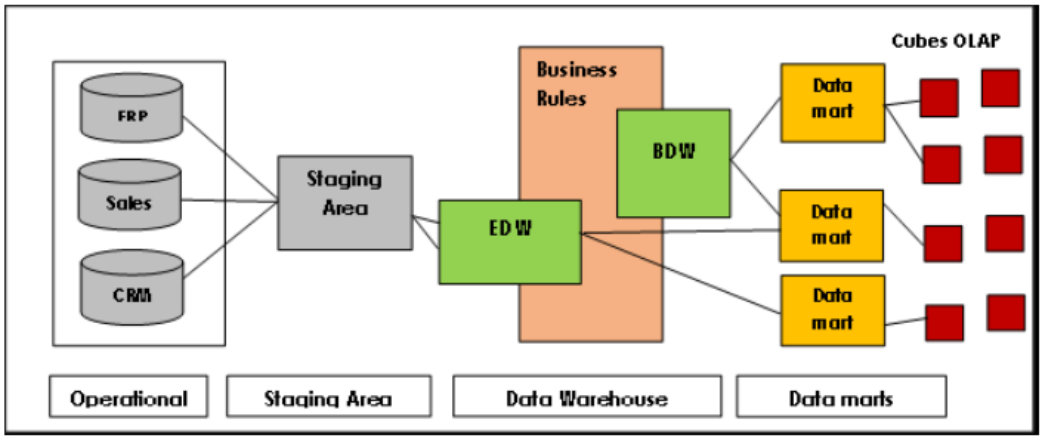


Figure 15. Data vault EDW architecture

The application of business rules in advance before the actual data loading, the extended functionality, flexibility and scalability considered to be the main advantage of data vault architecture where the complexity and the time requires to develop are the disadvantage of the architecture.

Considering the current initiatives of the bank to implement the data warehousing technology and for the above indicated advantages i.e. simplicity, initial cost, time required to build and query response time during report generation the Kimball approach is believed to be the appropriate design method selected to this study.

CHAPTER 4

4. DATA PRESENTATION AND INTERPRETATION

4.1. INTRODUCTION

This section presented and interpreted the perceived importance of application of data warehouse in the Ethiopian banking industry specifically at Dashen Bank, in terms of the data source and operational systems involved, Data integration & Report generation, Historic data incorporation and Varity and size of data.

The focus group discussion questions are designed in a way that the same questions forwarded to all the participants of the focus group to know the level of understanding of the participants on the concepts of data warehouse and its application and to understand the readiness of the bank to apply data warehouse. A total of 6 members of the database and MIS and digital banking section staff members are participated in the group discussion. The questions are categorized into three major categories as introductory, Key discussion points and conclusions in order to determine the importance of application of data warehousing technology, the readiness of the bank to apply the data warehouse technology and the gaps demonstrated in the bank on the analysis of historical data.

4.2. FOCUS GROUP DISCUSSION

In order for a focus group discussion to be most impactful, the following six step by step techniques are followed:

4.2.1. GROUP DISCUSSION PROCEDURES

Clarifying the purpose of the focus group.

- The purpose of the focus group discussion is to get a deeper understanding on the current system environment on the information being produced, data source and operational systems involved, Data integration & Report generation, Historic data incorporation, end user accessibility and Varity and size of data.

Defining the expected outcomes.

- The expected outcome of the discussion is identifying the current size data types in the data set of the bank, the number and type of operational databases that can be the data source

for the proposed data warehouse system, identify the current status of data integration and reporting, identify the integration of historical data and status of end user accessibility.

Selection of the Participants of the discussion

- The Selection of the participants was based on who is most exposed the current system.
- The size of the group is made to be between 6 which is believed to be the ideal group size
- The Invitation for the participants is made based on their willingness to patriciate in the discussion.

Manage the Atmospherics

- The meeting location selected in a way that it will be an appropriate for discussion to avoid outside interruptions and considering the convenience of the discussion group participants.
- The discussion questions are designed in order to make the duration of the meeting between 25 – 30 minutes

Develop the Questions

- The discussion questions are prepared in three major categories into as introductory, Key discussion points and conclusions. The discussion organizer is assigned as a note taker for the focus group.
- The questions are designed to be open-ended which do not lead participants to certain answers in order to have a deep understanding on the current environment focused on the why, how and what.

Conduct the Session

- During the discussion the scope, purpose and desired outcome of the focus group is clearly stated to the participants.
- The discussion is made in such a way that much of the discussion time to be on the key discussion topics of the questions.

Analyze the Results

- The discussion group Reviewed the minutes taken by the organizer and reach on a consensus on the contents of the discussion and data gained during the discussion.

- A summary report of the key findings of the discussion points are developed and organized in a table form for interpretation.

During the discussion all participants of agreed on the importance of application of data warehouse in the Ethiopian banking industry this shows that the need for the application of data warehouse in the bank is undeniable. Moreover, the responses on the current system of the Bank is summarized, presented and interpreted accordingly with the above indicated themes. The perceived importance of application of data warehouse in the Ethiopian banking industry was presented based on the themes of information being produced, data source and operational systems involved, Data integration & Report generation, Historic data incorporation, end user accessibility and Varity and size of data. The most and least applied operational systems as well as the magnitude of data increase rate of the operational databases are also discussed. Then findings of inclusion of historic data is discussed between the different operational data source systems. (APENDEX C)

The process of moving an organization from different collections of data to a position where many members of the organization can query the data to make sound decisions requires careful planning and cooperation across the organization, In this regard the readiness of the bank for the application of data warehousing is demonstrated by the initiative to organize Database and MIS as a department under director of Application development and support division and implement a report generating system using some operational databases, nerveless although more than 105 TB of data stored from different systems since the bank is established these historical data of the bank is not used, and also out of many operational databases currently only two are integrated with the reporting system.

4.3. INFORMATION BEING PRODUCED

As per the participants of the discussion currently the bank is integrating few of the operational databases to satisfy the increasing report need of the Bank. although the current system has the capability to work as a data warehouse if carefully designed which include ETL, its own database and an oracle BIP, it has been using for report generating purpose only. As per the participants of the discussion and observation made most of the information(reports) generated by the system is financial like GL performance of the branches, new Account opened, deposit made and cash withdrawn which helps to evaluate the day to day, weekly and monthly performance of the branches.

Regarding the banks current customers, Branches, POS and ATM terminals, although there is a Online transactional processing system which is capable of managing and storing the customers, transactions and terminals day to day involvement in the banks process, there are few or no analysis reports generated by the system to know the merchant performance, ATM and POS terminal performance, geographic profitability of terminals and branches.

4.4. END USER ACCESSIBILITY

As stated by Gray and Watson and described by Katherine Liddell Avery (Avery, 2003) People in a variety of organizational positions are potential users of a data warehouse. The positions include Executives, the senior managers of the organization. Managers, range from middle to lower management. Analysts, who analyze data and present their analyses to others, Operational personnel, who perform operational tasks throughout the organization and typically use custom-built or packaged applications, Customers and suppliers, as warehouses have been opened up to customers and suppliers, they have become the most recent users of warehouse data. They normally access data through web-based applications

The Dashen bank reporting system provides its service to branches, departments, CEOs and CIOs based on the requirement collected from each branch and department using requirement collection survey, To help these end users access the report the system includes oracle business intelligence publisher (BIP) reporting tool which provide three layers' architecture, Administrative layer for system administrators, Report authoring layer to design the required reports and presentation layer for end user to access more than 1700 daily, weekly and monthly reports designed in Report authoring layer in different formats supported by Oracle BIP which includes Microsoft excel, PDF, HTML and CSV.

Although most of the end users of the system are internal (direct system users), as a financial organization there are also external local and international organization (indirect users) which request reports on the operation and performance of the bank. Hence these local and international organizations are also considered as indirect users of the system and needs to take their requirements into consideration during design. And the user list is presented as internal and external as follows.

- Internal users include CEO's, CIO's and Departments and Branches
- External users include (indirect users) NBE and International networks

4.5. THE DATA SOURCE AND OPERATIONAL SYSTEMS INVOLVED

As per the participants and is confirmed during the interview with the Manager currently there are more than 30 operational databases excluding the CBS and OLTP/SWITCH in the different departments (APENDEX B) out of these databases only two of them are integrated in the system which lead to lack of analysis reports on the customer, ATM and POS terminal health and performance and foreign currency generating analysis.

According to the Database and MIS Manager, CBS become the most involved system because all transaction which comes through the switch, Mobile banking and internet banking should eventually end up on CBS for transaction authorization; hence all data regarding account, transaction and customer will be captured at CBS. However, ATM terminal performance with regard to the geographic location cannot be analyzed since switch data is not integrated and once the customer transfers from CBS Account to wallet account in Mobile and internet banking the behavior of the customer cannot be analyzed since the effect of wallet account is not reflected on the CBS.

Since only two databases are currently involved, the current system cannot satisfy all the business needs of the bank. Generally speaking, since the current system is satisfying only the reporting needs of the bank and not the banks business requirements needs which instead demands the analytics part of the data warehousing functionality to predict future business trends, the current system is practically failed short to provide the full services provided by a data warehouse.

According to (Gallo, 2010) there are three categories of data warehouse application, individual productivity, data query and reporting and Planning and analysis which are categorized as per their scope and complexity.

Individual productive are used to process and present data on a user's PC and are usually developed in an independent environment and access and manipulate limited volumes of data. Data query and reporting used for limited complexity queries and reports on historical or recent enough data. Whereas Planning and analysis are used for complex analysis of historical data and for planning and forecasting the future based on historical data. The result is the planning and forecasting of future events, simulations, assessments of processes and opportunities.

Since the bank is not currently using complex analysis tools to analyze and forecast its business future, the current effort of data integration of the Bank falls in the category of data query and reporting. Although the bank is on the way to include some other system in the integration process, the ETL and analysis tool applied will also be the other challenge of the bank since these tools are used for limited complexity queries.

4.6. REPORT GENERATION

During the discussion the participants described that report requirement was gathered from the different department and branches using survey through the portal of the bank focusing on the core banking and with this although the bank is able to satisfy its report needs, though it satisfies the report needs, it is not able to satisfy its business requirement needs with the current implementation since the business requirements of the bank are much more than report generating, which require data analytic.

The report creation and generation is done using Oracle Business intelligence publisher (BIP) reporting tool which have a three-layer capability, i.e. administrative layer for administrators, Report Authoring layer for report designers and presentation layer for end users to access the report. Oracle BIP reporting tool also have the capability to produce the reports in different formats as the need of the end user using Microsoft excel, PDF, CSV or HTML formats which are scheduled on daily, weekly, Monthly, quarterly and annually.

4.7. HISTORIC DATA INCORPORATION

Historical data, in a broad context, is collected data about past events and circumstances referring to a particular subject. Historical data includes data generated either manually or automatically within an enterprise. Sources of historical data include log files, reports, product documentation and others.

Since the Bank started in 1996 E.C different type of historic data is accumulated in the banks backup storage area, as per the datacenter engineer currently this accumulated data exceeds 120TB with a diverse format like structured data from relational database backup of core banking, unstructured data from the flat file system switch and pictures of ATM customers taken during cash disbursement at the terminal location.

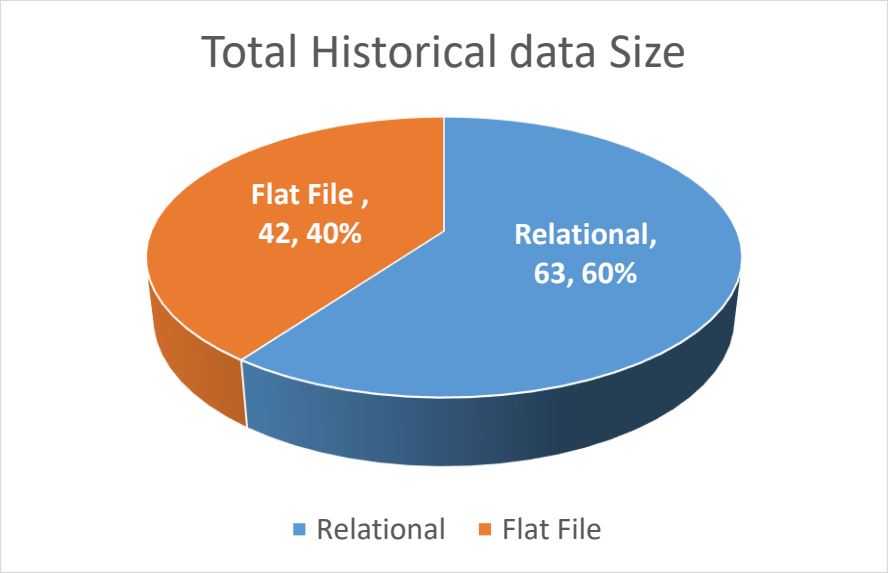


Figure 14. Graphical representation of Historical Data Accumulation

As it is trying to satisfy the current report need of the bank and only operational databases are involved and it is not using complex analysis tools to analyze and forecast its business future, and as described earlier the data warehouse implementation status of the Bank is at the level of data query and reporting according to Gallo (Gallo, 2010). Moreover, although more than 120 TB of data is stored in the backup storage from different systems these historical data of the Bank is not analyzed.

4.8. DATA SIZE ANALYSIS

In order to determine the most prominent systems to collect high volume of data from the current system, according to the interview with the MIS manager about the size of data collected by each system out of the 30 most of the systems were implemented within the past two years and the data collection status of the systems is insignificant as compared to the top five which are

- Online transaction processing systems which collects data about the transactions, terminal, products and merchants.
- Digital Banking system which collects data about every transaction, account and customer who used on the digital channels.
- Core banking systems which collects data about transactions, accounts and customers.
- Dispute management system which collects data about disputed transactions and their evidences

- Card authorization and processing system which collects data about customers who are registered for acquiring ATM card.

	STORAGE	TOTAL SIZE
1	OLTP	42 TB
2	DIGITAL BANKING	23 TB
3	CBS	13.5 TB
4	DMS	10.5 TB
5	CATPS	10 TB
6	OTHERS	11.5 TB
	TOTAL	**110.5**

Table 4. Data accumulation status of the top 5 systems in the Bank

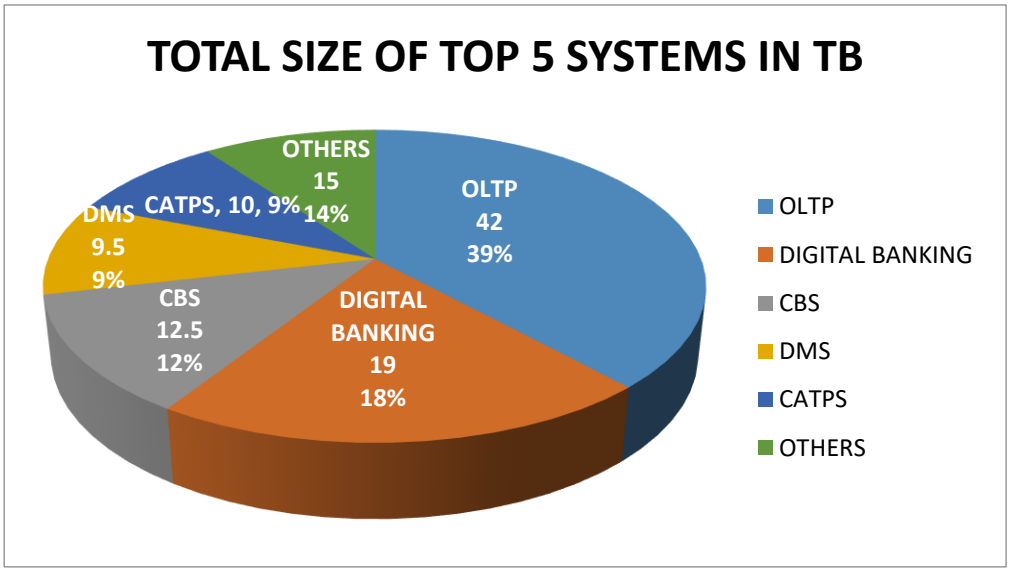


Figure 15 Data accumulation status chart in the Bank

As we can observe from the above data accumulation information in Banking Business beside the security control systems there are three Main systems which are common to most banking business to carrying out the day to day operation of the Bank this day are the core banking system and online transaction processing system working in front of the core banking system and Mobile and internet banking system. These three systems enable the bank to acquire, authenticate, route, switch, and authorize financial transactions across multiple channels. Hence the three systems are able to collect and store each and every transactional as well as customer data.

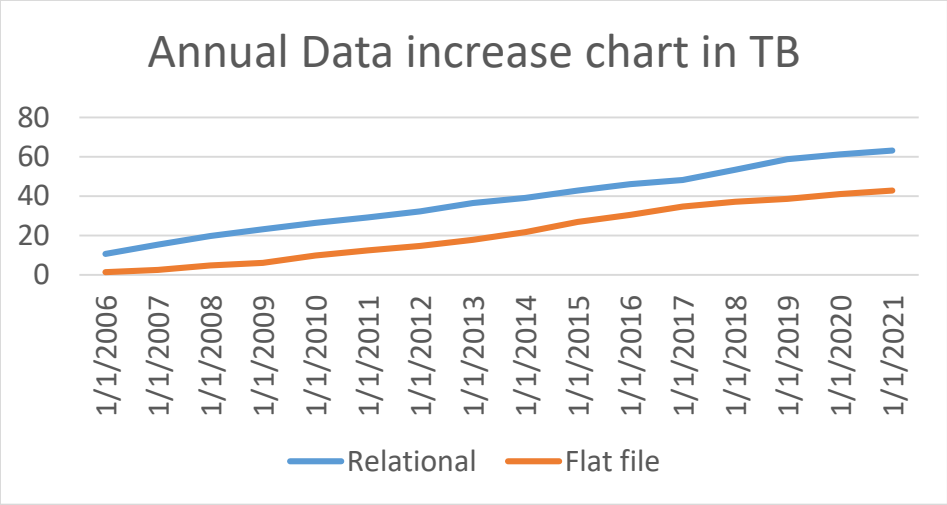


Figure 16. Graphical representation of Annual data Increase rate

Since the data to be extracted from the dispute management system and card processing system for the data warehouse can be found from the core banking and the online transaction processing system, only the top three systems are believed to be the most relevant systems for this design which controls the majority of historical as well as operational data concentration in the Bank.

4.9. GENERAL SYSTEM ARCHITECTURE OVERVIEW

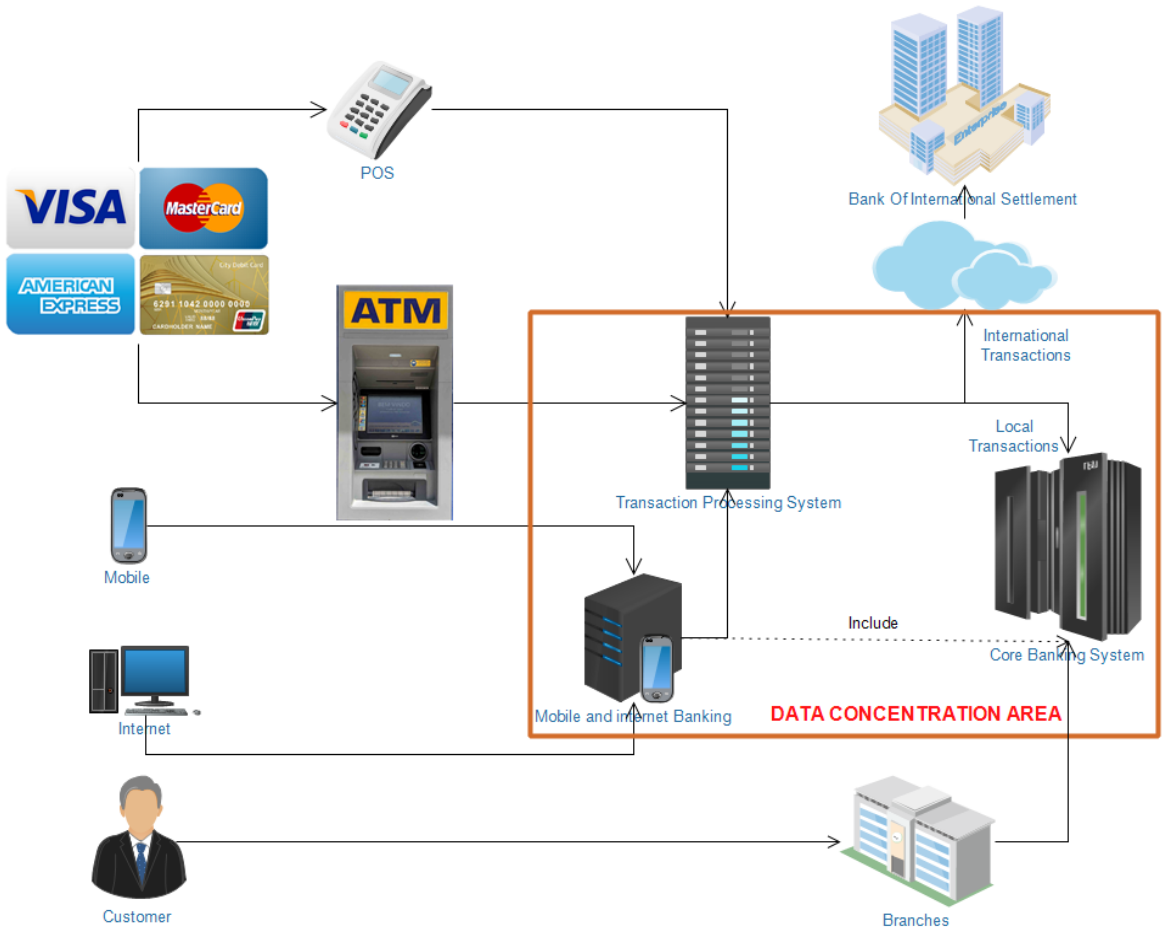
Core Banking System is the software used to sustain banks, most common dealings which include providing service loans, opening new accounts, processing cash deposit, withdrawals, calculating interests, client relationship management actions and maintenance of records for the banks transactions (Hariharan 2015). It is a complex system for commercial banks and financial institutions. It provides administration of clients, accounts, credits and deposits, accounting and daily closings, i.e. it covers all day-to-day processes of financial institution.

The online transaction processing system platform is used to acquire, authenticate, route, switch, and authorize financial transactions across multiple channels providing the bank with robust high-performance platform. Besides acting as a payment switch it is also used for managing card products as well as for POS and ATM terminals, running on high availability Non-stop servers.

The Mobile and Internet Banking is browser based application provided to bank's customers to access banking information and make transactions. The application is optimized for desktop, laptop and tablet platforms too, and both mouse and touch screen navigation. Mobile channel on the other hand use Smartphone application providing similar functionality to internet Banking but

in form of installed native application. The application is optimized for Smartphone with additional functionalities like branch and ATM locations features whereby customers can be navigated to nearby ATM and branches.

Figure 17. Diagrammatic representation of current system data concentration



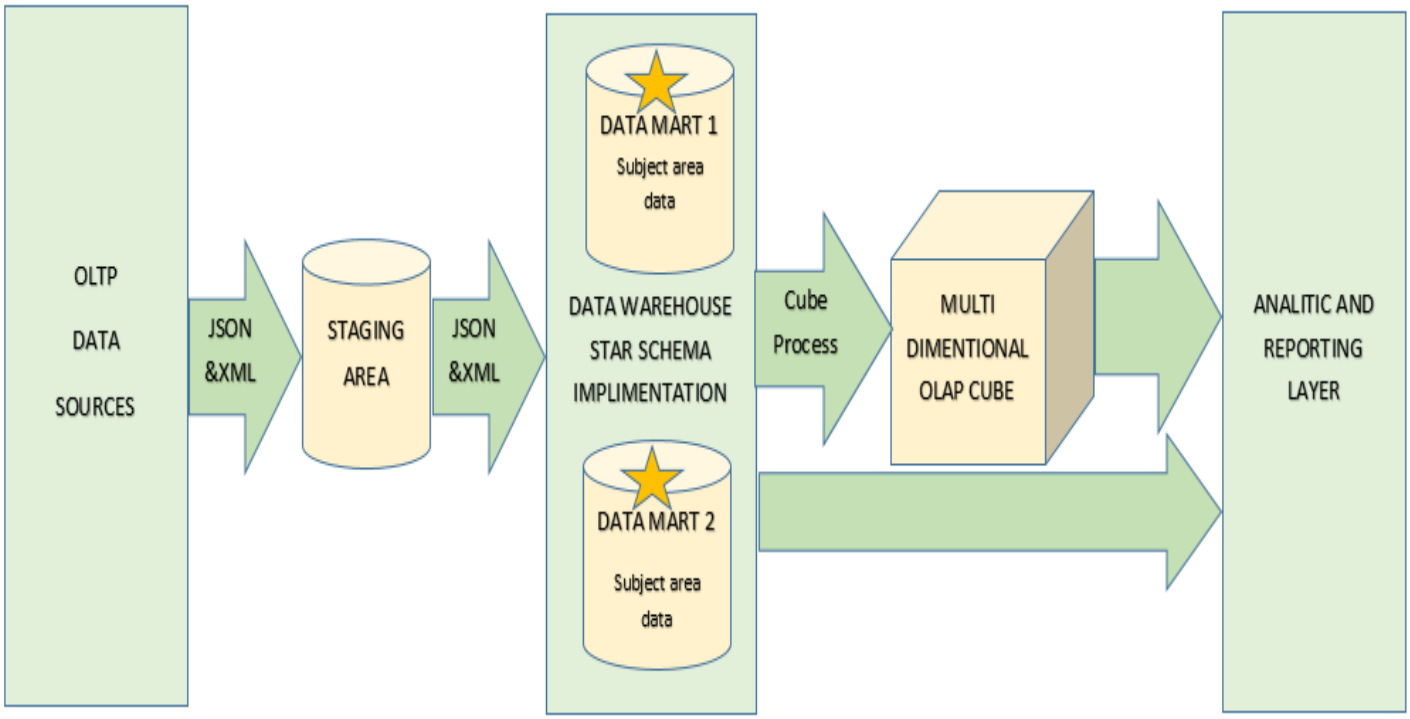


Figure 18. Diagrammatic representation of data warehouse architecture

CHAPTER 5

5. PROPOSED SYSTEM DESIGN

5.1. DATA VARITY IN THE BANKING

Designing and implementing an enterprise data warehouse means that we need to be aware of the different databases and how we might best integrate them with the data warehouse. Bank's main business processes are all aimed at supporting the five key transaction initiating channels; Branch, Internet, Mobile, ATM and POS. (Kaput 2018). All transactions performed on the indicated terminals passes through these channels or finally touch one of the three systems indicated in section 4.8. Since different type of date inflows through these channels the variety of data that is accumulating in the banks data storage is diverse.

When it comes to data storage, there is almost as much diversity in the types of databases as there is in the data that they contain. Most of the data concentration is on the three systems of the bank where two of them implement the same relational database model whereas the transaction processing system uses flat file system data. Flat files are collection of information stored and accessed sequentially in a database, typically consist of text files with no markup, representing relational data by separating it with a comma or other delimiter. Flat file is a single lined plain text file containing delimiters comma, tabs, tube (|) and many other characters. It is also called mixed text file. The main advantage of the flat file is that it takes less space as compare to the structured file. (Inayat 2011)

5.2. SOURCE OPERATIONAL DATABASES

Identifying data sources and their associated origins allows developers to predict specific sources of error and emphasize certain areas that should have greater scrutiny. (Emilio et. al 2008)

Operational databases are designed to run the day to day operations or transactions of banks. Such databases are used to manage and store data in real time and can also be used as a source for a data warehouse. Since they are processing data in real time and a huge amount of data is stored in these databases online analytical processing cannot directly be performed on them.

As illustrated in the above diagram the data concentration is on the core banking system, Mobile and internet Banking system and the online transaction processing system databases which contains a flat file system are the three operational system databases would be an ideal source of

data for the intended data warehouse since most of the historical as well as operational data accumulation is with these three system.

5.3. ISSUES INETL PROCESSES

According to (Gallo et. el 2010) the process of data flow within and out of the data warehouse is classified as Inflow, Up-flow, Down-flow and Down & Out. As per the direction of the data flow. Data are collected inside the transactional systems and are sent to the DW. This process is called Inflow. Data in a transactional system are normally at a detail level. Part of it is aggregated and summarized to give users a faster response time. Data thus processed is transferred into the DW and this process is called Up-flow. Data can be made available from a DW to end users which, through query and analysis tools, can receive an output data stream from the DW. This process is called Outflow. Data stored in a DW can be re-stored in it with a format of further gathering before being removed. This happens when data are old and are no longer used with sufficient frequency to justify their presence. This process is called Down-flow.

The original data of a Down-flow process are removed from the DW and transferred to external storage media. Data are not simply deleted and then lost forever. If that were true, we would lose the historical memory of the company in contravention of the basic principles of data warehousing. This process is called Down & Out.

A well designed ETL process provide Information clarity during the extraction process by avoiding unclear and ambiguous raw data, Information completeness by including all the business sources which are relevant to the operation and Information quality by validating data at extraction or correct/discard data at transformation, in general ETL process brings structure to the organizations information. There are always a number of ETL issues that arise when we bring data from multiple systems on multiple platforms. One of the most pressing of these issues is how to standardize (conform) data elements. (Emilio et. al 2008). And the other is how to choose a well-designed ETL tool that can connect and fetch the data from the sources with different data format.

As technology and data integration processes have evolved, different types of ETL solutions have entered the market. Some are designed to work in an on-premises data environment, some in the cloud, and some in hybrid environments.

- Batch processing or legacy ETL tools: These tools extract, transform and load the data into the target data warehouse or data lake in batches of ETL jobs. Until recently, these batch processing tools were a popular, practical, and cost-effective way to do ETL since they use limited resources for a limited time.
- Real-time ETL tools most organizations require real-time access to data from different sources. Cases like this have resulted in a shift to real-time ETL tools. Apache Kafka has emerged as the leading platform for processing streaming data in real time, and organizations are building modern ETL solutions on top of it, either as a SAAS platform or an on-premises solution.
- Open-source ETL tools: These ETL tools have source code freely available and therefore help organizations keep their costs low while providing similar functionalities as other ETL tools. Most open-source ETL tools were created as a modern management layer for scheduled workflows and batch processes.
- Cloud-native ETL tools: With more enterprise businesses moving to the cloud, they need a way to extract, transform and load data from sources directly into a cloud data warehouse. Cloud-native ETL tools let organizations gain key cloud benefits such as flexibility and agility in the ETL process.

On the other hand, Pall and his colleague's categorized ETL tools into two broad categories Hand coded ETL tools and Tool Based ETLs. Hand-Coded ETL tools are tools that are in-house developed to extract data from multiple source files, transform the data, and load to the target databases. The programs written using this method were lengthy and hard to document. Hand-Coded ETL tools have the advantage that the metadata created can be managed directly and they give the flexibility to the developer to manipulate to new needs and unit testing is much easier. However, it has also limitations to accommodate the continuous changes in the high volumes of data generated through various sources the programs need to be modified frequently which causes a burden on the overall project. Since, the hand-coded tools involve overheads and are slow in execution hence many vendors developed Tool Based ETL tools to be purchased by the organizations. These ETL tools started from simple extractions on mainframes to target database and now-a-days they are available in full GUI's with added functionalities and performances. These are the ETL tools of today that provide transformation features, support multiple input or output database or flat files, multi-dimensional designs, surrogate key generation, various transformation functions and native database or O/S utility. (Pall et. el 2013)

Moreover, Pall and his colleagues provide a list of criteria to consider, when organizations think of acquiring an ETL tool some of them are listed below

- Platforms: These criteria signify how many platforms are supported by the ETL product e.g Windows (all versions have been counted as one), Linux, Solaris etc.
- Ease-of-Use: Ease of use includes how easy is it to use the product, how quickly can it be learnt, number of training days required for the developer and the user to learn the product, screen element designs and GUI interface.
- Debugging: does it provide a good debugging facility either step by step or row by row?
- Separate Modules: Usually the tool is made up of at least two modules the real time module and the batch module. Now can they be bought separately?
- Data Mechanism: The data changes when it's extracted and transformed. So the question is how is it recognized i.e. how is the changed data recognized.
- Native connections: How much and which native connections does the ETL tool support? (ODBC, OLE DB and flat files excluded).
- Scheduler: whether or not there is an ability in the tool to schedule jobs based on interdependencies. Or in other words is the scheduler capable of handling dependencies.

5.4. ISSUES WITH FLAT FILE DATABASES

Flat file is a file that does not contain any internal hierarchy. Initially developed by IBM in the early 1970s, a flat file database is used to store unstructured data in plain text format. A flat file database is a database stored in a file called a flat file. Individual lines of the text file hold one record with additional data fields separated by delimiters. These delimiters can be commas or tabs to help you define a structured format. Most database programs can, however, import flat file databases so that they can be used in a relational database.

Flat file are text files stored in computer. Data in flat file is simple and can ported to any program. The basic characteristics of a flat file are that data are stored as plain text, even the number are plain text, and that each line of the file contains one record or case in the data set. Each line a flat file, several contain the values for the different variables in the data set. Fields within a record are separated by a special character, or delimiter. Each line after the header consists of two fields separated by a colon (the character ":" is the delimiter). Alternatively, we can used "white space" (one or more space tabs) as the delimiter. (Mohd et. al 2018)

It’s common for organizations to move large volumes of data such as customer information and product catalogues into a database. A flat file is used for importing data into a business system, most commonly during the synchronization of a large volume of data is needed. Despite there being no data manipulation on the data that is stored in the flat file, it can carry relevant information from the server to the desired database. It’s simple structure means that it uses minimal space compared to more structured files. The data within a flat file maintains its original form until it is transferred into a staging database within a business system. Once the data is in its staging area, it is then transformed into its assigned field(s).

Since most of the system now a day’s using relational databases the most common questions in the extraction process of a data warehouse is how to extract data from a flat file system. According to Ye s., Sun w. and Tian z the flat file architecture by itself is designed to make such data movement simple. According to the architecture, flat file architecture (FFA) is invented to facilitate the transformation from flat file to XML and vice versa. FFA includes three core function components, Flat File Reader (FFR), Flat File Writer (FFW), and Extension Controller. According to the predefined data extraction and transformation rule, the FFR could transform flat file to XML. The FFW could transform XML file to flat file. (Ye et. al 2004).

5.5. SELECTING THE BUSINESS PROCESS

As indicated in the previous chapter this study follows Kimball’s data warehouse design model. According to Kimball data warehouse development model, data warehouse development passes through four key steps of selecting the business process, declaring the grain, identifying the dimensions and identifying the facts.

Business processes are the operational activities performed by an organization; Business process events generate or capture performance metrics that translate into facts in a fact table. The Bank business processes are categorized in four broad categories as card processing, Customer handling, Loan processing, and Risk and compliance.

These business processes of the Bank will allow the bank to make analytics on Terminal performance, Product performance, Customer performance and Risk assessment and loss prevention

5.6. DECLARING THE GRAIN

According to Ralph Kimball's dimensional data modeling fact tables are categorized into three based on the grains they contain as Transactional fact tables, Periodic snapshot fact tables and Accumulating Snapshot Fact tables

Transactional fact tables

The transaction fact tables represent an event that occurs at point in time. A line exists in the fact table for the customer or product when the transaction occurs. Many rows in a fact table can connect to a customer or product because they are involved in multiple transactions.

Periodic snapshot fact tables

A row in a periodic snapshot fact table captures some sort of periodic data, for instance, a daily summary snapshot of financial metrics, or a weekly summary of accounts receivable, or a monthly inventory numbers.

Accumulating Snapshot Fact tables

This type of fact table is used to show the activity of a process that has a well-defined beginning and end. For example, the processing of an order. An order moves through specific steps until it is fully processed. As steps towards fulfilling the order are completed, the associated row in the fact table is updated.

CHARACTERISTIC	TRANSACTION GRAIN	PERIODIC SNAPSHOT GRAIN	ACCUMULATING SNAPSHOT GRAIN
Time period represented	Point in time	Regular, predictable intervals	Indeterminate timespan, typically short-lived
Grain	One row per-transaction event	One row per period	One row per life
Fact table loads	Insert	Insert	Insert and update
Fact row updates	Not revisited	Not revisited	Revisited whenever activity
Date dimensions	Transaction date	End of period date	Multiple dates for standard milestones
Facts	Transaction activity	Performance for predefined time interval	Performance over finite lifetime

Table 5. Characteristics of grains

As the study aims show the application of data warehouse for data analysis that show the periodic financial performance of the bank, the facts table of this study will follow the type of fact table which shows periodic summary snapshot of events.

GRAINS
Sum of Number of transactions and amounts per customer
Sum of Number of transactions and amounts per account status
Sum of Number of Accounts and their of current Loan amount
Sum of Number and amount of each transaction type per terminal
Sum of Number and amount of transaction performed at Merchants with different product types.

Table 6. List of grains in the proposed data warehouse

5.7. IDENTIFYING THE DIMENSIONS

DIMENSIONS	DESCRIPTION
Terminal dimension	POS or ATM Terminals prepared by the bank for their customer
Branch dimension	Branches legally identified by the NBE
Product dimension	Card products issued by the bank and acquired by the bank
Customer dimension	Individuals and merchants registered by the bank as customer
Transaction Type dimension	The detail data of the transaction performed by the banks customer on the banks infrastructure.
Account dimension	The detail of Account opened by the banks to their customers.
Merchant dimension	Detail of customers registered by the bank as Merchant.
Date dimension	A duration of time used to perform any functionality in the bank
Channel dimension	The kind of channels used to perform transactions(Terminals, Internet & Mobile banking)

Table 7. List conformed dimensional tables of the proposed data warehouse

5.8. IDENTIFYING THE FACTS.

A fact table is a primary table in dimension modeling, the table is the center of the star schema which contains the business metrics (i.e., numerical measurements) and foreign key to dimension table. Fact tables are the largest tables (in number of rows) in the star schema design. As per Kimball's multi-dimensional data model design a fact table contains a numeric measures produced

by an operational measurement event in the real world. At the lowest grain, a fact table row corresponds to a measurement event and vice versa.

The below table shows the six different Fact tables and their numeric measures the table contains during the Load process of ETL process.

FACTS	MEASURES OF FACT TABLE
Transaction Fact	Total number of daily transactions and amount of each customers
Account Fact	Total number and amount of the different type of account status at each branch
Loan Fact	The total daily count and amount of each type of loans provided by Branches
ATM Terminal Fact	Total number and amount of transactions performed on each terminal per day.
Merchant Fact	Total daily count and amount of transactions performed at Merchants with different product types.
POS Terminal Fact	Total number and amount of transactions performed on each terminal per day.

Table 8. Facts and their Measures of the proposed data warehouse

FACT DETAIL					
ACCOUNT FACT	**TRANSACTION FACT**	**ATM TERMINAL FACT**	**LOAN FACT**	**POS TERMINAL FACT**	**MERCHANT FACT**
ACCT_FACT_ID	TRAN_FACT_ID	ATM_TERM_FACTCT_ID	LOAN_FACT_ID	POS_TERM_FACT_ID	MERCH_FACT_ID
ACCT_ID	PROD_ID	TERM_ID	LOAN_ID	TERM_ID	PROD_ID
BRN_ID	CHENNEL_ID	BRN_ID	BRN_ID	MERCH_ID	MERCH_ID
DATE_ID	DATE_ID	TRAN_TYPE_ID	DATE_ID	TRAN_TYPE_ID	DATE_ID
TOTAL_ACCT_BAL	TOTAL_AMT	DATE_ID	TOTAL_LOAN_CUNT	DATE_ID	TOTAL_TRAN_CUNT
	TOTAL_COUNT	TOTAL_TRAN_COUNT	TOTAL_LOAN_AMT	TOTAL_TRAN_COUNT	TOTAL_TRAN_AMT
	TOTAL_CHRG_GAIN	TOTAL_TRAN_AMT	TOTAL_LOAN_PAID	TOTAL_TRAN_AMT	MERCH_CAT
		TOTAL_CHARGE_GAIN	TOTAL_LOAN_OUT	TOTAL_CHARGE_GAIN	PROD_GROUP
		TOTAL_CHARGE_LOST	TOTAL_INTEREST_CHG	TOTAL_CHARGE_LOST	INTEREST_GAIN

Table 9. Facts and their Attributes of the proposed data warehouse

5.9. CONCEPTUAL DESIGN

Conceptual design models information gathered from business requirements and defines around the business's need. The conceptual data model is designed by identifying the various entities or classes and the highest level relationships between them as per the given requirements.

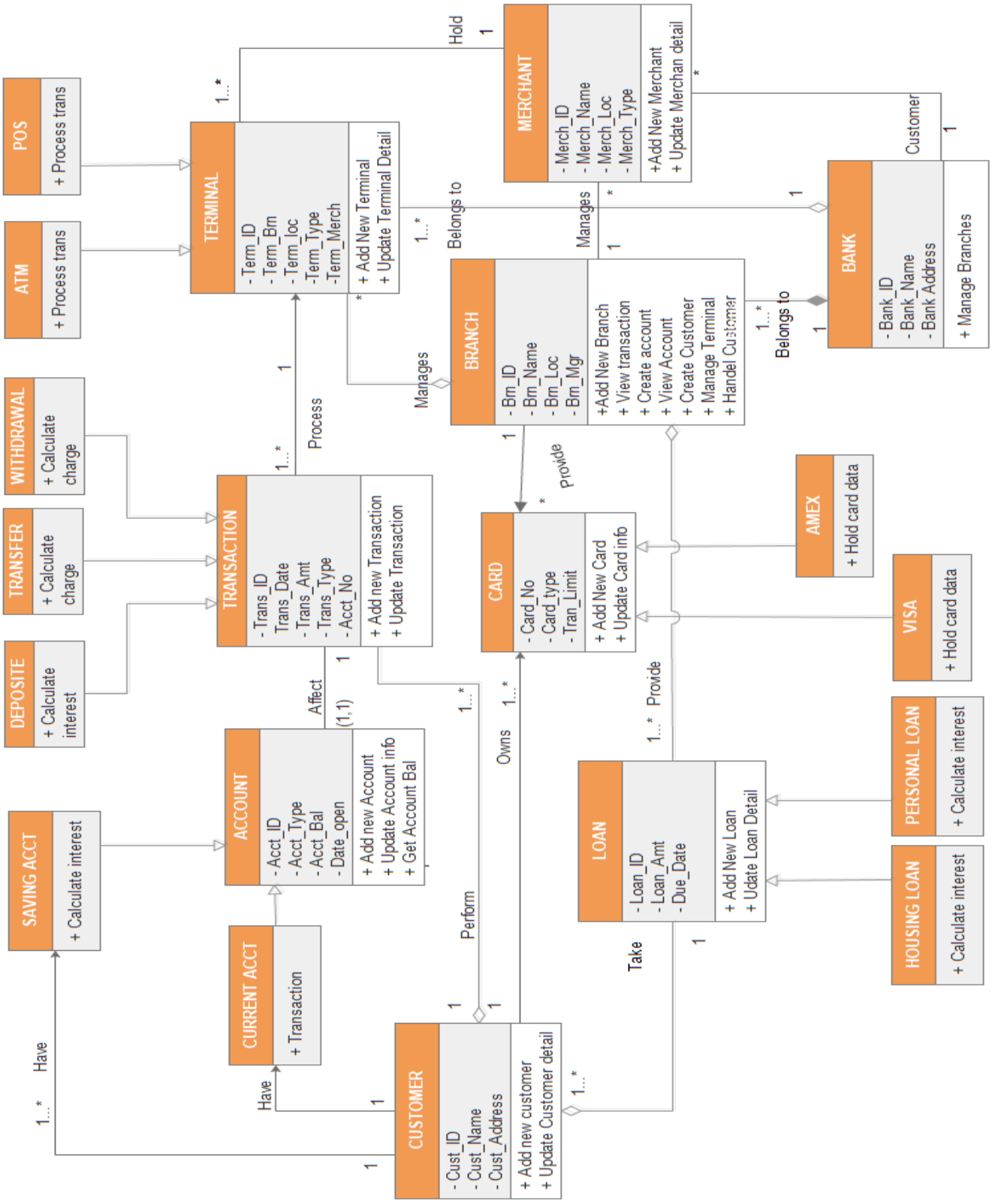
Conceptual modeling is widely recognized to be the necessary foundation for building a database that is well-documented and fully satisfies the user requirements; usually, it relies on a graphical notation that facilitates writing, understanding and managing conceptual schemata by both designers and users in the context of design, a basic role is played by conceptual modeling that provides a higher level of abstraction in describing the warehousing process and architecture in all its aspects, aimed at achieving independence of implementation issues. (Rizzi, 2007)

While it is universally recognized that a Data warehouse leans on a multidimensional model, little is said about how to carry out its conceptual design starting from the user requirements. (Golfarelli et. al, 1998)

The way to describe the conceptual design is also the other discussion point whether the traditional ERD or the Object oriented way is appropriate. While Bill Inmon believes that the traditional ERD model is appropriate to describe the conceptual model of an enterprise data warehouse whereas Ralf Kimball states that the object oriented class diagram is more appropriate to describe the conceptual model of a multi-dimensional data warehouse. As per Juan Trujillo the object Oriented approach is not restricted to using flat UML class diagrams to model large, complex data warehouse systems. UML's package grouping mechanism groups classes into higher-level units, creating different levels of abstraction and simplifying the final model. In this way, a UML class diagram improves and simplifies the system specifications created with classic semantic data models such as the Entity-Relationship model. Further, class diagram expressions can embed operations and constraints in the class diagram. (Trujillo et al. 2002)

Based on the Kimball's suggestion and from the above data availability and concentration analysis, the below conceptual design of the banks transactional process is designed using UML class diagram.

Figure 19. Class diagram of the current system process



5.9.1. DIMENSIONAL FACT MODEL (DFM)

Dimensional fact model is a graphic conceptual data model developed to Supports conceptual design for a data warehouses, Independent from the logical models by Matteo Golfarelli and his colleague Stefano Rizzi in University of Bologna. According to Golfarelli the Dimensional fact model models analytic information by a specific notation that is used in BI Applications to Create an environment where user queries can be formulated in an intuitive way which Enables communication between designer and end user helping to refine requirements specification in a better way moreover it Provides a stable platform for logical layer design.

The main element of Dimensional fact model is fact which is a focus of interest for the decision-making process; typically, it models an event occurring in the enterprise world. Measures are continuously valued (typically numerical) attributes which describe the fact from different points of view; for instance, each sale is measured by its revenue. Dimensions are discrete attributes which determine the minimum granularity adopted to represent facts. Hierarchies are made up of discrete dimension attributes linked by a-to-one relationships, and determine how facts may be aggregated and selected significantly for the decision-making process. (Golfarelli et. al 1998)

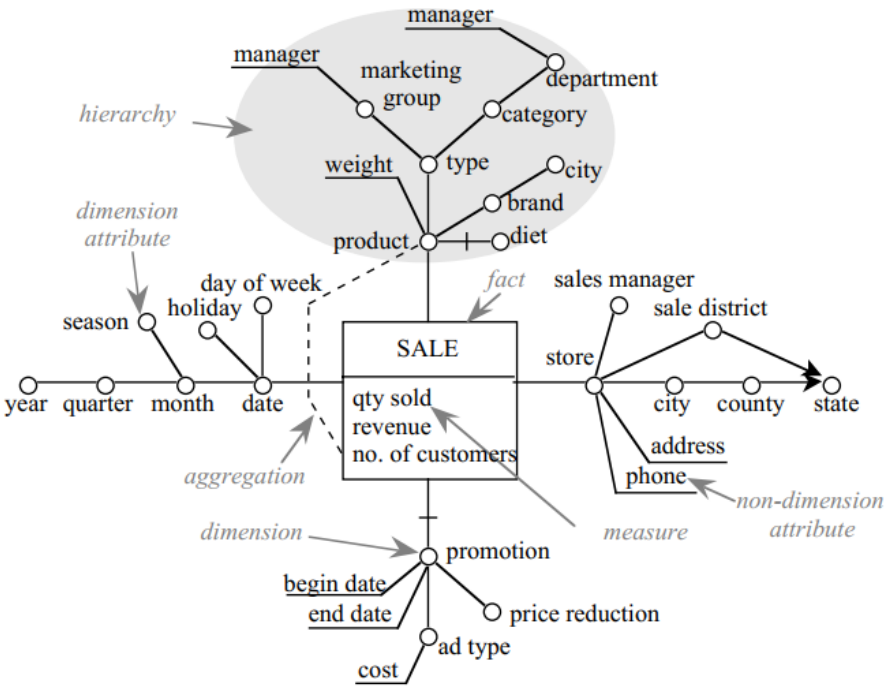


Figure 20. Golfarelli's Fact schema for Sales Fact

In the DFM, a fact scheme is structured as a quasi-tree whose root is a fact. A fact is represented by a box which reports the fact name and, typically, one or more measures. Dimension attributes are represented by circles. Each dimension attribute directly attached to the fact is a dimension. Non-dimension attributes are always terminal within the quasi-tree, and are represented by lines. Sub-trees rooted in dimensions are hierarchies. They are connecting two attributes represents a -to-one relationship between them. (Golfarelli et. al 1998)

5.10. GOLFARELLI FACT REPRESENTATION

5.10.1. TRANSACTION FACT

Transaction Fact will allow the Bank to analyze the total count and amount of transactions performed by each product through each channel for the specified period of time. Moreover, this fact will also allow the calculation of charges gained by specific product and terminal. Hence these helps the bank to evaluate the performance of each product issued and acquired and makes decisions on the profitability of the product.

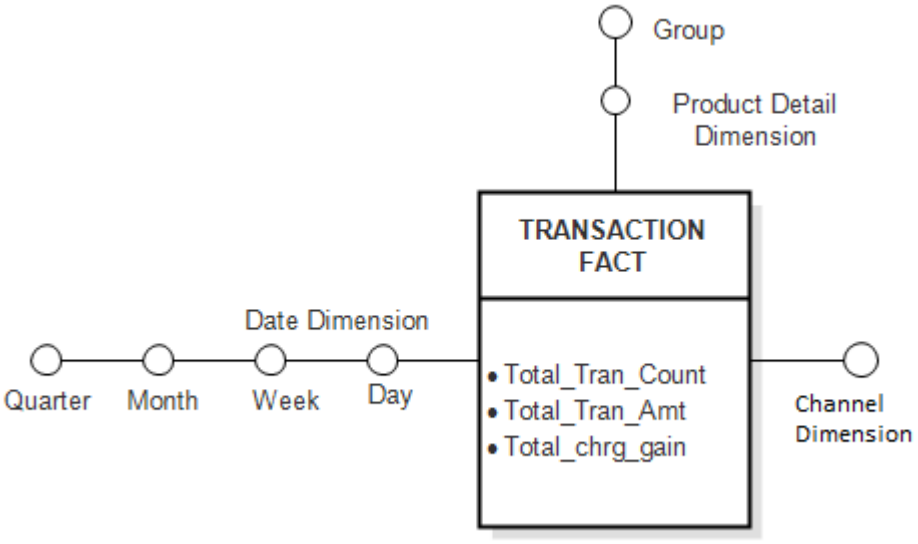


Figure 21. Simplified Transaction Fact schema representation

5.10.2. ACCOUNT FACT

Account Fact helps the bank to evaluate the performance of its branches by identifying the number of accounts which are Active, Dormant, Freeze and Blocked with their respective total amounts. The analysis helps in deciding whether the branch is gaining new customers, retaining current customers or losing its customers within a specified period of time.

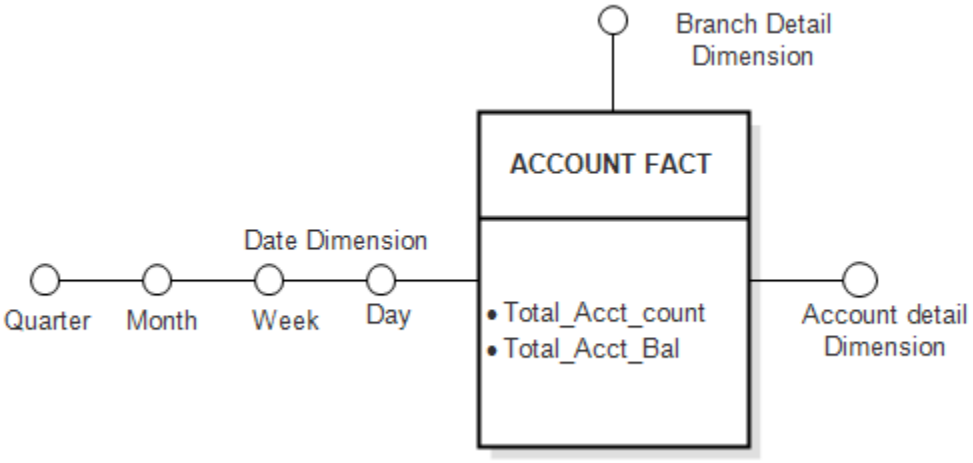


Figure 22. Simplified Account Fact schema representation

5.10.3. LOAN FACT

Loan Fact helps the bank to evaluate the performance of its loan repayment from each branches by identifying the total number loans, repayments and interest gained from each branch based on the type of loan provided. The analysis helps in deciding whether the branch is following the loans it provided to its customer and collect the repayments timely within a specified period of time.

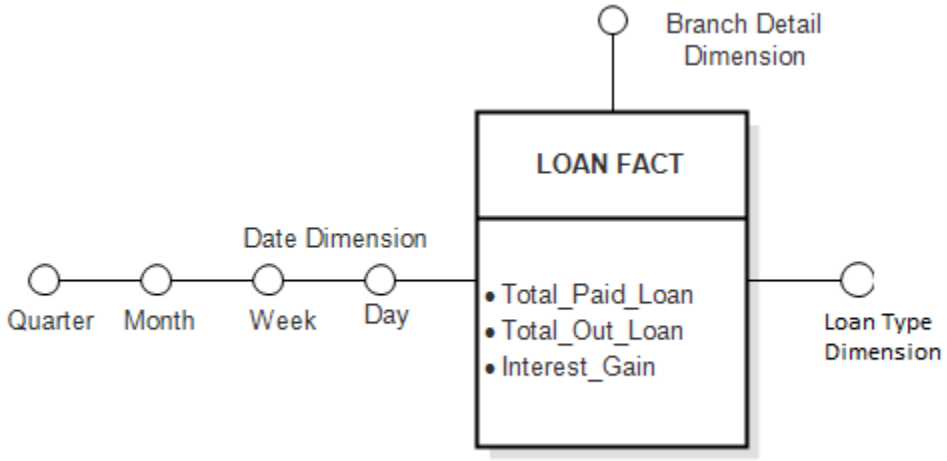


Figure 23. Simplified Loan Fact schema representation

5.10.4. TERMINAL FACT

ATM Terminal Fact helps the bank to evaluate the ATM performance as well as terminal health by identifying the total number of transactions and amounts performed on the specific terminal based on the status of each transaction as Approved, declined, reversal and suspect. The analysis helps in deciding the location profitability of the terminal and maintenance cost of each terminal within a specified period of time. The fact also calculates the total transaction charge gain from the terminal and the loss of potential transaction charge gain from the terminal.

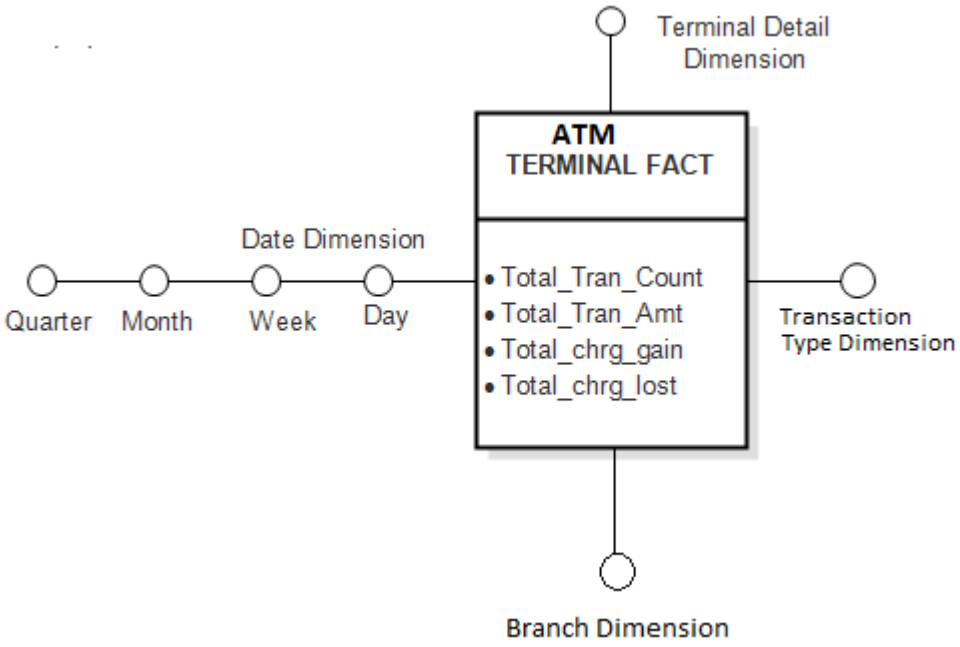


Figure 24. Simplified ATM Terminal Fact schema representation

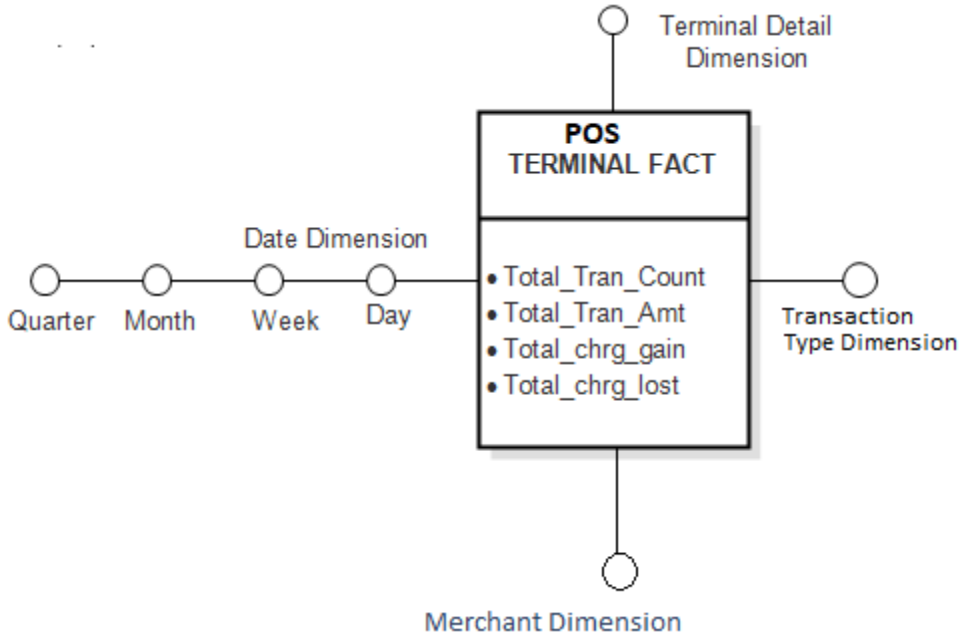


Figure 25. Simplified POS Terminal Fact schema representation

5.10.5. MERCHANT FACT

Merchant Fact helps the bank to identify total number of counts and amounts of transactions performed at each merchant location with its POS terminals based on merchant category, product group and product type. This fact can be used to analyze its merchant performance, in terms of merchant category, product type and product group.

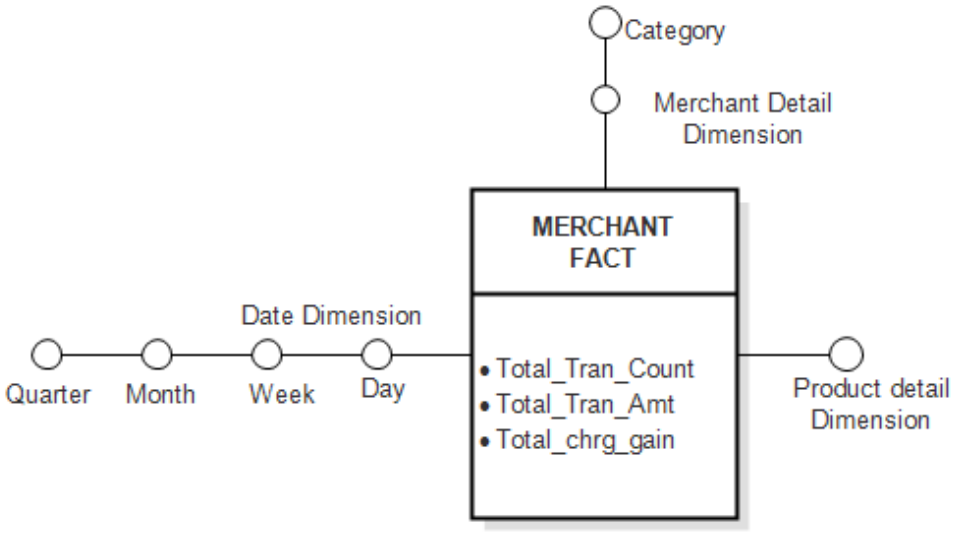


Figure 26. Simplified Merchant Fact schema representation

5.10.6. DIMENSIONS

A dimension table or dimension entity is a table or entity in a star or snowflake schema that stores details about the facts in terms of columns and attributes, which are used to describe business processes. Dimension tables also describe the different aspects of a business process and groups the data in the database when the business creates reports.

TERMINAL DETAIL DIMENSION	BRANCH DETAIL DIMENSION	PRODUCT DETAIL DIMENSION	TRANSACTION TYPE	CHANNEL
TERM_DIM_ID	BRN_DIM_ID	PROD_DIM_ID	TRAN_TYPE_ID	CHANNEL_ID
TERM_ID	BRN_ID	PROD_ID	TRAN_TYPE_NAME	CHANNEL_NAME
TERM_NAME	BRN_NAME	PROD_GROUP	TRAN_TYPE_DESC	CHANNEL_DESC
TERM_LOCATION	BRN_LOCATION			
MERCH_ID	BRN_MGR			
TERM_TYPE	BRN_DESC			
TERM_VENDOR	BRN_STAT_CODE			
LAST_USAGE_DATE	BRN_GL			
REC_INS_DATE				

LOAN TYPE	ACCOUNT DETAIL DIMENSION	MERCHANT DETAIL DIMENSION	DATE DIMENSION
LOAN_TYPE_ID	ACCT_DIM_ID	MERCH_DIM_ID	DATE_DIM_ID
LOAN_TYPE_NAME	ACCT_ID	MERCH_ID	DAY
LOAN_TYPE_DESC	CUST_ID	MERCH_NAME	WEEK
	ACCT_OPEN	MERCH_LOC	MONTH
	ACCT_BAL	MERCH _PHONE	QUARTER
	ACCT_LIMIT	MERCH _MAIL	YEAR
	ACCT_TYPE	MERCH _REG	
	ACCT_STATUS	MERCH _SUBCITY	
		MERCH _WOREDA	
		MERCH _HNO	

Table 10. Dimension and their Attributes of the proposed data warehouse

Product ID attributes Identifies whether the product is one of the listed issued or acquired products

- Visa local (VL) or Visa international (VI)
- Amex green local (AGN), Amex gold local (AGO) or Amex international (AI)
- MasterCard international (mc)
- China Union pay (CUP) and Interest Free Banking (IFB)

5.11. LOGICAL DESIGN

Conceptual models are useful to design data warehouse since it is used as the communication between the stakeholders in a project. However, conceptual models must be translated into logical ones for their implementation on a database management system. A Logical Data Model is a representation of the business concepts described in the conceptual model laid out in a visual format that clearly shows these concepts and their various relationships. Logical Data Models are independent of the underlying database implementation.

According to Kimball in his Data Warehouse Toolkit, designing a fact table with too many dimensions leads to significantly increased fact table disk space requirements. Although de-normalized dimension tables consume extra space, fact table space consumption is a concern because it is the largest table by orders of magnitude. Most business processes can be represented with less than 20 dimensions in the fact table. If a design has 25 or more dimensions, you should look for ways to combine correlated dimensions into a single dimension. Perfectly correlated attributes, such as the levels of a hierarchy, as well as attributes with a reasonable statistical correlation, should be part of the same dimension. (Kimball et. el 2013).

BUS MATRIX									
	DIMENSIONS								
FACTS	Date	Merchant	Product	Account	Tran Type	Loan Type	Terminal	Branch	Channel
ATM TERMINAL FACT	✓				✓		✓	✓	
POS TERMINAL FACT	✓	✓			✓		✓		
ACCOUNT FACT	✓			✓				✓	
LOAN FACT	✓					✓		✓	
MERCHANT FACT	✓	✓	✓						
TRANSACTION FACT	✓		✓		✓				✓

Table 11 Bus Matrix of the proposed Data warehouse design

The logical design of this data warehouse design contains nine conformed dimensional tables and five Fact tables which are connected by a star's schema model of Ralph Kimball. As per the

suggestion of Kimball which states a very large number of dimensions typically are a sign that several dimensions are not completely independent and should be combined into a single dimension. It is a dimensional modeling mistake to represent elements of a single hierarchy as separate dimensions in the fact table. Most of the correlated dimensional tables which are used to help the intended goal of the bank in this design are combined in order to create a simplified logical model. In this study each dimension is connected with at least one fact table through a one to many relationships and represented by a foreign key in the fact table. Each fact table has a primary key that is composed of the keys of all the dimensions; this establishes a one-to-many mapping from each dimension tuple to the (many) fact tuples which share the same value as the dimension's key (Bonifati et al 2001)

5.11.1. TRANSACTION FACT

As stated in the conceptual design section Transaction Fact will allow the Bank to analyze the total count and amount of transactions performed by each product through each channel for the specified period of time. Moreover, this fact will also allow the calculation of charges gained by specific product and terminal. Hence these helps the bank to evaluate the performance of each product issued and acquired and makes decisions on the profitability of the product. This fact will provide analysis on the following issues

- General product performance per Time period
- Product performance per channel type
- Transaction load of per channel types

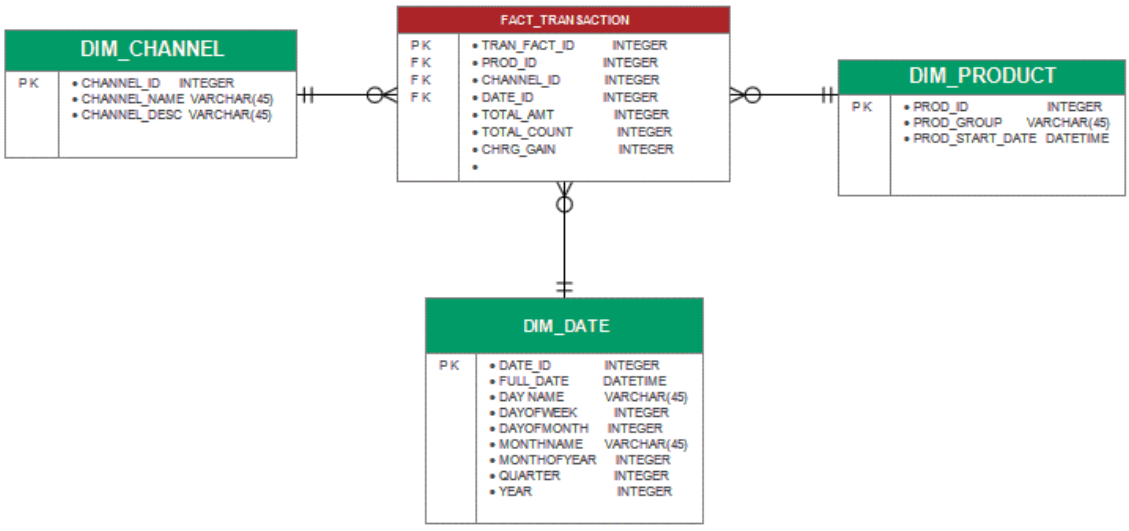


Figure 27. Star schema Logical Data Model for Transaction Fact

The fact connected with channel, date and product dimension in order to accumulate the aggregate transaction count and amount of each product per specified time of period day, week, month, quarter or year. Products in this study are the different card types issued or acquired by the bank.

5.11.2. TERMINAL FACT

As stated in the conceptual design section Terminal Fact helps the bank to evaluate the performance of ATM and POS terminals as well as terminal health by identifying the total number of transactions and types of transactions performed on the specific terminal based on the status of each transaction as Approved, declined, reversal and suspect. The analysis helps in deciding the location profitability of the terminal within a specified period of time. The fact also calculates the total transaction charge gain from each terminal and the loss of potential transaction gain in the case of ATM terminal. This fact will provide analysis on the following issues

- General Terminal Performance,
- Terminal Health Monitor
- Charge Gain and Charge Lost due to Terminal Health problem

Although both ATM and POS terminal share some common properties as a transaction channel they also differ in some of properties, functionality and mobility hence the need for its own logical design for POS arises from this deference. Some of the differences are ATMs are managed by Branches at different locations in the city whereas POS terminals are managed by the Merchant at the merchant's location, POS terminals are able to perform a refund transaction for customers but ATMs are not able to do refunds.

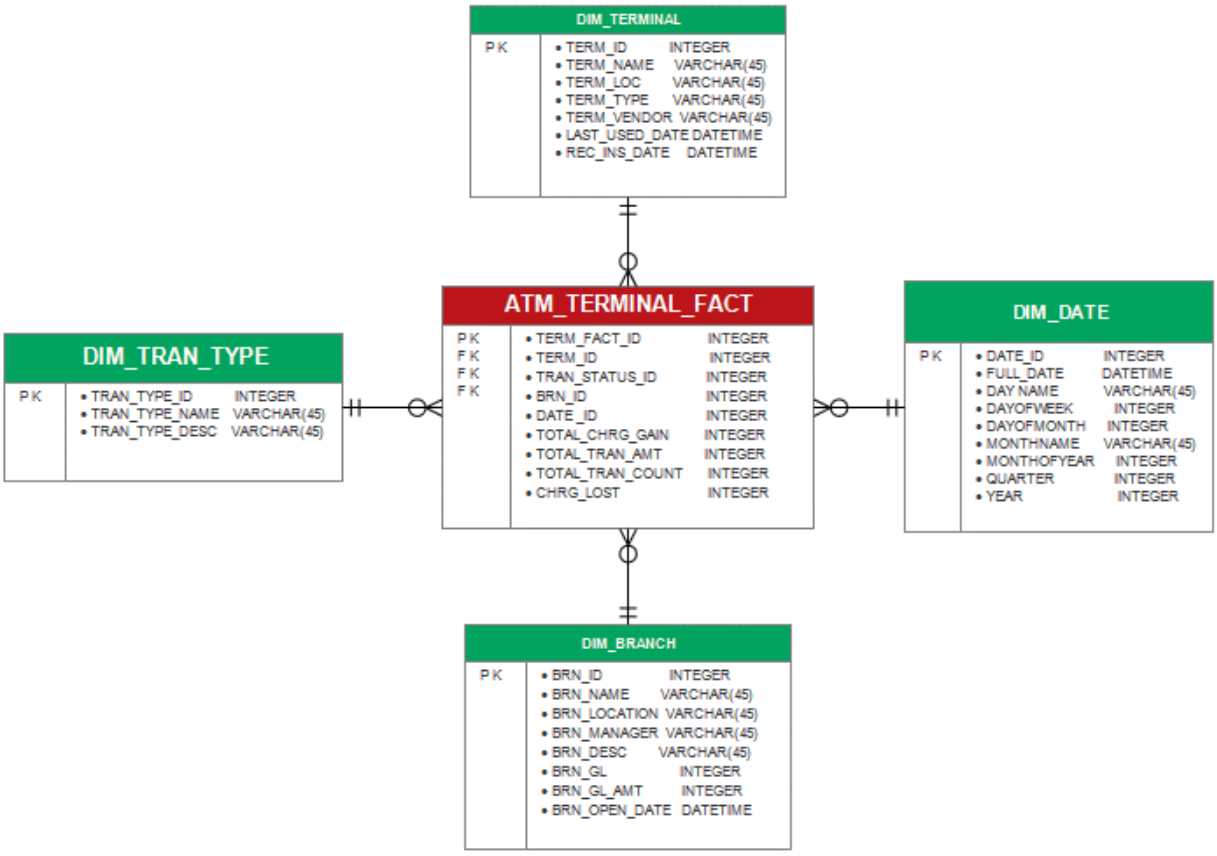


Figure28. Star schema Logical Data Model for ATM Terminal Fact

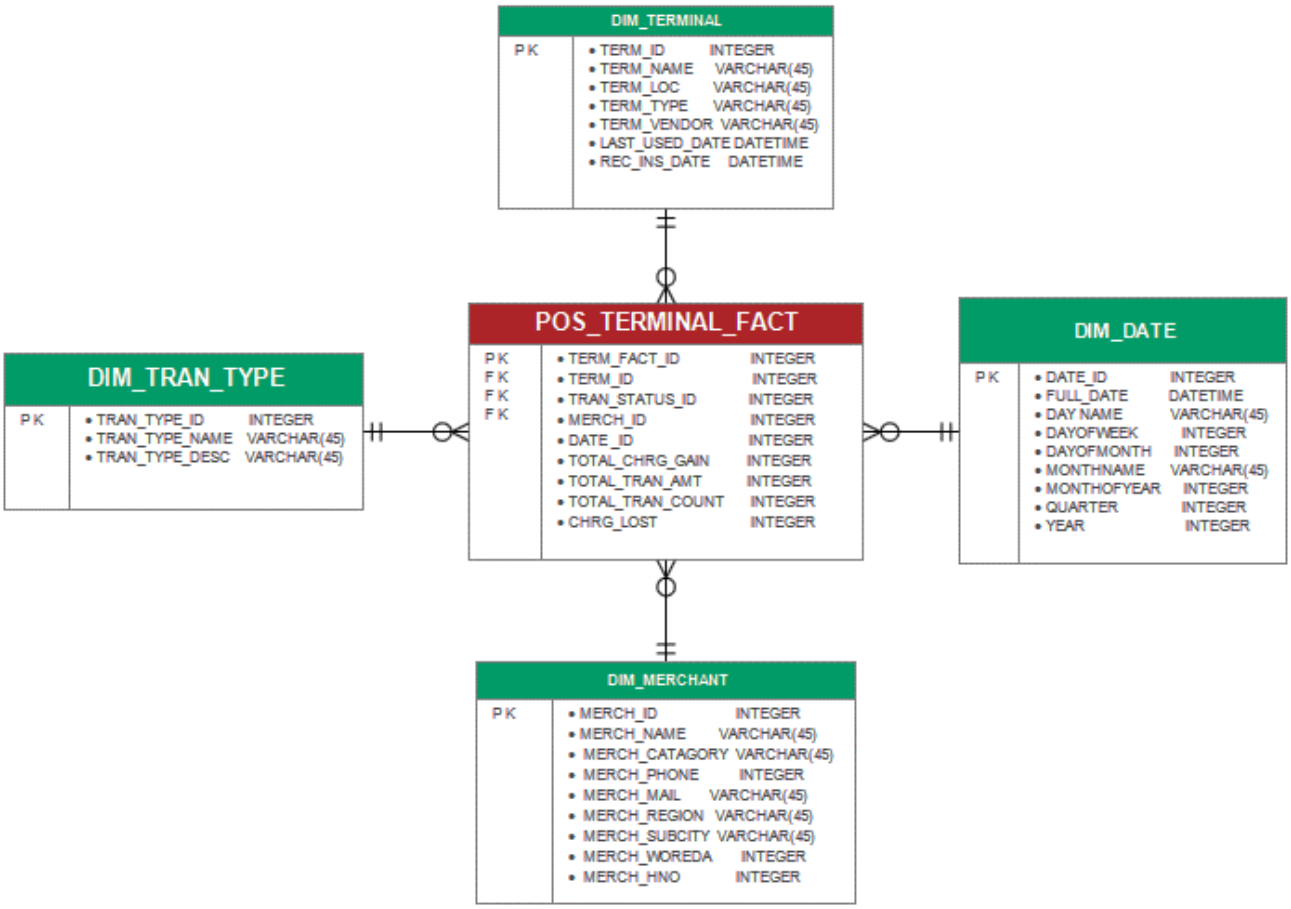


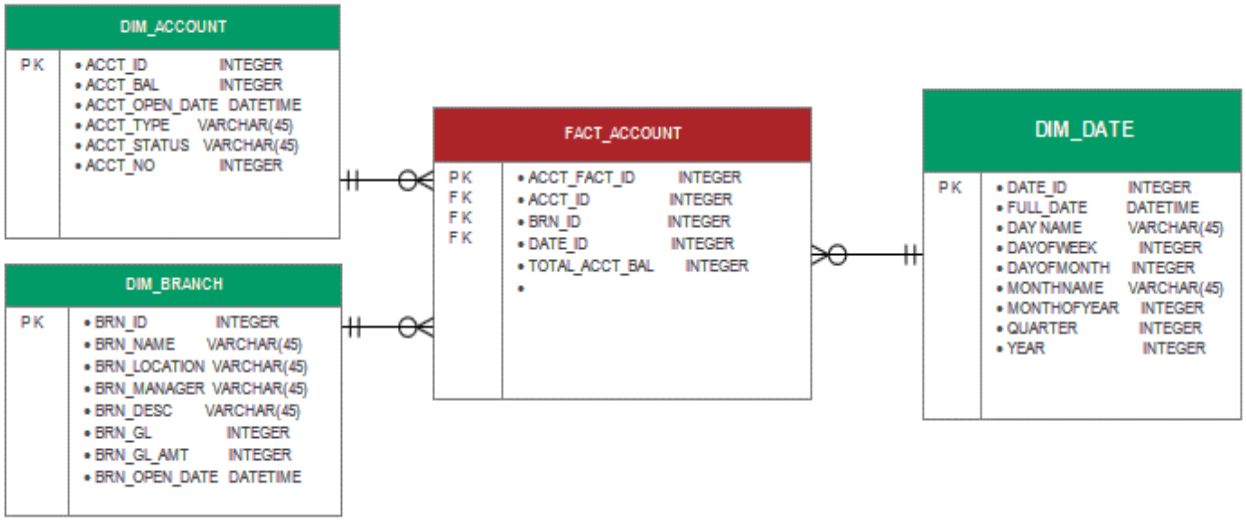
Figure 29. Star schema Logical Data Model for POS Terminal Fact

5.11.3. ACCOUNT FACT

As stated in the conceptual design section account Fact helps the bank to evaluate the performance of its branches by identifying the number of accounts which are Active, Dormant, Freeze and Blocked with their respective total amounts. The analysis helps in deciding whether the branch is gaining new customers, retaining current customers or losing its customers within a specified period of time. This fact will provide analysis on the following issues

- General Branch performance
- Branch performance per account type
- Branch performance per account status
- Branch customer retention performance

Figure 30. Star schema Logical Data Model for Account Fact



5.11.4. LOAN FACT

As stated in the conceptual design section Loan Fact helps the bank to evaluate the performance of its loan repayment from each branches by identifying the total number loans, repayments and interest gained from each branch based on the type of loan provided. The analysis helps in deciding whether the branch is following the loans it provided to its customer and collect the repayments timely within a specified period of time. This fact will provide analysis on the following issues

- General Branch Loan performance per time period
- Branch Loan performance per Loan Type
- Branch Loan interest charge per Loan type
- Branch Loan outstanding per Loan Type

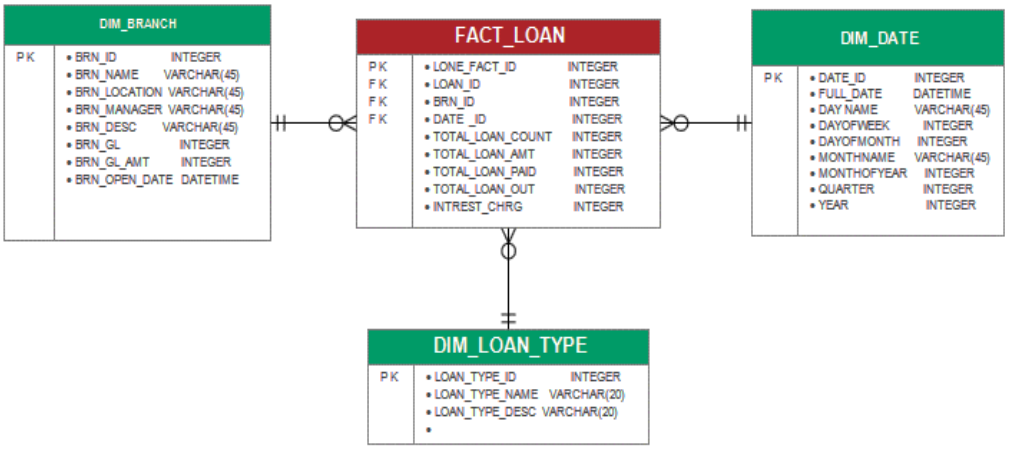


Figure 31. Star schema Logical Data Model for LOAN Fact

5.11.5. MERCHANT FACT

As described in the conceptual section Merchant Fact helps the bank to identify total number of counts and amounts of transactions performed at each merchant location with its POS terminals based on merchant category, product group and product type. This fact can be used to analyze its merchant performance, in terms of merchant category, product type and product group. Some of the questions which can be answered by this fact are

- Which merchant is using more acquirer products?
- Which merchant is using more issuer products?
- Which merchant generate more revenue to the bank?
- Which merchant is not using the terminal provided by the bank?
- Which product group is more profitable? Acquiring or Issuing
- Which product type uses the POS terminal at merchant location frequently?
- Which merchant category (Business type) is more profitable? Hotel, supermarket etc...
- Which product is frequently used at a specific merchant category?

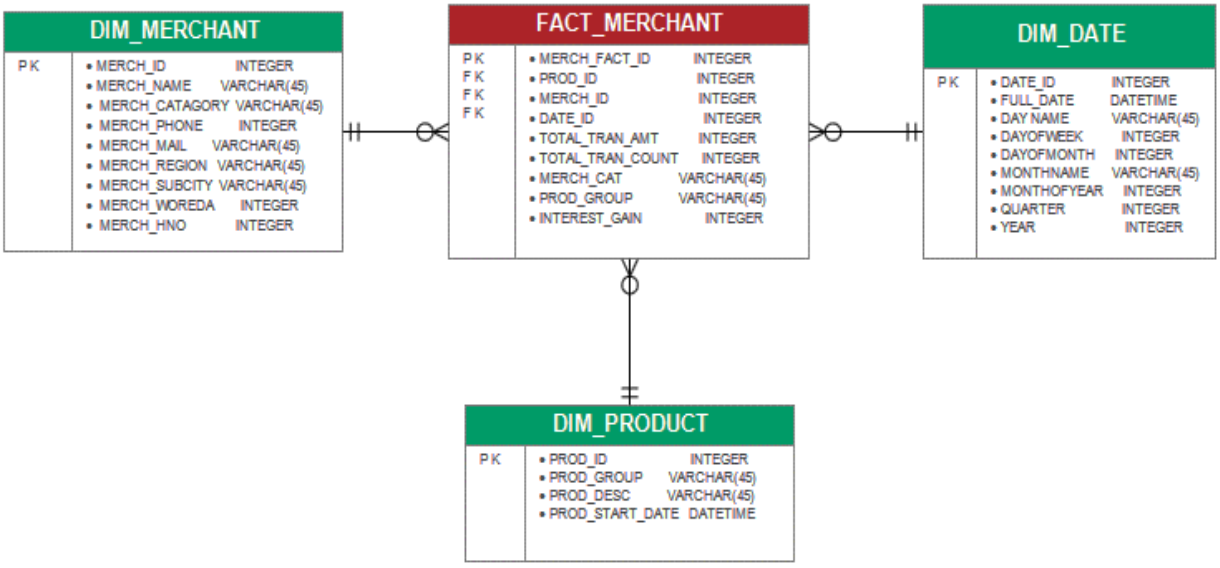


Figure 32. Star schema Logical Data Model for Merchant Fact

Terminal fact table contain a measure of aggregate count and amount of transactions performed by each terminal with the type of the transactions and includes an additional calculative attributes of charge gain and charge lost which calculates its value from the aggregate amount of the

transaction. Hence this helps to identify the number of transaction performed by each terminal, the total amount of cash flow through each terminal and the total amount of transaction charge gain by the successful transactions and possible transaction charge lost due to the unsuccessful transaction of reversal, suspect and declined status within a specific period of time. The below table shows the Terminal fact table with sample data.

5.12. PHYSICAL DESIGN

Having designed the logical model of both the fact and dimensional tables, it is now time to design the physical design of the data warehouse system. The main key column is a primary key in the fact and dimensional table, and it is a foreign key on the fact table. This is known in the database world as referential integrity. The main key in the dimensional table are usually the surrogate key, they are unique and not null, it uniquely identifies the record in a dimension tables. We made use of the surrogate because the data to each of the dimensional table are from different sources and there is need to have a unique key to identify the record. This is where the referential integrity is important.

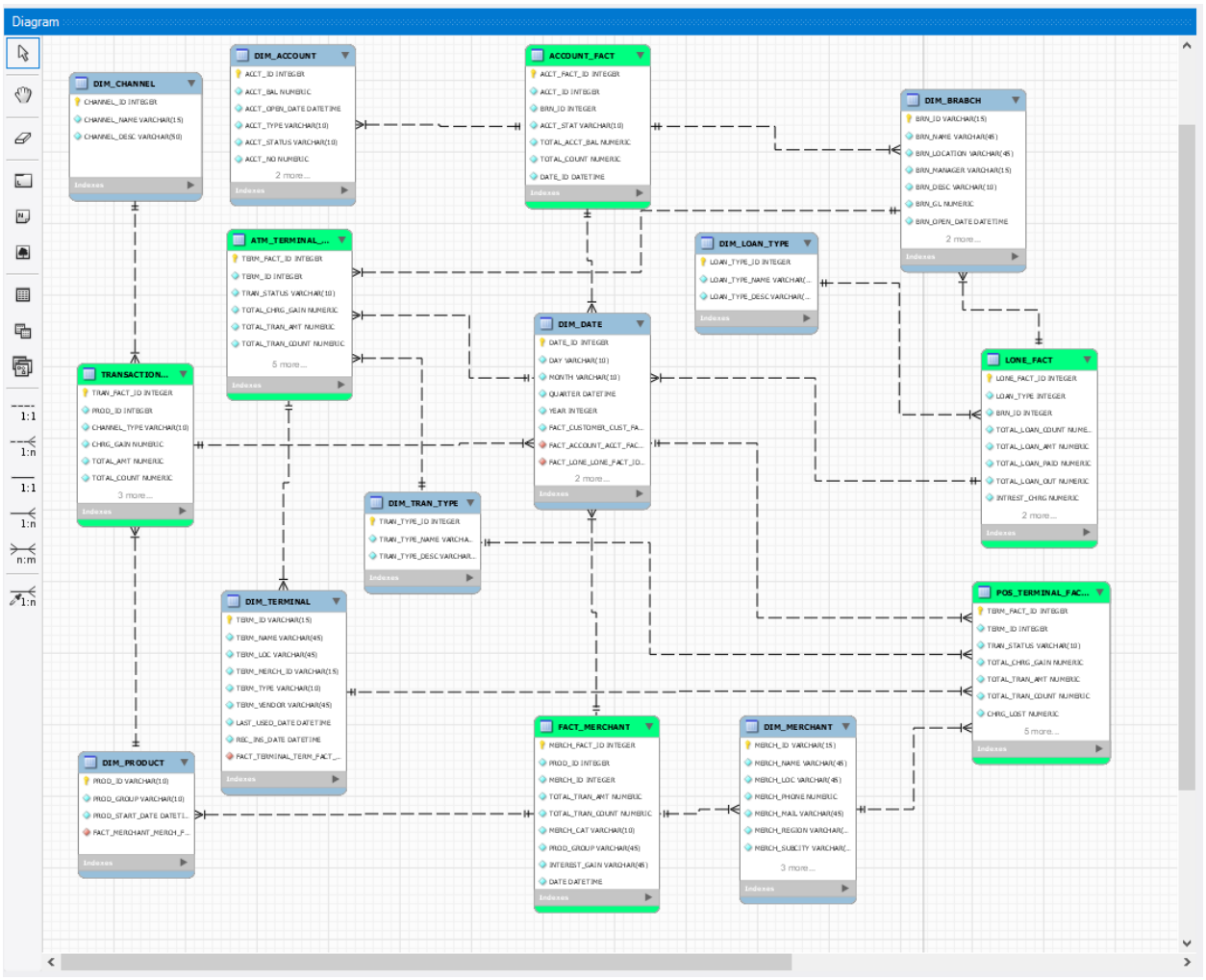


Figure 33: Physical design of the Fact and Dimension Tables on My SQL Workbench

Data warehouse is optimized for data retrieval and it is very important that users are able to run their reports as quickly as possible. In the data storage, it is good to have a database structure and right index. Indexes are the pointers to the record stored in a database. In the concept of the data warehouse, indexes are important and the help in the loading and data retrieval of the data warehouse. Indexing can significantly improve the query and loading performance of data warehousing.

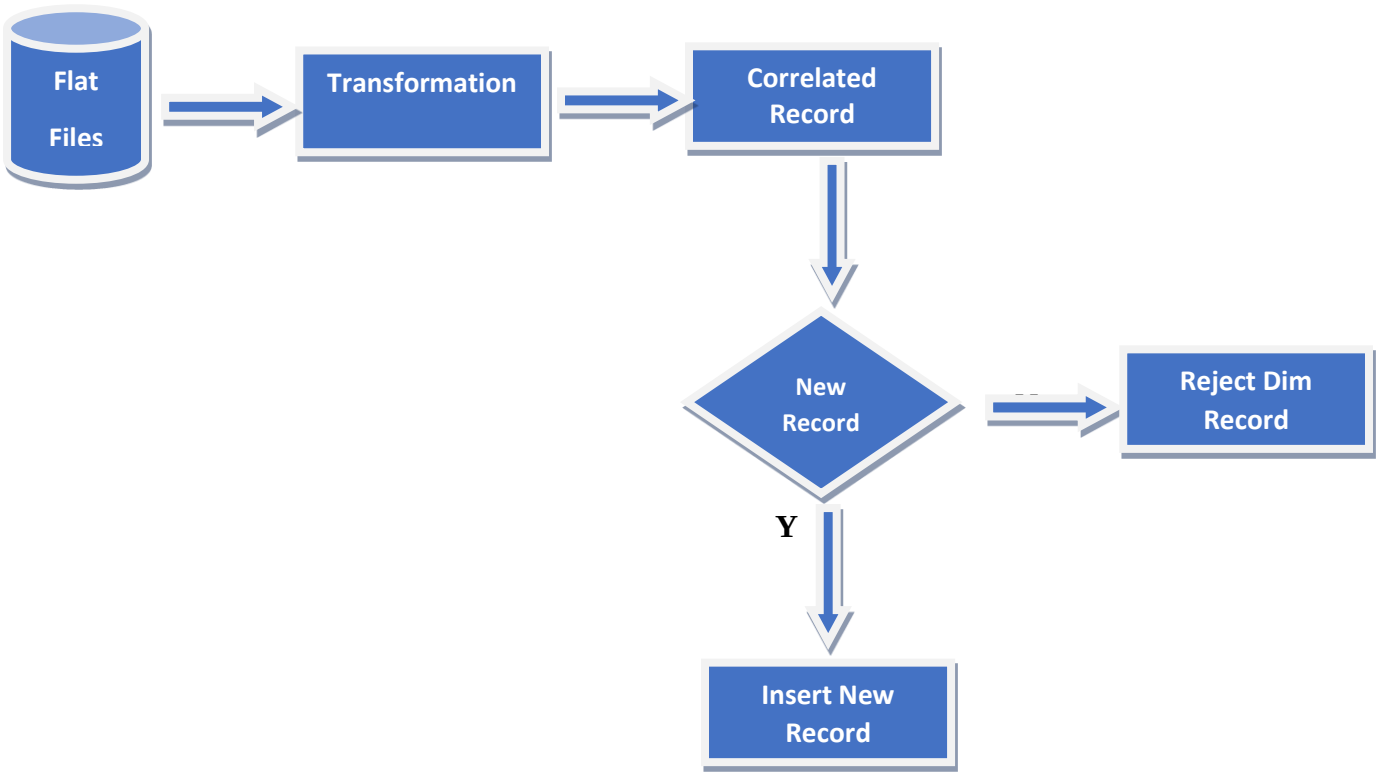


Figure 33a: Flow Chart for loading the Dimension Tables

After successful loading of the dimension tables, it is now time to load the fact table which is the last step in the process of data warehouse loading. Data are loaded from the Normalized Data store dimensional tables.

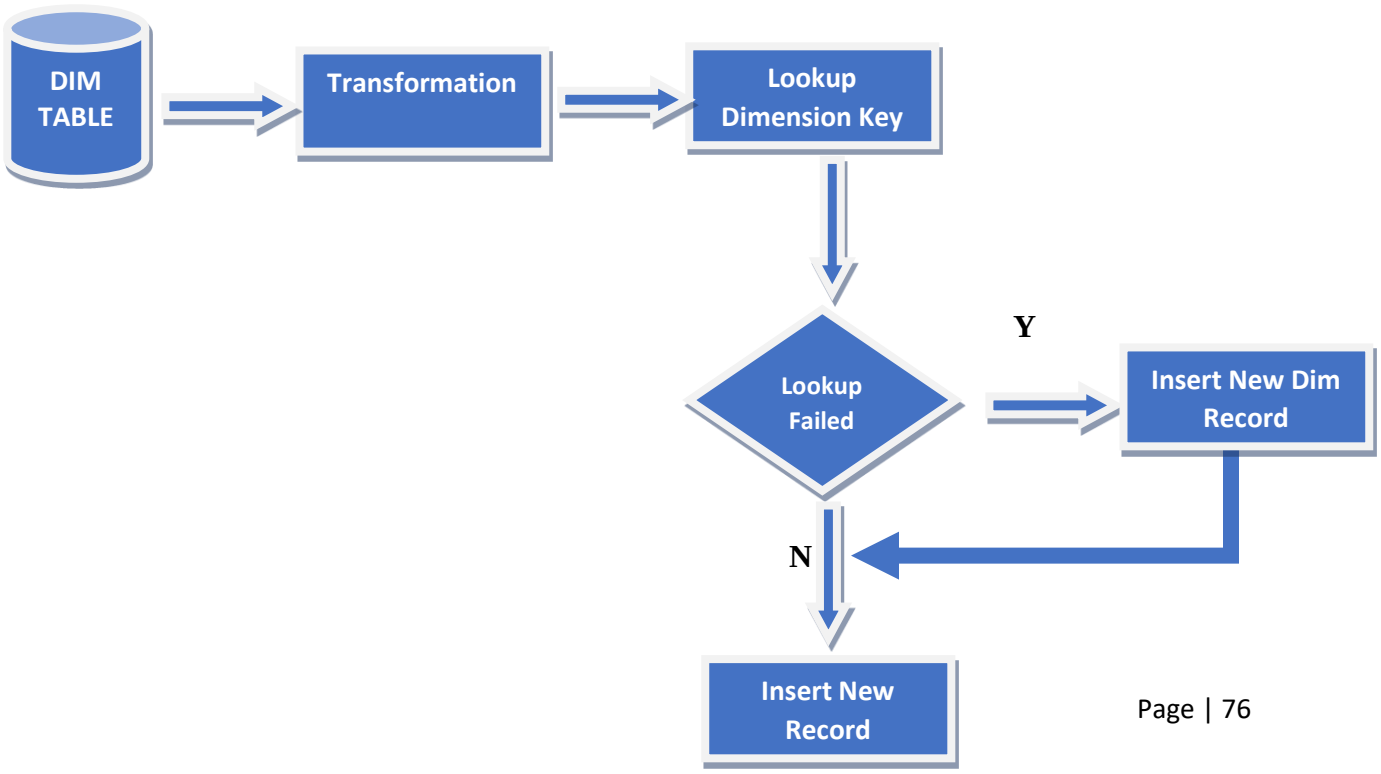
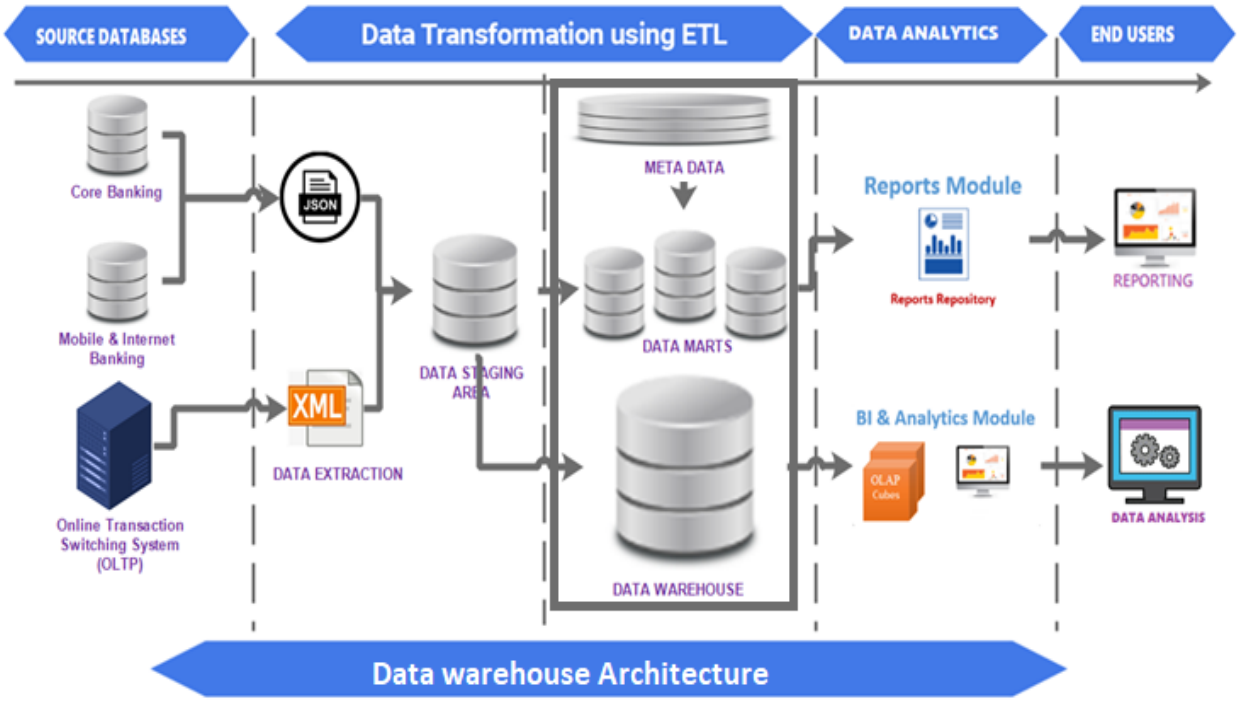


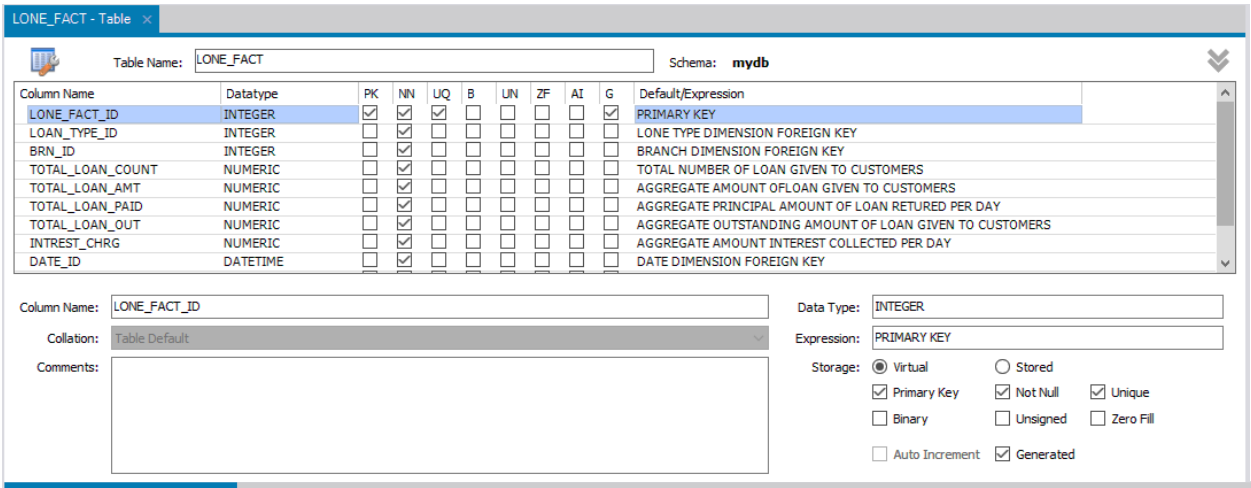
Figure 34: Flow Chart for loading the Fact Tables

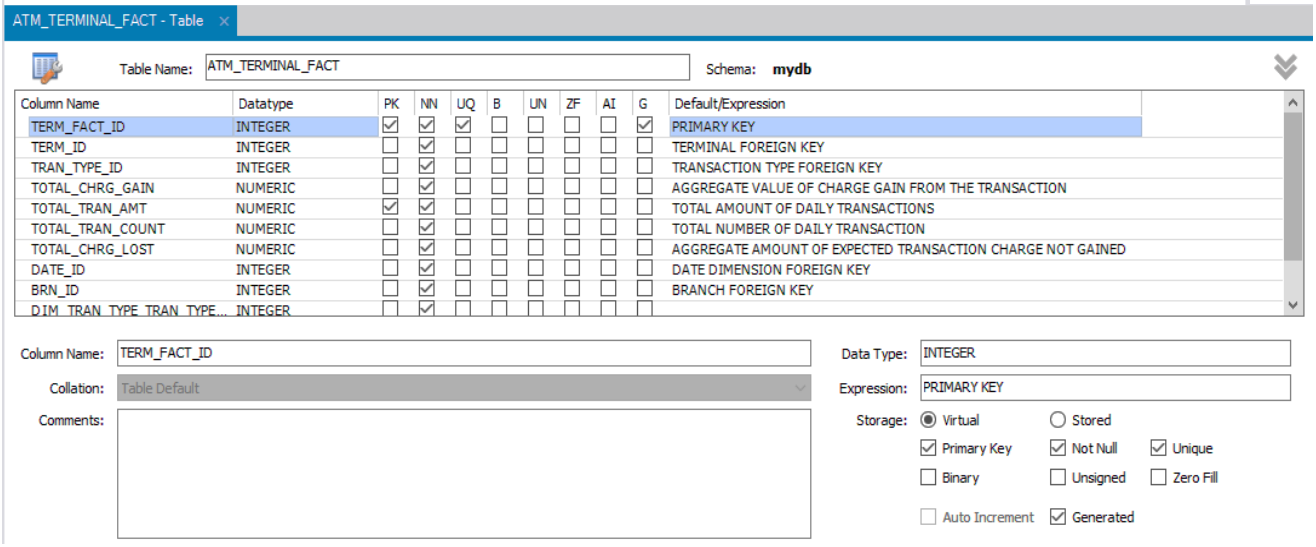
In the Dimensional Data Store, we have fact tables and we have dimension tables. They require different indexing and primary keys. We made this surrogate key column the primary key of the dimension table. We also used this surrogate key column the clustered index of the dimension table. The reason for doing this is because in the DDS, the dimension tables are joined to the fact table on the surrogate key column. By making the surrogate key a clustered primary key, we will get good query performance.

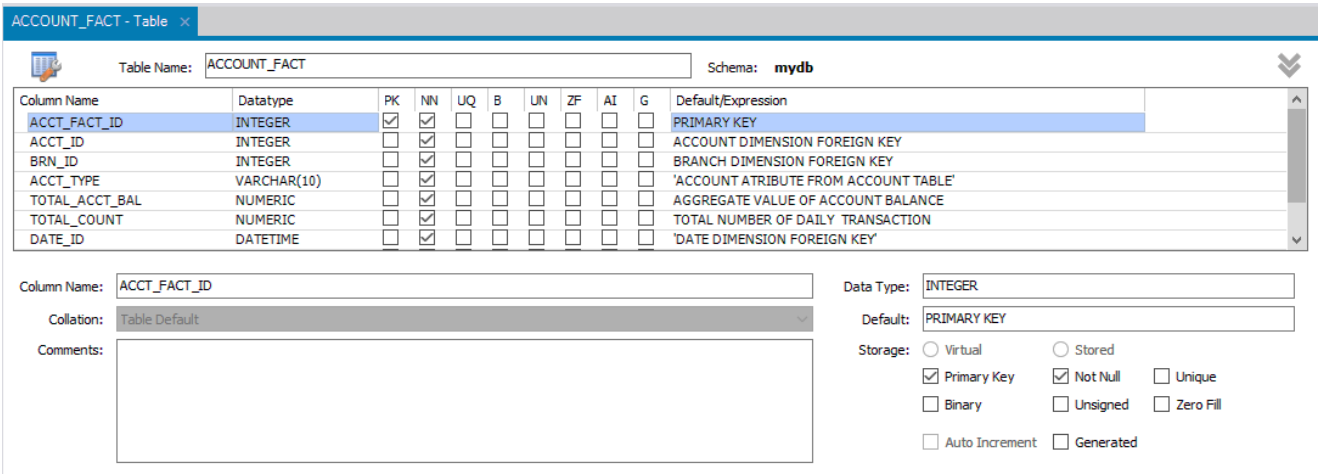
Figure 35. The data warehouse architecture



For this study, we have used six instances of servers to diagrammatically as data staging, data marts data warehouse and OLAP servers to represent a physical box of server as shown above. From the architecture above, the database engine can be installed on the data warehouse or data marts instance. This is where the database and tables would reside after the ETL process would have extracted the data from the different sources. The physical data marts can be created from one or more of the Fact tables created below.


LONE_FACT - Table
Table Name: LONE_FACT
Schema: mydb
Column Name | Datatype | PK | NN | UQ | B | UN | ZF | AI | G | Default/Expression
LONE_FACT_ID | INTEGER | PRIMARY KEY
LOAN_TYPE_ID | INTEGER | LONE TYPE DIMENSION FOREIGN KEY
BRN_ID | INTEGER | BRANCH DIMENSION FOREIGN KEY
TOTAL_LOAN_COUNT | NUMERIC | TOTAL NUMBER OF LOAN GIVEN TO CUSTOMERS
TOTAL_LOAN_AMT | NUMERIC | AGGREGATE AMOUNT OFLOAN GIVEN TO CUSTOMERS
TOTAL_LOAN_PAID | NUMERIC | AGGREGATE PRINCIPAL AMOUNT OF LOAN RETURED PER DAY
TOTAL_LOAN_OUT | NUMERIC | AGGREGATE OUTSTANDING AMOUNT OF LOAN GIVEN TO CUSTOMERS
INTREST_CHRG | NUMERIC | AGGREGATE AMOUNT INTEREST COLLECTED PER DAY
DATE_ID | DATETIME | DATE DIMENSION FOREIGN KEY
Column Name: LONE_FACT_ID
Collation: Table Default
Comments:
Data Type: INTEGER
Expression: PRIMARY KEY
Storage: Virtual Stored
Primary Key Not Null Unique
Binary Unsigned Zero Fill
Auto Increment Generated


ATM_TERMINAL_FACT - Table
Table Name: ATM_TERMINAL_FACT
Schema: mydb
Column Name | Datatype | PK | NN | UQ | B | UN | ZF | AI | G | Default/Expression
TERM_FACT_ID | INTEGER | PRIMARY KEY
TERM_ID | INTEGER | TERMINAL FOREIGN KEY
TRAN_TYPE_ID | INTEGER | TRANSACTION TYPE FOREIGN KEY
TOTAL_CHRG_GAIN | NUMERIC | AGGREGATE VALUE OF CHARGE GAIN FROM THE TRANSACTION
TOTAL_TRAN_AMT | NUMERIC | TOTAL AMOUNT OF DAILY TRANSACTIONS
TOTAL_TRAN_COUNT | NUMERIC | TOTAL NUMBER OF DAILY TRANSACTION
TOTAL_CHRG_LOST | NUMERIC | AGGREGATE AMOUNT OF EXPECTED TRANSACTION CHARGE NOT GAINED
DATE_ID | INTEGER | DATE DIMENSION FOREIGN KEY
BRN_ID | INTEGER | BRANCH FOREIGN KEY
DIM_TRAN_TYPE_TRAN_TYPE... | INTEGER
Column Name: TERM_FACT_ID
Collation: Table Default
Comments:
Data Type: INTEGER
Expression: PRIMARY KEY
Storage: Virtual Stored
Primary Key Not Null Unique
Binary Unsigned Zero Fill
Auto Increment Generated


ACCOUNT_FACT - Table
Table Name: ACCOUNT_FACT
Schema: mydb
Column Name | Datatype | PK | NN | UQ | B | UN | ZF | AI | G | Default/Expression
ACCT_FACT_ID | INTEGER | PRIMARY KEY
ACCT_ID | INTEGER | ACCOUNT DIMENSION FOREIGN KEY
BRN_ID | INTEGER | BRANCH DIMENSION FOREIGN KEY
ACCT_TYPE | VARCHAR(10) | 'ACCOUNT ATRIBUTE FROM ACCOUNT TABLE'
TOTAL_ACCT_BAL | NUMERIC | AGGREGATE VALUE OF ACCOUNT BALANCE
TOTAL_COUNT | NUMERIC | TOTAL NUMBER OF DAILY TRANSACTION
DATE_ID | DATETIME | 'DATE DIMENSION FOREIGN KEY'
Column Name: ACCT_FACT_ID
Collation: Table Default
Comments:
Data Type: INTEGER
Default: PRIMARY KEY
Storage: Virtual Stored
Primary Key Not Null Unique
Binary Unsigned Zero Fill
Auto Increment Generated

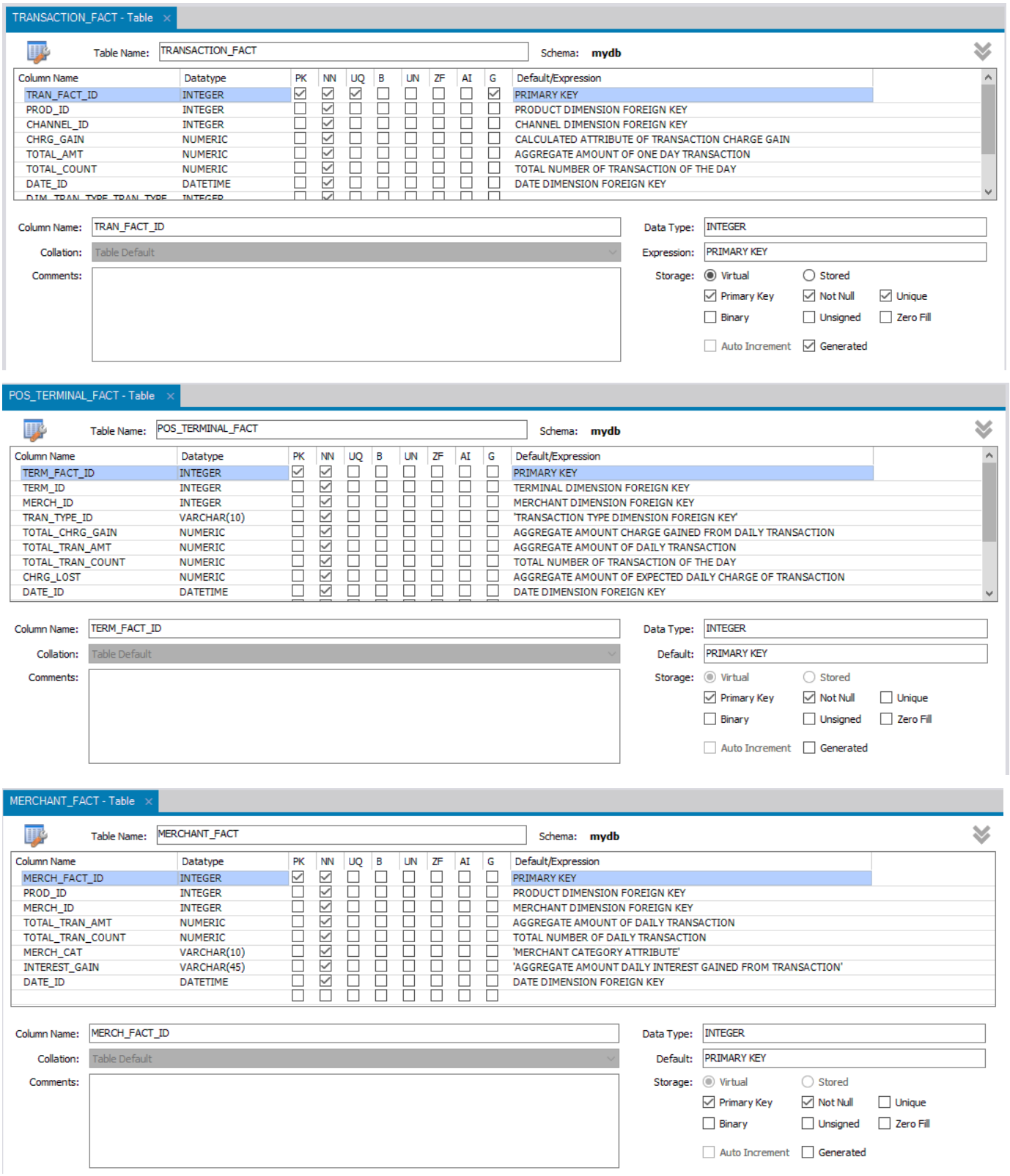


Figure 36. Physical data Model of Facts

5.13. DATA MARTS

A data mart is a simple form of a data warehouse and is designed to serve a specific purpose from the organizations objectives; the assumption is to divide the enterprise wide data warehouse to smaller and easily accessible warehouses, focused on a particular line of business, department, or subject area. Data marts can be simple because each of them are focused on a single aspect over the operational data [Kimball 1996 cited by A. Bonifati et al 2001].

Unlike an enterprise data warehouse that's expensive and complex to create, it offers a cost-efficient alternative. It also allows faster data access and is simple to use, as it's precisely designed according to the end user requirements and focuses on a single department/subject area. Data mart design is oriented toward the needs of end users. End users of a typical bank want to perform analysis and look at aggregated data, rather than at individual transactions.

Once the logical design of facts and dimension is completed the design of data marts from these fact tables will depend on the requirements of the end users and performance and access simplicity. A single data mart can contain one or more fact tables, depending on the correlation of facts and subject area, hence we can have four or five data marts by combining Transaction and Merchant Facts into one data mart since both facts are based on transactions, the other facts that can be integrated into one data marts can be ATM and POS terminal facts which are both dependent on terminal and transaction dimensions.

Data marts can be dependent or independent depending on the creation of the marts. Dependent data marts are the subdivisions of a larger data warehouse that serves as a centralized data source which is also known as the top-down approach. Whereas Independent data marts act as standalone systems, which can work without a data warehouse, they receive data from external and internal data sources directly. Since the study follow Ralph Kimball's approach the independent data marts are applied. Although it is possible to integrate more than one facts into one data marts, using a single fact table to one data mart also provide the highest performance update flexibility and low access response time.

6. CHAPTER SIX

6.1. CONCLUSION AND RECOMENDATIONS

In this information age data become the very essential tool to be successful in almost every business; the large volume of data collected every day from the different devices is a powerful tool in the success of the business. According to a survey from Deloitte, 49% of respondents say that analytics helps them make better decisions, 16% say that it better enables key strategic initiatives, and 10% say it helps them improve relationships with both customers and business partners. The survey concluded that in today's business the use of data for decision making process is not a luxury any more it has become an economic necessity.

Importance of DW cannot be denied due to its benefits because decisions at management level will no longer need to be taken on the limited and inaccurate data and it also helps the companies to avoid different challenges. So it becomes the need of every individual company to implement data warehouse. (Mohammed et al. 2016). With this thought in mind this study tried to show the application data warehouse in the Ethiopian banking industry by developing a framework for the application of data warehousing technology.

The study identified

- The study Identify the data that should be integrated in the Bank,
- Proposed nine conformed dimensional table which are basic in the banks day to day financial activity
- Proposed Six Fact tables to summarize and help develop data marts
- Developed conceptual, logical and physical design of the proposed system

which are very essential in serving the banks benefit from the accumulated data set by increasing the performance of its terminals, measuring the performance of each branch through interest collected from the lone given by each branches and customer handling performance and identify the performance of its merchant customers for each products provided and supported by the Bank.

6.2. RECOMMENDATIONS

The study from Addis Ababa University School of information science by Leule Berhe on the role of data mining technology in electronic transaction expansion at Dashen bank becomes the

initiative for this study. By identifying the data concentration area in the bank this study covers the conceptual, logical as well as physical design of a data warehouse for the banking industry.

6.3. LIMITATIONS OF THIS RESEARCH STUDY

First, the validation of both Application and appropriateness of the data warehouse were only aims at the case study company (Dashen Bank S.C) and it is not possible to generalize the design or the overall assessment to other Banks as there are differences in organizational factor, internal process and technological maturity.

Second, this thesis does not include all data from all available sources, but rather from the three area of high data concentrations which are created from the day to day transactional activity of the Bank. As mentioned in the implication for further research section the implementation of the data warehouse in the Ethiopian banking industry can be one of the research areas in the future. For further researches, I recommend researchers to validate the workability of the design as a continuation or extension of the framework and add their own methods to improve the design.

6.4. IMPLICATIONS FOR FURTHER RESEARCH

This study provides a basis of research into the application, design and implementation of a data warehousing technology in the Ethiopian banking industry, especially for those banks who are eager to benefits from their data accumulation and willing to study their customers, terminal devices, merchant's behavior and Branch performance.

As mentioned above, one of the limitations of this study is limited to some source databases only which makes the study limited even within a single case study organization, hence future studies of banking data warehouse can be done including all the available and valuable source databases. Banks should also consider Data warehousing technology as a solution for their lack of customer handling problem and many other current problems through the use of data analysis of historical data accumulated in their data store.

6.5. DATA WAREHOUSING TECHNOLOGY AWARENESS

Success in banking industry requires a growing customer base and being able to retain existing customers. But without a way to gain insight into customer trends and historical data, a growing banking business can be difficult. Developing the awareness of a data warehouse technology within the organization, a continuous training and development strategy on the new data

warehousing technology and maintaining the use of historical as well as operational data can be the first step in this process.

6.6. TO THE BANK

The bank owns many terminal devices, products and also acquires other products from other banks and also does transactions through different channels Digital, ATM, POS and Branch this transaction which come through different channels can provide valuable information to the Bank about the customers, the devices and performance of Branches if they are properly analyzed. Although the Bank started the initiative by starting building reporting system this initiative should also include data from all aspects of the banks data source and historical data and support the decision making process of the management by applying a fully-fledged data warehousing technology.

6.7. FURTHER IMPROVEMENT OF THE PROPOSED DESIGN

In this study, the Application of data warehouse is discussed using Ralph Kimball's multi - dimensional data warehouse development methodology the design is dependent on observation, focus group discussion, and interview on survey conducted by the banks IT department on the different branches and departments for the current reporting system. This was one-time survey conducted by the department one year before the initiation of this study. I believe, the design could be improved if the study can include consultants, customers, vendors and merchants so that it can assess the internal reliability, validity and perceived value.

LIST OF RELEVANT LITERATURE

1. Alemayehu Geda (2006) The Structure and Performance of Ethiopia Financial Sector in the Pre- and Post-Reform Period with a Special Focus on Banking, United Nation University UNU-WIDER 2006
2. Angela Bonifati, Fabiano Cattaneo, Stefano, Alfonso Fuggetta, Stefano Paraboschi Designing (2001) Data Marts for Data Warehouses, ACM Transactions on Software Engineering and Methodology, Vol. 10, No. 4, October 2001, Pages 452–483
3. Alejandro Vaisman and Esteban Zimanyi (2012) Data Warehouses Next Challenges, Lecture Notes in Business Information Processing, 978-3-642-27357-5
4. Amanpartap Singh PALL and Dr. Jaiteg Singh KHAIRA (2013) A comparative Review of Extraction, Transformation and Loading Tools School of Information Technology, APJIMTC, Jalandhar Chitkara Institute of Engineering and Technology, Rajpura, Database Systems Journal vol. IV, no. 2.
5. Arvind Kumar (2016) Exploring the Applications of Data Warehouse in Public Life. International Journal of Computer Science and Information Technologies, Vol. 7 (6), 2016.
6. Basel Committee on Banking Supervision (BCBS, 2001) Working Paper on the Regulatory Treatment of Operational Risk
7. Beate, Robert, Karl, Josef (2002) A Comparison of Data Warehouse Development Methodologies Case Study of the Process Warehouse, IBM Research Division, Thomas J. Watson Research Center York town Heights, NY 10598, United States
8. Bogdan and Emina (2011) Application of Business Intelligence in the Banking Industry, Management Information Systems, Vol. 6 (2011)
9. Chen Wenzhe (2015). The Application of Data Warehouse Technology in Modern Finance, International Conference on Advances in Mechanical Engineering and Industrial Informatics.
10. C. Suba (2018) Data Warehousing Methods and its Applications. International Journal of Engineering Science Invention (IJESI) ISSN (Online): 2319 – 6734, ISSN
11. Chuck Ballard, Dirk Herreman, Don Schau, Rhonda Bell, Eunsaeng Kim, Ann Valencic (1998) Data Modeling Techniques for Data Warehousing, IBM Corporation, International Technical Support Organization SG24-2238-00.

12. Crescenzio gallo, Michel Angelo de bonis, Michele perilli (2010) Data warehouse design and management: theory and practice, dipartimento di scienze economiche, matematiche e statistic Heuniversit `a di Foggia largo papa Giovanni paolo ii, 1 - 71121 foggia, Italy.
13. C.R. Kothari(2004) Research Methodology, methods and Techniques second revised edition Published by New Age International (P) Ltd., Pu
14. Daniel Fasel (2012) Concept and Implementation of a Fuzzy Data Warehouse University of Fribourg, Switzerland.
15. Deloitte Touché Tohmatsu Limited (2013) The Analytics Advantage, We're just getting started The first few steps of the analytics journey hold promise for the long term www.deloitte.com/analytics survey
16. Emilio Power & Greg Nelson (2008) ETL and Data Quality: Which Comes First? The Wave Technologies, LLC, Chapel Hill, NC, SAS Global Forum 2008.
17. Earl Von F. Lapuraa *, John Kenneth J. Fernandez, Mark Jonathan K. Pagatpata , Dante D. Dinawanao (2018) Development of a University Financial Data Warehouse and its Visualization Tool. Published by Elsevier, ICT center, MSU – Iligan Institute of Technology, Andres Bonifacio Avenue, Tibanga, 9200 Iligan City, Philippines
18. Fernando and G.N. Wikramanayake (2016) Application of Data Warehousing & Data Mining to Exploitation for Supporting the Planning of Higher Education System in Sri Lanka, University of Colombo School of Computing.
19. Gallo, De Bonis and Perilli (2010) Data warehouse design and management theory and practice. dipartimento di scienze economiche, matematiche estatistiche universit`a di Foggia largo papa giovanni paolo ii, 1 - 71121 Foggia, Italy
20. Garfield Kipruto Settim(2018) an approach for successful information system implementation: a case of core banking systems in Kenya, united states international university, Africa
21. H.J. Watson, B.H. Wixom, D.L. Goodhue (2002) The Benefits of data warehousing why some organizations realize exceptional payoffs. Information and management, Elsevier science 2002.
22. lker Etikan, Sulaiman Abubakar Musa, Rukayya Sunusi Alkassim (2016) Comparison of Convenience Sampling and Purposive Sampling. American Journal of Theoretical and Applied Statistics. Vol. 5, No. 1, pp. 1-4. doi: 10.11648/j.ajtas.20160501.11
23. James D. Torr (2003) The information age, Green haven Press. ISBN 0-7377-1186-8

24. Jacek Maślankowski (2013) The Evolution of the Data Warehouse Systems in Recent Years University of Gdańsk, Poland, Faculty of Management, Department of Business Informatics.
25. Jiawei Han (1997) OLAP Mining an Integration of OLAP with Data Mining Intelligent Database Systems. Research Laboratory School of Computing Science, Simon Fraser University, British Columbia, Canada
26. José Ferreira1, Fernando Almeida, José Monteiro (2017) Building an Effective Data Warehousing for Financial Sector Higher Polytechnic Institute of Gaya, V.N.Gaia, Portugal Faculty of Engineering of Oporto University, INESC TEC, Porto, Portugal Corresponding.
27. Juan Trujillo, Manuel Palomar, and Il-Yeol Song (2002) Designing Data Warehouses with object Oriented Conceptual Models.0018-9162/01/ 2001 IEEE
28. katherine liddell avery, A study of data warehousing end user training methods, their effectiveness, and challenges, *Athens, Georgia 2003*
29. Lokanath Mishra (2016) Focus Group Discussion in Qualitative Research, Techno LEARN Vol. 6: No. 1: p. 1-5, June 2016, Department of Education, Mizoram University, Aizawl, Mizoram, India
30. Mary Breslin (2004) Data Warehousing Battle of the Giants: Comparing the Basics of the Kimball and Inmon, Models business intelligence journal
31. Markus Tresch and Dirk Jonscher (2001) Evolution not Revolution: The Data Warehousing Strategy at Credit Suisse Financial Services, credit Suisse financial services Technology & Services 8070 Zurich, Switzerland.
32. Matteo Golfarelli, Dario Maio and Stefano Rizzi (1998)The dimensional fact model: A conceptual model for data warehouses, DEIS - University di Bologna, Viale Risorgimento 2, 40136 Bologna, Italy ,International Journal of Cooperative Information Systems Vol. 07, No. 02n03, pp. 215-247
33. Mohammed Kemal (2019), Intangible Advantages of a Data Warehouse (DW), https://www.researchgate.net/publication/331971360_Intangible_Advantages_of_a_Data_Warehouse_DW
34. Mohd Kamir Yusof and Mustafa Man (2018) Efficiency of Flat File Database Approach in Data Storage and Data Extraction for Big Data, Indonesian Journal of Electrical Engineering and Computer ScienceVol.9, No.2, February2018, pp. 460

35. M.G.N.A.S. Fernando and G.N. Wikramanayake (2016) Application of Data Warehousing & Data Mining to Exploitation for Supporting the Planning of Higher Education System in Sri Lanka, University of Colombo School of Computing {nas, gnw}@ucsc.cmb.ac.lk
36. Muhammad B. Shahid, Umber Sheikh, Basit Raza, Munam A. Shah, Ahmad Kamran, Adeel Anjum (2016) Application of Data Warehouse in Real Life: State-of the-art Survey from User Preferences Perspective. International Journal of Advanced Computer Science and Applications.
37. Nenad Jukic, Boris Jukic, and Mary Malliaris (2008) Online Analytical Processing (OLAP) for Decision Support, School of Business Administration, Loyola University Chicago, IL, School of Business, Clarkson University, Potsdam, NY, USA
38. N.P Hariharan PHD (2015) Challenges of Core Banking Systems Mediterranean Journal of Social Sciences MCSER Publishing, Rome-Italy Vol. 6 No 5 (ISSN 2039-2117 (online)
39. Radu Lovin (2004) Data Warehouse and OLAP. Methodologies, Algorithms, Trends, Economy Informatics.
40. Ralph Kimball and Margy Ross (2013) The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition Published by John Wiley & Sons, Inc.10475 Cross point Boulevard Indianapolis, IN 46256
41. Robert m. Bruckner, Beate list, Josef Schiefer Requirements for data warehouse systems with use cases, 2001 Seventh Americas Conference on Information Systems
42. Robert C. Bogdan and Sari Knopp Biklen (1998) Qualitative Research for Education: An Introduction to Theory and Methods, Third Edition from Syracuse University and Syracuse University
43. Stefano Rizzi, Conceptual Modeling Solutions for the Data Warehouse, DEIS-University of Bologna, Italy 2007, Idea Group Inc.
44. Surajit Chaudhuriand Umeshwar Dayal (1997) An Overview of Data Warehousing and OLAP Technology. Microsoft Research, Redmond and Hewlett-Packard Labs, Palo Alto, Appears in ACM Sigmod Record.
45. Torben Bach Pedersen (2016) Multidimensional Modeling, Aalborg University, Denmark, Encyclopedia of Database Systems.

46. Ugochukwu, Musbahu, Aminu and Abdullahi (2019) Data Mining Applications In Banking Sector For Effective Service Delivery, Computer Science Department, Hussaini Adamu Federal Polytechnic Kazaure, Jigawa Nigeria.
47. Van Scott (2000) Extraction, Transformation, and Load Issues and Approaches
48. Vaibhav, Shrinivas, Ujwal (2014) Data Warehouse Architecture for Financial Institutes to Become Robust Integrated Core Financial System using BUID, International Journal of Advanced Research in Computer and Communication Engineering.
49. V Postolache (2017) Optimization of business processes in banks through flexible workflow Alecu Russo Balti State University, Republic of Moldova, Department of Machine Manufacturing Technology, str. Puşkin, No. 38, MD-3100, Balti, Republic of Moldova
50. Wen-Yang Lin and I-Chung Kuo (2004) A Genetic Selection Algorithm for OLAP Data Cubes, Article in Knowledge and Information Systems · July 2004
51. Ye s., Sun w. and Tian z. (2004). Data extraction and transformation with flat file for business integration. In Proceedings of the Sixth International Conference on Enterprise Information Systems, pages 643-646DOI: 10.5220/0002652106430646

APPENDICES A: GROUP DISCUSSION

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

DEPARTMENT OF INFORMATION SCIENCE

Discussion Outline

In partial fulfillment of the requirements for the degree of masters of Science in information system, I Zelalem Tesfaye Kassaye undertaking a study on the "application of data warehouse technology in the Ethiopian banking industry: The case of Dashen Bank share company". At Addis Ababa University I have accordingly preparing a group discussion on the topic with the objective of exploring the current organizational commitment, current operational capability and cultural readiness issues that will affect the application of Data warehouse at Dashen Bank Share Company. The discussion points are categorized into introductory, key discussion points and closing points. Your honest responses to the discussion are extremely valuable to the outcome of this study. The discussion approximately will take not more than 30 minutes to complete and the result of the discussion will be used for academic study only. Hence all responses will be kept in strict confidentiality.

Welcome

- Good evening and welcome to this discussion session My name is Zelalem Tesfaye
- Thank you for being willing to participate in this discussion and taking your time to join Me to talk about data warehouse.
- .

Our topic

- The results will be used as an input for a thesis paper of partial fulfillment for master's degree in information systems
- You were invited because you all work in Dashen Bank and you are working in one or the other way with different databases in the bank and some of you working directly on the banks initiative project of a data warehouse, so you're familiar with what data warehouse is does.

Guidelines

- There are no wrong answers but rather differing points of view. Please feel free to share your point of view even if it differs from what others have said.
- Keep in mind that we're just as interested in negative comments as positive comments, and at times the negative comments are the most helpful.
    - We're tape recording, one person speaking at a time
    - we won't use any names in our reports
    - You don't need to agree with others, but you must listen respectfully as others share their views
    - Rules for cellular phones and pagers if applicable. For example: We ask that your turn off your phones or pagers. If you cannot and if you must respond to a call, please do so as quietly as possible and rejoin us as quickly as you can.
    - My role as moderator will be to guide the discussion
    - Let's find out some more about each other by going around the table. Tell us your name and your title in the Bank.

APPENDEX B

List of Available operational data bases in the Bank

- Core Banking System (CBS)
- Online transaction processing system (BASE 24)
- Dispute Management system(DMS)
- Payment and Settlements Follow-up Management System
- Cash load and Count Management System
- Stock Receiving and Issuing Management System
- Budget Management System
- Dispute & Settlements Database
- Engineer valuation (EVPMS)
- Innovative idea collection system (IICS)
- Risk Registration Management System
- School fee collection management system
- Shareholder Management System
- Credit Operation Management System
- International money transfer management system (remittance)
- Loan committee meeting registration system
- Incident Management System
- Audit Finding Management System
- Learning management system (LMS)
- Import & export related and foreign currency permits system
- Fleet management system (vehicles, fuel records, maintenance, insurance related, etc.)
- Project Cost Monitoring Management System
- Incoming SWIFFT Messages Registration System
- Human resource management System
- Manage Lease Contract Information Based on the IFRS-16 reporting standards
- Legal Person Customer Information Registration System
- Swift transaction Management Legacy System (SIMRS)

QUESTION TYPE	PURPOSE	QUESTIONS	ANSWERS
Introduction	General understanding about DW	1. What is your understanding about data warehouse?	A. Data warehouses a database which contain historical, current, and critical enterprise data. They form the storage and processing platform underlying reporting, dashboards, business intelligence, and analytics.
Introduction	Understanding the importance of DW	2. What do you think about the application of DW technology in the organization (Why is data warehousing important)?	Warehousing is important because it allows organizations to: A. *In organizations like banks and other financial organization since there is a high accumulation of data from the day to day transactions and operations, in order to handle and also use to the benefit of the organization the importance of data warehouse is unquestionable.*
Introduction	Understanding need for the implementation of DW.	3. How do you evaluate the need for the implementation of DW in the Bank?	A. *As I indicated previously Banks have the ability to collect a huge amount of data from different channels. Data warehousing technology mainly benefits banks.*
Introduction	DW Understanding with respect to the Bank	4. Why do you think the implementation of DW is essential for the Bank especially at this moment?	*A. Considering the current market competition, it is very important to use the historic data of the bank in order to take advantage of the growing market competition through informed decision using data analytics.* *B. As the current Banking sector is very vulnerable for fraud the Bank needs to implement current technology and analyze the data of suspicious intrusions using system logs.*

			C. The huge amount of data collected from the different system is becoming a liability in terms of storage space, process to keep safe and secure the data and human resource and the bank need to change this into opportunity. *D. Moreover the bank should be in track of implementing new IT technology in order to be competitive.*
Key discussion Point	Current DW implementation status	5. How many database system are there in the bank currently	*E. Currently there are around 32 different database systems available in the bank used by the different departments and branches.*
Key discussion Point	Current DW implementation status	6. How many data source /operational databases/ are involved?	A. Currently only one source databases are involved. B. There are also one other database ongoing to be integrated.
Key discussion Point	Current DW implementation status	7. What are the source data databases?	A. Although CBS (core banking system) is the most involved source database.
Key discussion Point	Current DW implementation status	8. How do you evaluate the current status of data integration in the Bank? a. What do you mean by small scale? b. Why is it incorporated few source databases?	A. The bank is already started integrating its operational data in small scale to satisfy the increasing report need of the branches. a. I said Small scale because the current integration incorporated few source databases for report generation only. b. Because of the capacity of the servers and the human resource allocated for the project and the urgency for the request for reports from the end users side. B. The bank planned to incorporate more and more source databases step by step after evaluating of the current implementation.

Key discussion Point	Current DW implementation status	9. How is the data transformation from the data source working?	A. Since the currently the project is to satisfy the report need of the different departments and branches the data transformation is done with scheduled jobs, stored procedures using SQL DML.
Key discussion Point	Current DW implementation status	10. Who are the end users of the report generated by the system? a. Can you give us some example departments which actively use these reports?	A. The end users are the different branches and departments of the Bank B. The different CEOs and CIOs of the Bank are also requested different generalized reports a. Apart from Branches, HR, Digital Business End Users (Merchants) and Marketing department are few of them.
Key discussion Point	Current DW implementation status	11. How do you collect user requirement? a. Did you implement all requests or select some of them? If you select how was the selection process?	A. During the early stage of the project we conducted a requirement collection survey on all of our branches and departments to identify the various report needs of Branches and departments, during that time the Banks Portal was very instrumental in collecting the requirements. a. We were able to collect more than 700 report requests from Branches and departments most of them are daily base report requests. b. Considering the limited capacity and performance of our server's we couldn't satisfy all the report needs of our users. c. The selection process was based on priority and request Evaluation, hence we first send a message for each department and braches to resend their requirements with

			priorities and then we evaluated each of the prioritized requests which can be accommodated in the project with the currently incorporated source databases and tables and views in the database.
Key discussion Point	Current DW implementation status	12. Are the project deliverables clearly identified?	A. Yes, after the requirement collection is done the deliverables are designed in consideration with the requirements.
Key discussion Point	Current DW implementation status	13. What is the BI tool used? a. How the end users access the system for report?	A. The system uses oracle BIP (Business intelligence publisher) reporting tool to generate the different reports. a. The oracle BIP reporting tool Has three layers' Administrative layer for system administrators, Report authoring layer to design the required reports and presentation layer for end user to access the reports designed in Report authoring layer, hence end used this presentation layer to access their reports.
Key discussion Point	Current DW implementation status	14. Is historic data analytics incorporated?	A. As we discussed before because of the server's capacity and the urgency of the request we are not able to include historical data to the project. B. let alone the historical data we are not able to incorporate other operational databases sources that should be included as it takes longer time and resource to prepare and process the data.
Key discussion Point	Current DW implementation status	15. Which data sources (operational databases) that should be included but not included currently?	A. Reconciliation databases from finance department and HR databases are the one which should be involved but not.

Key discussion Point	Current DW implementation status	16. What are the BI tools used? a. How the end users access the system for report? Rough	A. The system uses oracle BIP (Business intelligence publisher) reporting tool to generate the different reports. a. The oracle BIP reporting tool Has three layers' Administrative layer for system administrators, Report authoring layer to design the required reports and presentation layer for end user to access the reports designed in Report authoring layer, hence end used this presentation layer to access their reports.
Key discussion Point	Current DW implementation status	17. How many table and views involved?	A. The current implementation More than 2000 tables are involved
Key discussion Point	Current DW implementation status	18. How many reports are generated?	A. As per the analysis made before one month currently more than 1700 reports are generated for the different departments and branches
Key discussion Point	Current DW implementation status	19. What is the report generation frequency? Daily/weekly/monthly/quarterly/annually	A. The reports are generated as per the request submitted in different time range, there are reports generated daily, there are reports generated weekly, and there are reports generated monthly and yearly.
Key discussion Point	Current DW implementation status	20. How is the document creation work for end users Excel/PDF/HTML	A. The reports are provided to the end users through the available formats of Oracle BIP which includes Microsoft excel, PDF, HTML and CSV. B. The report can also be generated with Microsoft word as the need arises.
Key discussion Point	Current DW implementation status	21. How do you evaluate the current data warehouse initiative of the bank with respect to the business needs?	A. Since currently any data analytics is not done by the system we cannot say the current implementation satisfy the business needs of the bank.

			B. Since the definition of business needs is gaps between the current state of the company and its goals, it is not near to the satisfaction of its business needs.
Key discussion Point	Current DW implementation status	22. Have business requirements been satisfied by the current implementation?	A. No, b/c the current implementation is working on only in satisfying the report needs of the bank but a data warehouse is much more than generating reports. B. The business requirements needs the data analytics part of the data warehousing functionality to predict future business trends.
Key discussion Point	Current DW implementation status	Are the right tools deployed and effectively used to support environment specific needs? • Consider tools used for ETL, • performance and usage monitoring, • metadata management, and • Information access, delivery, and analysis.	A. Oracle BIP is used to support the ETL process. B. As part of oracle Business Intelligence & Data Warehouse package Oracle BI Publisher is flexible and powerful tool which can also be used for future implementation.
Key discussion Point	Current DW implementation status	23. Does the technical architecture provide the right range of data delivery services? (e.g. report publishing; queries; ad hoc querying; analytical models, etc.)	C. The oracle BIP reporting tool Has three layers' Administrative layer for system administrators, Report authoring layer to design the required reports and presentation layer for end user to access the reports designed in Report authoring layer, hence end used this presentation layer to access their reports.
Closing		24. If there is someone who want to add some point on our discussion or if you have comments, you are all welcome.	