CHAPTER 1

INTRODUCTION

1.1. Introduction

This chapter introduces the background, problem statement, research objectives, proposes solution, scope and describes the thesis organization.

1.2. Research Background

Information assets are immensely valuable to any enterprise, and because of this, these assets must be properly stored and readily accessible when they are needed. However, the availability of too much data makes the extraction of the most important information difficult, if not impossible (Golfarelli ,Rizzi ,2009).

The term "Data Warehouse" was first coined by Bill Inmon in 1997. According to Inmon(1996) a data warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data. This data helps analysts to take informed decisions in an organization.

Data warehousing is a phenomenon that grew from the huge amount of electronic data stored in recent years and from the urgent need to use that data to accomplish goals that go beyond the routine tasks linked to daily processing. In a typical scenario, a large corporation has many branches, and senior managers need to quantify and evaluate how each branch contributes to the global business performance. The corporate database stores detailed data on the tasks performed by branches. To meet the managers' needs, tailor-made queries can be issued to retrieve the required data (Golfarelli,Rizzi, 2009).

Many years ago, database designers realized that such above approach is hardly feasible, because it is very demanding in terms of time and resources, and it does not always achieve the desired results. Moreover, a mix of analytical queries with transactional routine queries inevitably slows down the system, and this does not meet the needs of users of either type of query. Today's advanced data warehousing processes separate online analytical processing (OLAP) from online transactional processing (OLTP) by creating a new information repository that integrates basic data from various sources, properly arranges data formats, and then makes data available for analysis and evaluation aimed at planning and decision-making processes (Lechtenbörger, 2001).

1.3. Problem Statement

The problem that is to be addressed is "how data warehousing is support the decision making in sales". To address this problem narrowed the scope of the research to an investigation focusing on paint sales data mart. The major causes for the problem are the database management systems in branches are separated .That mean:

I. Small number of data used for support of crucial paint sales.

II. Similar process is very slow in operational database, as the process of query writing and interpreting them are durable.

III. An expert in information technologies and the mangers think in different categories and as a corollary tend not to understand each other .mangers need clear information to make decisions in paint sales.

IV. Lack of time for manager to find significant sales numbers.

1.4. ResearchObjectives

1. To build and analysis the data warehouse in paint sales.
2. To evaluate the ability to make quick and better decision about the paint sales.
3. To use integrated views of company historical data to optimize the decision.
4. To provide accurate reports from paint sales data warehouse.
5. To forecast future sales.

1.5. Propose Solution

Build data warehouse to copy data from the multiple source and integrate data into a single repository, so a single query engine can be used to present data,build dimensional model to Analyze data which in data warehouse to support decision making processing with historical data and data mining to predict sales in futur.

1.6. Scope

Build data warehouse prototype to organization and analyze impact data warehouse for improve decision making using historical data.

1.7. Thesis organization

This research contains five chapters. Chapter one includes introduction. Chapter two reviews the literature review. Chapter three include methodology, tools used in research, system analysis and research design. Chapter four implementation data warehouse, analysis data warehouse, and made data mining .Chapter five conclusion.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews the literature review, background theory and related work.

2.2. Background Theory:

2.2.1. The Data Warehouse Concept:

Data warehousing technology aims to structure the data in a appropriate way to access the data, and use it in an efficient and effective manner (Diaset al., 2008). The data warehouse is responsible for the consistency of information. The integration of tools such as query tools, reporting tools and analysis tools provide opportunity to handle the coherence of information. The aim of data warehousing is to organize the gathering of a wide range of data and store it in a single repository (Kerkri et al., 2001). Currently, data warehousing plays a major role in the business community at large. Furthermore, integrating data from the different sources and converting them into valuable information is a way to obtain competitive advantage (del, Lees, 2002).Data warehousing is “a collection

of decision support technologies aimed at enabling the knowledge worker (executive, manager, analyst) to make better and faster decisions" (Chaudhuri,Dayal, 1997) . According to Inmon (2005), data warehouse is a "subject-oriented, integrated, time-variant and non-volatile collection of data in support of management decisions". March and Hevner (2007) Argued that the three components of intelligence namely understanding, adaptability and profiting from experience are important considerations when designing the data warehouse. Also, these authors mentioned that the data warehouse should allow managers to gather information such as identifying and understanding different situations and the reasons for their occurrence. Further, they have argued that the, data warehouse should "enable a manager to locate and apply the relevant organizational knowledge and to predict and measure the impact of decision over time" (March, Hevener, 2007). However, as mentioned by March and Hevner (2007),these arguments forms the challenges that need to be consideredwhen implementing a data warehouse.

2.2.2. MainComponents Of The Data Warehouse

According to Kimball and Ross (2002), a few components can be identified to form the data warehouse environment (Figure 2.1). Each component of the data warehouse provides a specific function. The main components are:

- Operational source system
- Data Staging Area
- Data Presentation Area
- Data Access Tools

i. Operational Source Systems

The Operational source system is mainly concerned about processing performance and availability. Generally, the source system maintains a small amount of historical data. The queries designed against source systems are narrow. On the other hand, one-record-at-a-time queries which operate as part of the normal transaction flow and act according to the demands on the operational system (Kimball, Ross, 2013).

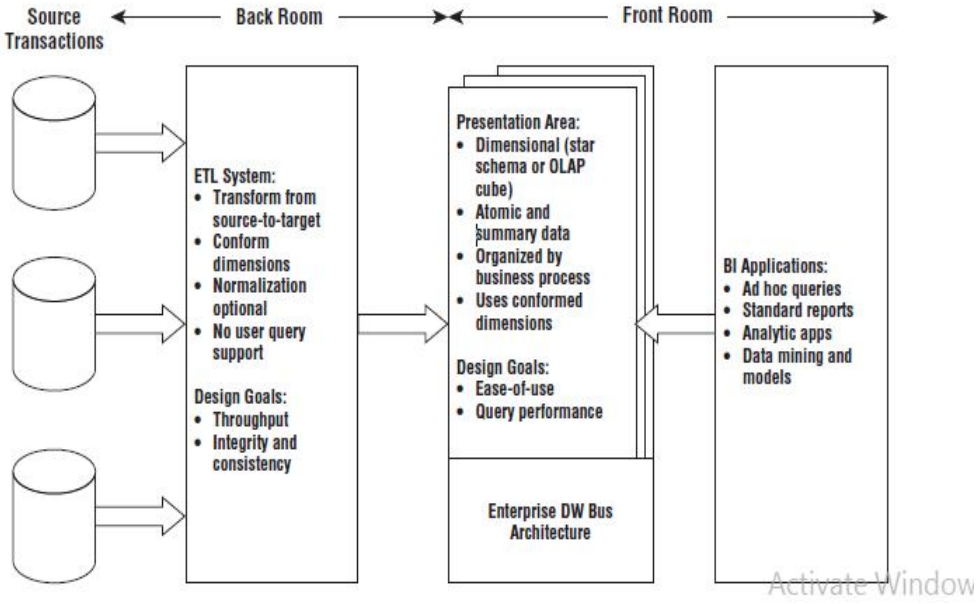


Figure (2.1): Components of the Data Warehouse

ii. Data Staging Area

The data staging area is the place that keeps the data as temporary storage (Kimball, Ross, 2002). Also, this area is known as the Extract Transformation Load (ETL) because it is conducting the data extraction, transformation and loading. In other words, the data staging area can be referred to as everything between the operational source systems and the

data presentation area (Kimball, Ross, 2002). The first process of transferring data to the data warehouse is extraction. During this process it is important to read and understand the source data and copy them to the staging area of the data warehouse for further management. After extracting the data to the staging area many alterations such as cleansing the data (correcting misspellings, resolving domain conflicts, dealing with missing elements, or parsing into standard formats), combining data from multiple sources, deduplicating data, and assigning warehouse keys take place .Then the load the data to the presentation area of the data warehouse (Kimball, Ross, 2002).

iii. Data Presentation

The data presentation area is the place where data is organized, stored, and made available to the users. In addition, the data presentation area is the place where business communities see data and gain access using data access tools. As stated by Kimball and Ross (2002), this area can be referred as series of integrated data marts. A each of this data mart presents the data from a single business process.

iv. Data access Tools

The data access tools element is the final element of the data warehouse. This element provides many capabilities for the business users to control the presentation area for analytic decision-making. Generally, the data access tool can act as a simple query tool or can be complex as a data mining application (Kimball and Ross, 2002).

A DB is simply a collection of planned information, usually as a set of related lists of similar items. The data is often structured so that it is easily manageable. For example, a school DB would have several table as

teachers, students, and classes where each table would have records that specify information about each item. A DB often involves a software system called Database Management System (DBMS) that is responsible for storing and managing the data in the DB. MySQL, Oracle, Microsoft SQL Server are some well-known DBMS.

A data warehouse is a special type if DB used for analysis of data. It collects and stores integrated sets of historical data from multiple operational systems and feeds them to one or more data marts. It may also provide end-user access to support enterprise views of data.

2.2.3. Key DifferenceComparison Between Database and Data Warehouse:

This table (2.1) compare database and data warehouse in terms of the type of data stored, the purpose of data and the method of uses data.

Table (2.1): Comparison between Database and Data Warehouse

Database	**Data Warehouse**
Stores current data	Stores historical data
Data cannot be used for analysis or reaching decision	Extracts data and reports them to analyze and reach decisions.
provides a detailed relational view	Provides a summarized multidimensional view.
Database can do a lot of concurrent transactions	Data warehouse is not designed for such tasks.
used for Online Transactional Processing	used for Online Analytical Processing
Tables in a database are normalized to achieve efficient storage	Tables are usually demoralized to achieve faster querying.

2.2.4. Data Warehouse Modeling:

In the data warehouse, after the business queries and subject area have been identified the information stored in the data warehouse/data mart is designed (Borysowich, 2007). Designing the data warehouse/data mart structure is different from designing the operational systems. According to Mohania, Samtani, Roddick and Kambayashi (2007), operational systems consist of simple pre-defined queries. On the other hand, in data warehousing environments queries join with more tables and more computation time and informality (Mohania et al., 2007).

2.2.5. Star schema

Star schemas are dimensional structures deployed in a relational database management system (RDBMS). They characteristically consist of fact tables linked to associated dimension tables via primary/foreign key relationships. The star schema gets its name from the physical model's resemblance to a star shape with a fact table at its center and the dimension tables surrounding it representing the star's points(Kimball, Ross, 2013).The star schema separates business process data into facts, which hold the measurable, quantitative data about a business, and dimensions which are descriptive attributes related to fact data. Examples of fact data include sales price, sale quantity, and time, distance, speed, and weight measurements. Related dimension attribute examples include product models, product colors, product sizes, geographic locations, and salesperson names.as shown in Figure (2:2)

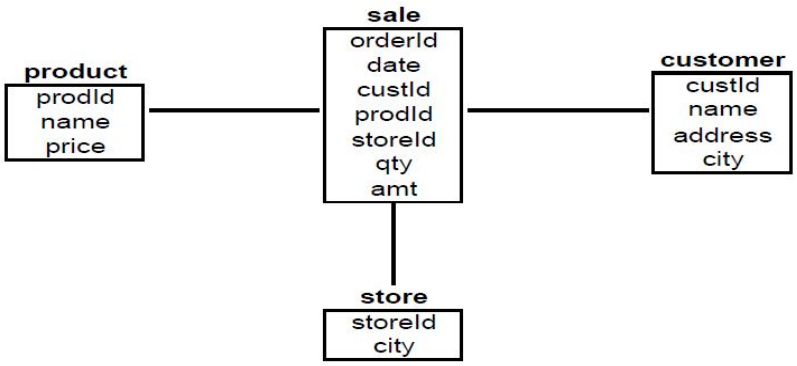


Figure (2:2): Star Schema

2.2.6. Dimensional modeling:

Dimensional modelingnames a set of techniques and concepts used in data warehouse design. It is considered to be different from entity relationship model. Dimensional modeling does not necessarily involve a relational database. The same modeling approach, at the logical level, can be used for any physical form, such as multidimensional database or even flat files. According to Kimball and Ross (2013), Data warehousing consultant, Dimensionalmodeling is a design technique for databases intended to support end-user queries in a data warehouse. It is oriented around understandability and performance. According to him, although transaction-oriented ER is very useful for the transaction capture, it should be avoided for end-user delivery.

Dimensional modeling always uses the concepts of facts (measures), and dimensions (context). Facts are typically (but not always) numeric values that can be aggregated, and dimensions are groups of hierarchies and descriptors that define the facts. For example, sales amount is a fact;

timestamp, product, register#, store#, etc. are elements of dimensions. Dimensional models are built by business process area, e.g. store sales, inventory, claims, etc. Because the different business process areas share some but not all dimensions, efficiency in design, operation, and consistency, is achieved using conformed dimensions, i.e. using one copy of the shared dimension across subject areas. The term "conformed dimensions" was originated by Kimball and Ross (2013).

2.2.7. Fact tables

Fact tables record measurements or metrics for a specific event. Fact tables generally consist of numeric values, and foreign keys to dimensional data where descriptive information is kept. Fact tables are designed to a low level of uniform detail (referred to as "granularity" or "grain"), meaning facts can record events at a very atomic level. This can result in the accumulation of a large number of records in a fact table over time. Fact tables are defined as one of three types:

- Transaction fact tables record facts about a specific event (e.g., sales events)
- Snapshot fact tables record facts at a given point in time (e.g., account details at month end)
- Accumulating snapshot tables record aggregate facts at a given point in time (e.g., total month-to-date sales for a product)Fact tables are generally assigned a surrogate key to ensure each row can be uniquely identified. This key is a simple primary key.

2.2.8. Dimension tables

Dimension tables usually have a relatively small number of records compared to fact tables, but each record may have a very large number of attributes to describe the fact data. Dimensions can define a wide variety of characteristics, but some of the most common attributes defined by dimension tables include:

Time dimension tables describe time at the lowest level of time granularity for which events are recorded in the star schema Geography dimension tables describe location data, such as country, state, or city Product dimension tables describe products Employee dimension tables describe employees, such as sales people Range dimension tables describe ranges of time, dollar values, or other measurable quantities to simplify reporting Dimension tables are generally assigned a surrogate primary key, usually a single-column integer data type, mapped to the combination of dimension attributes that form the natural key(Kimball, Ross, 2002).

2.2.9. Data Marts

A data mart and a data warehouse have different architectural structures. On some occasions there is a need to perform a standardized data analysis and organizing data to identify simple usage patterns. As a result of this, data warehousing is arranged in to small units called data marts (Bonifati et al., 2001). As mentioned by Inmon (1999), “a data mart is a collection of subject areas organized for decision support based on the needs of a given department”. Therefore, each department has its own way of understanding how the data mart should look. Each data mart is designed according to the department’s needs.

2.2.10. Data warehouse VS Data mart:

This table definesthe Differences betweenData warehouse VS Data mart.

Table (2.2): Differences between Data warehouse VS Data mart.

Data warehouse	Data mart
enterprise-wide data	department-wide data
multiple subject areas	single subject area
difficult to build	easy to build
takes more time to build	less time to build
larger memory	limited memory

2.2.11. Datamining

The objective of any data mining process is to build an efficient predictive or descriptive model of a large amount of data that not only best fits or explains it, but is also able to generalize to new data (Mukhopadhyay et al., 2014).Based on a broad view of data mining functionality, data mining is the process of discovering interesting knowledge from large amounts of data stored in either databases, data warehouses, or other information repositories. Data mining refers to extracting or “mining” knowledge from large amountsof data (Jiawei Han, Micheline,2006).

2.2.12. MicrosoftTime Series Data Mining Algorithm

The Microsoft Time Series algorithm provides regression algorithms that are optimized for the forecasting of continuous values, such as product sales, over time.Whereas other Microsoft algorithms, such as decision trees,

require additional columns of new information as input to predict a trend, a Time Series model does not. A Time Series model can predict trends based only on the original dataset that is used to create the model. New data can be added to the model making a prediction and automatically incorporate the new data in the trend analysis.

Another unique feature of the Microsoft Time Series algorithm is that it can perform cross prediction. The algorithm can be trained with two separate, but related, series, and the resulting model created can predict the outcome of one series based on the behavior of the other series. For example, the observed sales of one product can influence the forecasted sales of another product. The Microsoft Time Series algorithm uses both methods, ARTXP (Autoregressive Tree Models with Cross Prediction) and ARIMA (Autoregressive Integrated Moving Average), and blends the results to improve prediction accuracy. The ARTXP algorithm can be described as an autoregressive tree model for representing periodic time series data. The ARIMA algorithm improves long-term prediction capabilities of the Time Series algorithm (Ankuret al., Chandra).

2.2.13. Data Mining Extensions (DMX)

Data Mining Extensions (DMX) is a language that you can use to create and work with data mining models in Microsoft SQL Server Analysis Services used to create the structure of new data mining models, to train these models, and to browse, manage, and predict against them. DMX is composed of data definition language (DDL) statements, data manipulation language (DML) statements, and functions and operators (Lynn Langi,otger, 2008).

2.3 Related Studies :

There was a research entitled (Design and Implementation of an Enterprise Data Warehouse).The problem of this research is that the information is available in some isolated databases. The proposed solution is to collect and integrate that information into a single repository. The motivation for this research was to come up with better reports because they were coming out from one warehouse. The Research involves a description of data warehousing techniques, design, and extracting from transactional databases. The Research also discusses how the data from databases and other data warehouses could integrate. Separately, an important piece of this research takes an actual example of data and compares the performance between them by running the same queries against separate databases, one transactional and one data warehouse. As the queries expand in difficulty, larger grows the gap between the actual recorded times of running that same query in the different environments (Edward, 2011).

As there was another study conducted byTemitopeAdeoye and RaufuOlalekan.This thesis seeks to develop DW and BI system to support the decision makers and business strategist at Crystal Entertainment in making better decision using historical structured or unstructured data. This thesis shown that the data warehouse collect, consolidate, organize, and summarize this structured and unstructured data.so, that this intelligent data can be used to inform business decisions.The thesis focused on designing and implementation of Data Warehouse and business intelligent system for retail. The research has shown how data can be integrated from different sources to data warehouse which is used for delivering business intelligence to the end users and executives. The research have developed data analysis template that user can interact with to get an immediate answers to the business questions. The reports can be generated with click of a button (Temitope ,Raufu ,2011).

In the field of health provided this studyconducted byPubudikaKumari (2011).This thesis used the cardiac surgery unit at the Prince Charles Hospital (TPCH) as case study. The cardiac surgery unit at Hospital uses a stand-alonedatabase of patient clinical data, which supports clinical audit, service management and research functions. The main objectives of this research are to improve access to integrated clinical and financial data, providing potentially better information for decision-making.Difficulty of integrating data from otherdatabase was identified as problem. The researcher proposed building data warehouse prototype based on the cardiac surgery unit needsand integrating the databases in one repository. The methodology was used in this research consisted of survey methods (questionnaire and unstructured interviews).This study comes to, implementing centralized data warehouse will minimize the issues faced in the current decision-making process and also, providemany benefits such as improved access to data, improved quality and safety monitoring, provide data for clinical effectiveness and evaluation research and improved decision-making (Pubudika,2011).

In the side of data mining there was a study on the time series algorithm,this Study conducted byKaziand Mohammad. The purpose of this research is to propose an estimation of possible sales to assist Telecom Sales Management. Sales Management is always critical in broadly liberalized especially in telecom market. In order to keep the competitive advantage in telecom market, mobile service providers must be able to predict the rate of New Subscriber Activations. In this research paper, the researchers utilized Microsoft Time Series algorithm with SQL server to predict New Subscriber Activations and proved how the using of this model can predict the activation of new subscribers in upcoming days (Kazi, Mohammad, 2013).

The last study that was reviewed in this research is the study carried out byMahesh Kumar Yadav (2016).The problem of the Study is Organizations have huge amounts of data but have found it increasingly

difficult to access it and make use of it. This is because, it is in many different formats, exists on many different platforms, and resides in many different file and database structures developed by different vendors. The solution of the problem of this study is building data warehouse for e-Government system that can help to predict overall organization Performance and activity for making healthier decision for government activities through ICT.The methodology of this study focus on different theoretical and implemented architectures, which architecture suits the best for e-Gov system in Nepal. This study comes to, The use of a Data Warehouse allows the Ministries executives to use information in making appropriate decisions making process , planning process and deriving the compared results. About 16% of the total ministries are planning to implement the warehousing system in the future for e-Government system(Mahesh Kumar Yadav et al., 2016).

According to previous studies which mentioned above .So we have built the data warehouse to data gathering, integrated it for several sources after data cleaned. Addition to that, data were analyzed and come out reports.

2.4. About Company and Data Set:

National Paints Factories was founded in 1969 in Amman-Jordan. In 1977, the company established the first paint factory in the U.A.E. Sharjah , now factories are also established in Qatar, Sudan, Oman, Egypt, Kyrgyzstan Kazakhstan, Romania, Palestine, Russia and India .In 2013 established a state of art manufacturing plant in UAE Capital Abu Dhabi.

We used the data of sales department in Sudan for research purposes. We used the data of two (Khartoum and Omdurman).

CHAPTER 3

RESEARCH METHODOLOGY

3.1. Introduction

This chapter include Research Methodology (Research Design) and Tools used in research.

3.2.Research Design

For the purpose of this research, the researcher used SQL server 2014 which is the product in Microsoft technology that support the enterprise data warehouse and business intelligence. The application has the relational database management system that is capable of storing all the data required for the data warehouse. It has the functionality that can extract data from different sources and consolidate it into one single location for better analysis. This is known as the integration services (SSIS). After integration and build data warehouse we used SSAS to analyze and make sense of information. We used SSAS to create cubes using data from data marts (data warehouse) for deeper and faster data analysis and mining in data warehouse to predict sales in the future by using Time Series Algorithm. The front end application for this research would be the Reporting services (SSRS). It was used to design reports according to users.Thefigure (3.1) below shows this:

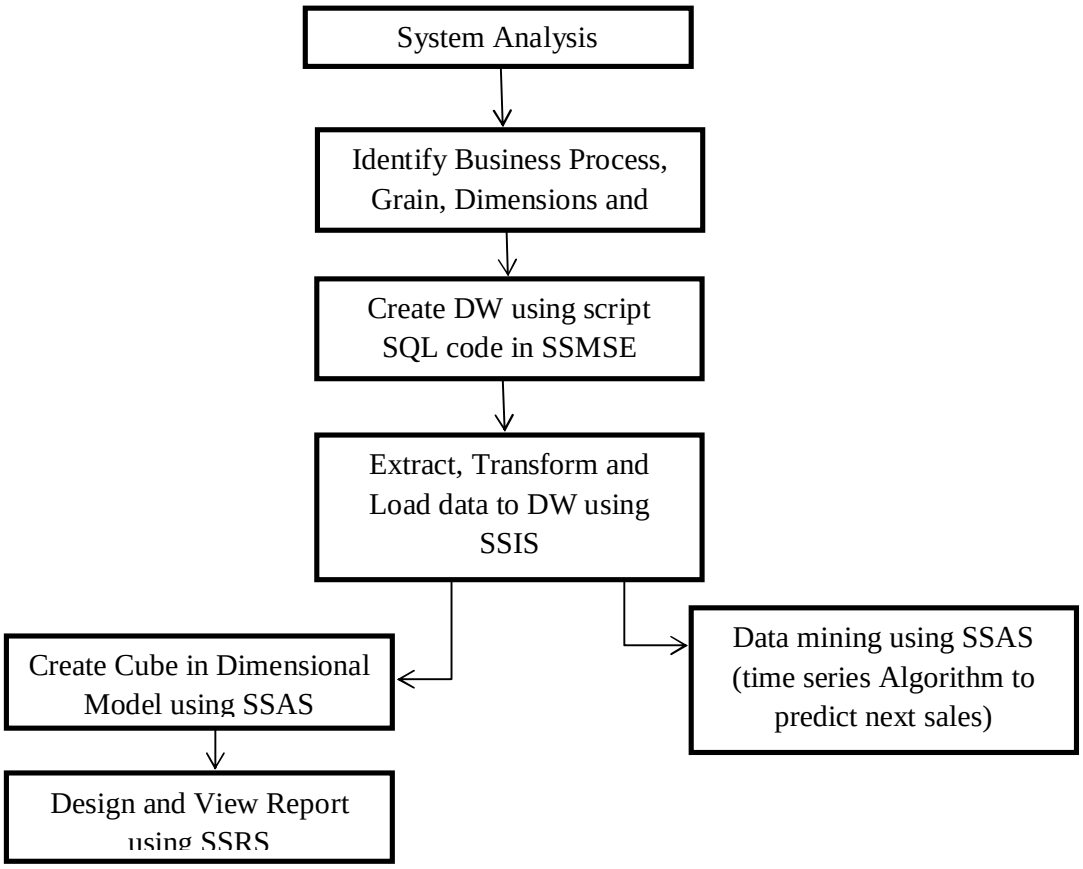


Figure (3.1):Research Design

3.2.1. SystemAnalysis

Analysis involved a detailed study of the current system, leading to specifications of a new system. During analysis, we studied the activities of the company and we choose one department (Sales) and two locations (Khartoum and Omdurman) to design the data mart and data were collected on the available files and database. Interview is the tool used for collect the requirement and system analysis.

3.2.2. Preparation to Designof a dimensional model

According to (Kimball, Ross, 2013) the design of a dimensional model by consistently considering four steps:

Step 1: Select the Business Process

Select paints sales in two locations.

Step 2: Declare the Grain

One row per one product sells.

Step 3: Identify the Dimensions

Date Dimension, Product Dimension and Store Dimension.

Step 4: Identify the Facts

Quantity and Sales Cost.

3.2.3. The Data Mining Process

According to (Hornick et al, 2006), the lifecycle of a data mining project consists of six different phases (Business Understanding,Data Understanding,Data Preparation,Modeling,Evaluation and Deployment). The process in general is iterative, but also foresees stepping back between certain phases to adjust some of the decisions made. From the data mining perspective, the (business) user is mainly involved in the phases Business Understanding and Deployment, while the other phases are mostly performed only by the data miner. In terms of integration, it has to be distinguished between the (technical)deployment of the data mining solution as a whole, which might be done only once for a given business process, and the deployment of new data mining models, which might be done frequently.

3.2.4. Data Required for Time Series Models

The requirements for a Time Series model are as follows:

- A single key time column: in this model usedyear column.
- Predictable columns: in this model used cost and quantity column.

3.2.5. How the Algorithm Works

In SQL Server 2008, the Microsoft Time Series algorithm uses both the ARTXP algorithm and a second algorithm, ARIMA. The ARIMA algorithm is optimized for long-term prediction. By default, the Microsoft Time Series algorithm uses a mix of the algorithms when it analyzes patterns and making predictions. The algorithm trains two separate models on the same data: one model uses the ARTXP algorithm and one model uses the ARIMA algorithm. The algorithm then blends the results of the two models to yield the best prediction over a variable number of time slices. Because ARTXP is best for short-term predictions, it is weighted more heavily at the beginning of a series of predictions. However, as the time slices that we are predicting move further into the future, ARIMA is weighted more heavily. We can also control the mix of algorithms to favor either short- or long-term prediction in the times series.

3.3. Tools Used in Research:

This research aims to integration data to one database to support decision making so we used SQL Server Management Studio and SQL Server Data Tools.

- **SQL Server Management Studio:**

  SSMS is tool includes both script editors and graphical tools that work with objects and features of the server. Use to create data warehouse in this research.

- **SQL Server Data Tools**

  SQL Server Data Tools is a modern development tool that use to Integration Services packages, Analysis Services data models, and Reporting Services reports. With SSDT, we design and deploy data warehouse.

CHAPTER 4

THE PROPOSED DATA WAREHOUSE IMPLEMENTATION

4.1. Introduction

This chapter includes implementation data warehouse, analysis data warehouse, and made data mining and made report to decision maker.

4.2. System Design

The study started the process of data warehouse design with the High level plan according to (Kimball, Ross, 2013) Dimensional modeling. The figure (4.1) below shows the High-Level Plan:

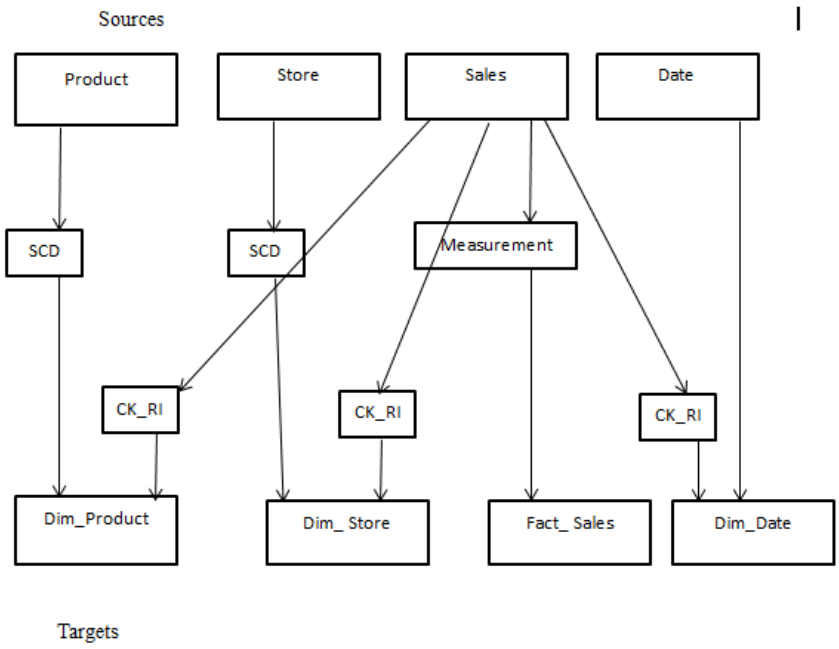


Figure (4.1): High –Level Data Staging Plan Schematic

4.3. LogicalModels

The study started the design of the data marts through the fact and dimension tables. All database design start with logical design.

4.3.1. Facts and Dimensions Tables

This research looked at the design of the data mart within the National Paints Factories (paint sales). The dimensional model has three dimension tables (store, date and product) and one fact table (sales fact table). Next step is to determine which column combination will uniquely identify a fact table row. This is important as it is required for both logical and physical design in order to determine the primary key. We design the

dimension table for the fact table to complete the data mart. A dimension table is a table that contains various attributes explaining the dimension key in the fact table. The link between the fact and dimension table is through the referential integrity. The dimension table has the primary key while the fact table has the foreign key. The former is enforced at the table level while the later can be enforcing through the ETL. Thefigure (4.2) below shows this:

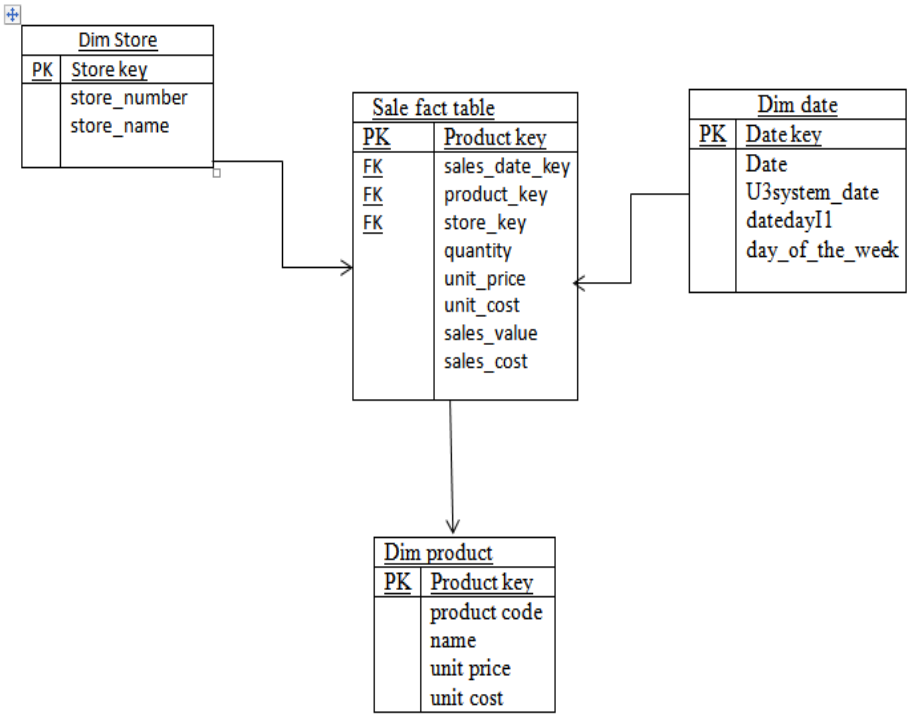
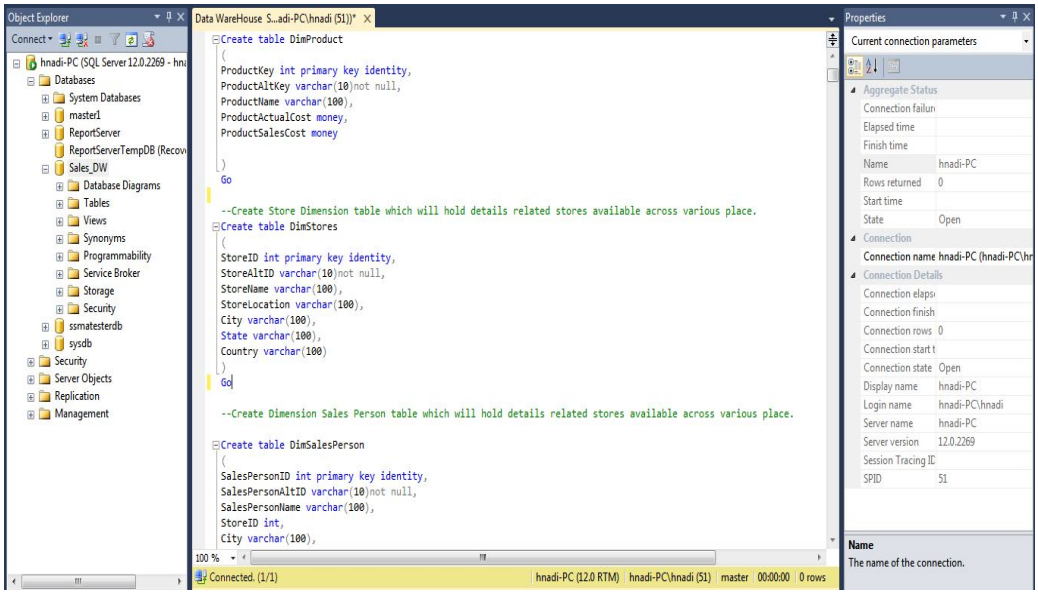


Figure (4.2):Logical Model of Dimensional Model

4.4. Design of the Physical Database

The actual design of the database is carried out by executing scripts in the management studio of SQL server 2014. after run the executing scripts, the data warehouse is built and the dimensions tables (DimStore,

DimProudct, DimDate) and the fact table (Sales) are created.Thefigure(4.3) below shows this.



The figure (4.3): The Scripts to Build Data Warehouse in SQL Server Management Studio

4.4.1. Facts and Dimensions Tables

The main key in the dimensional table are usually the surrogate key (Store key, Date key, product key), they are unique and not null, it uniquely identify 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. The diagram blow illustrates the relation between dimension tables and fact table.

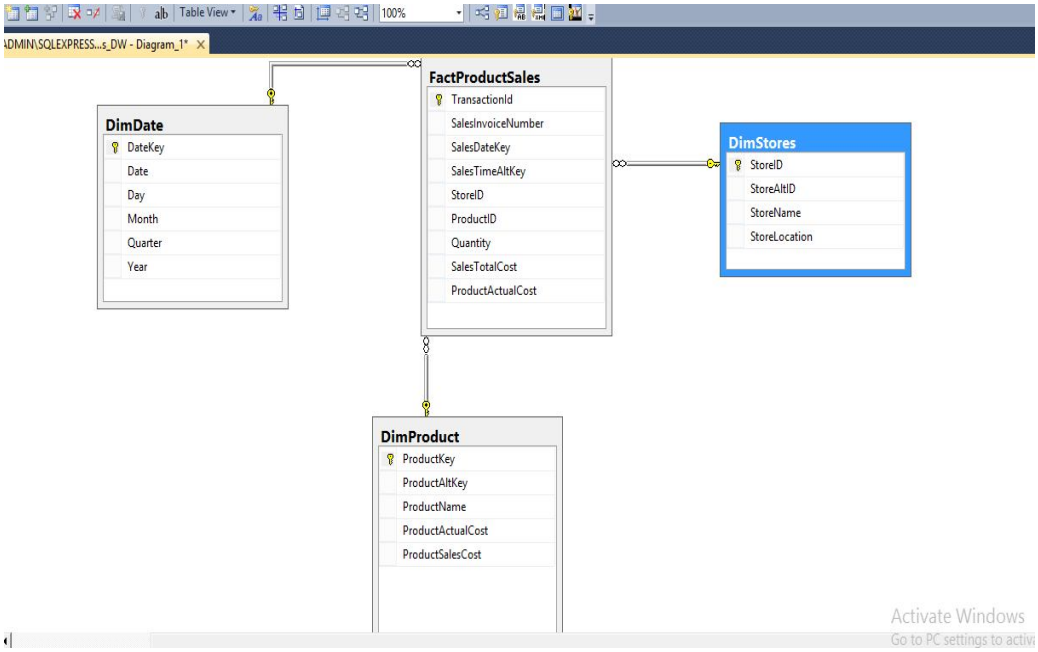


Figure (4.4): Dimension Tables and Fact Table Diagram

4.5. Design of the ETL Process

The following subsection describe the design of ETL Process

4.5.1. Dimension Table Incremental Processing:

According toKimball and Ross(2013) The DW/BI team may not be successful in pushing the responsibility for identifying new, updated, and deleted rows to the source system owners. Wherefore, we used ETL tool to do that.

In the beginning, we select the sources to extract the data. In our case the data sources are an Excel 2010 File and SQL Server. At this stage,

we select just the important columns from sources. The structure screen quality is checked by the tool. If there is any failure the process is suspended and stopped, until the correct referential integrity is confirmed.We check columns quality by replacing null value with (-1)value according to Kimball and Ross(2013), also we need to convert some data types of data to be compatible with data type of data warehouse. Then the transformation is finished. After that data is ready to next step .It is passed to slowly changing dimension (SCD) and then the data is either entered or modified. After the dimension data is properly prepared, the load process into the target tables is fairly straightforward.The following ETL design loads the data into the dimensional tables (Product, Store, and Date). The figure (4.5) below shows Extract, Transform and Load product data from excel to target warehouse.

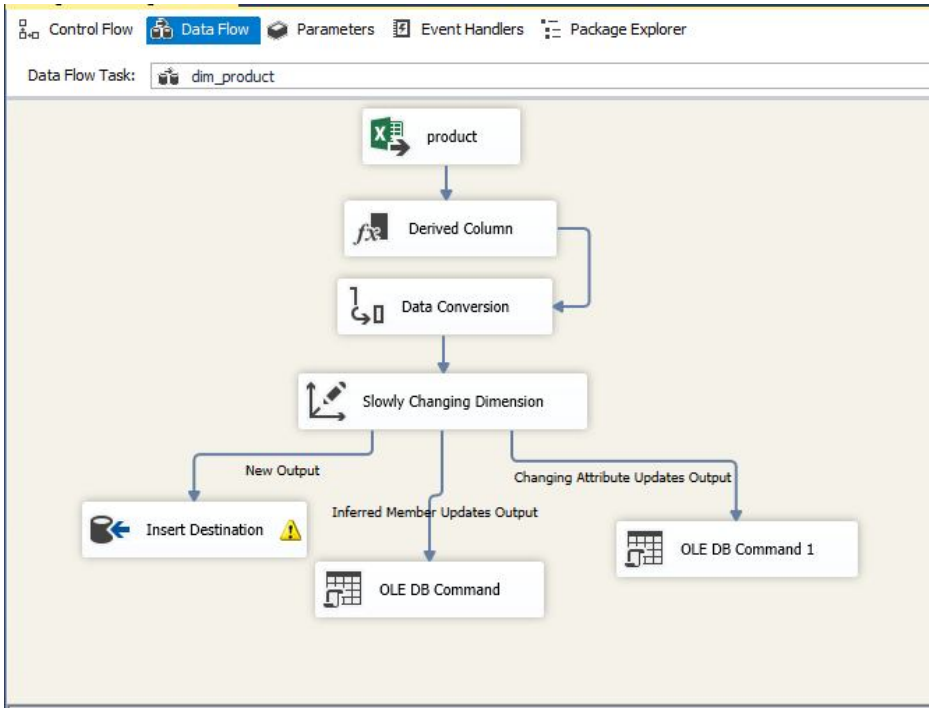


Figure (4.5): ETL Product Data

Figure (4.6) below illustrates Extract, Transform and Load Date data from excel to target warehouse.

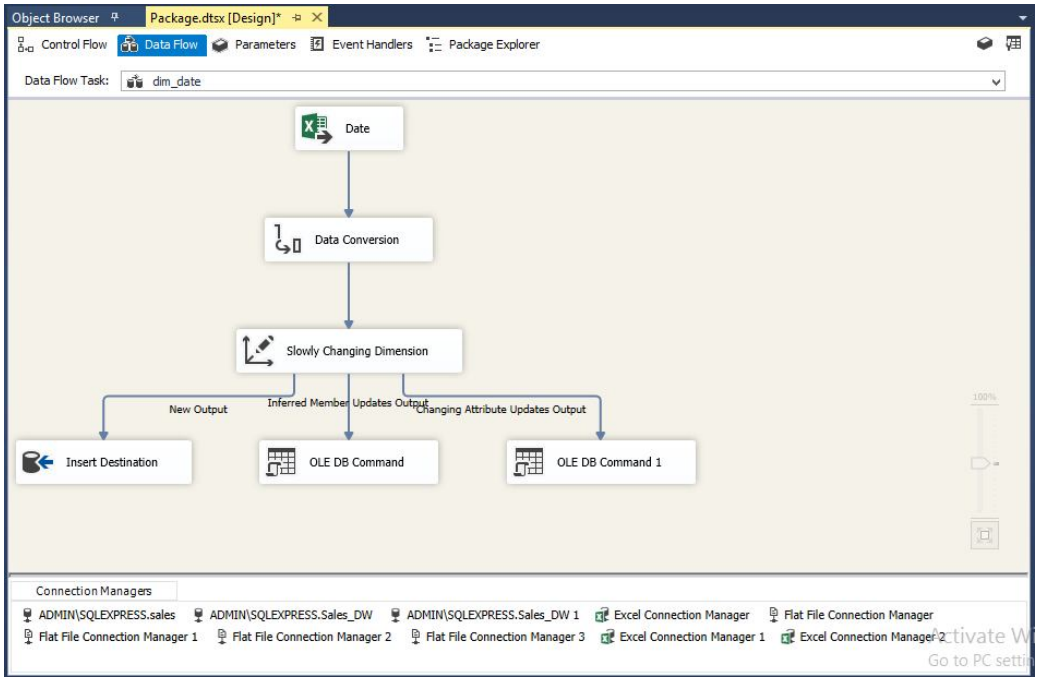


Figure (4.6): ETL Date Data

Figure (4.7) below illustrates Extract, Transform and Load Store data from SQL Server to target warehouse.

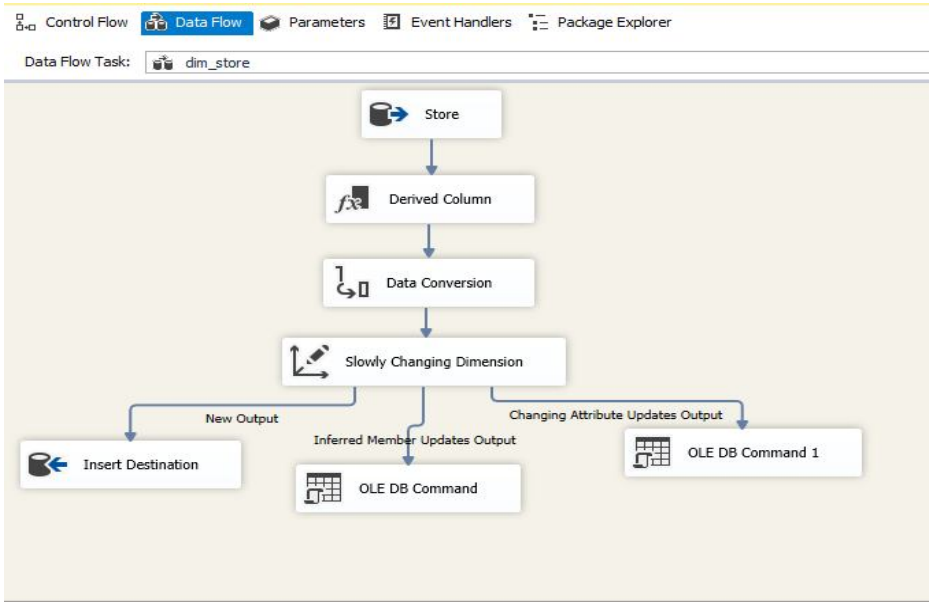


Figure (4.7): ETL Store Data

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.Whenwe load data and facts to the fact table, we use a look up operator to check the referential integrity between the fact table and dimension tables.The surrogate key pipeline is the final operation before we load data into thetarget fact table. The incoming fact data should look just like the target fact table inthe dimensional model, except it still contains the natural keys from the sourcesystem rather than the warehouse's surrogate keys. We are handling referential integrity violations by written error rows to a file for later analysis as shown in the figure (4.8) below:

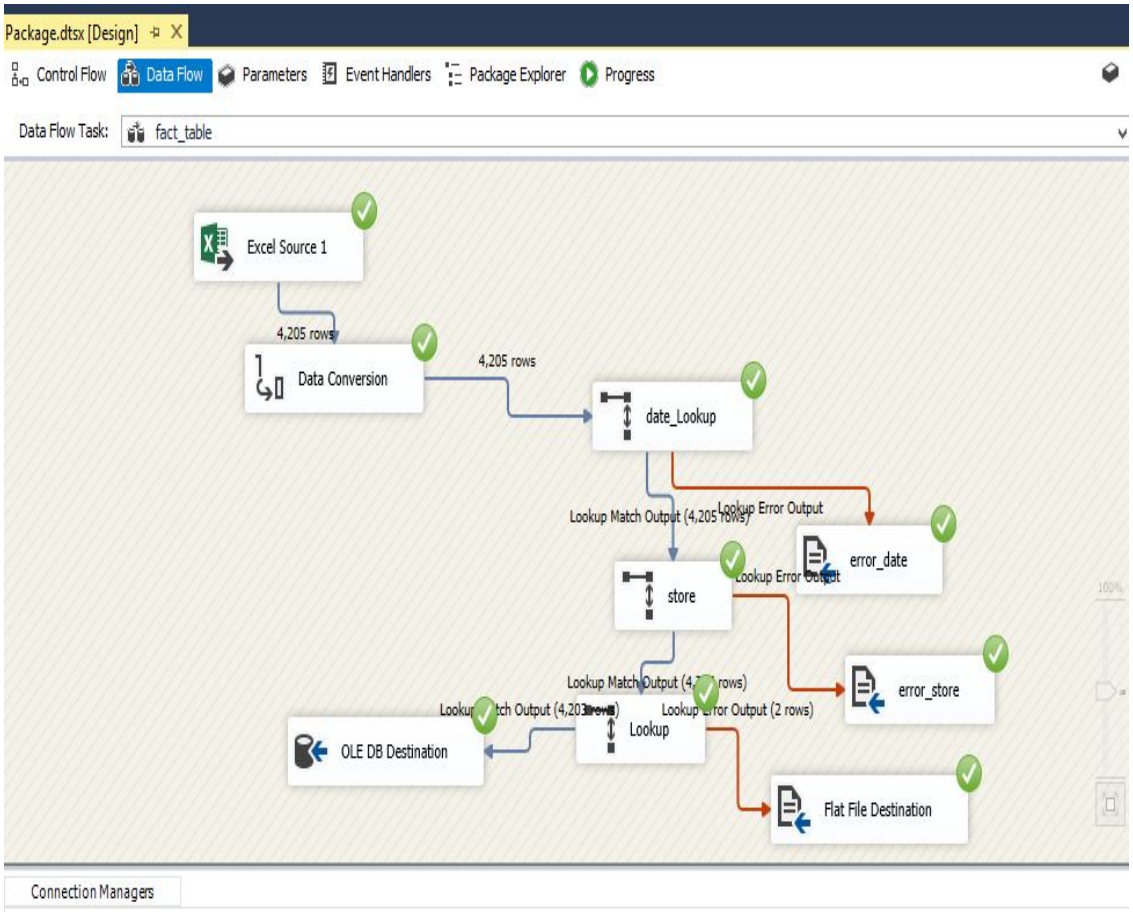


Figure (4.8): Load Fact Table

4.6. SQL Server Analysis Services (SSAS):

We chose SSAS to analysis because it's become one of the most robust data analysis platforms available for the enterprise. Building on its reputation for being very easy to start new projects in, the releases of SQL Server 2014 have shown incredible growth in functionality and scalability of the platform. To this end, many developers are choosing to base their BI solutions in SSAS because they can quickly develop and deploy their warehouse solutions.

4.7. Cube

We started SSAS with the creation of data warehouse dimensions. If we consider a dimension as a table, all the fields in this table can be perceived as attributes. We created a dimension using the three dimension tables (Product, Store and Date) which we have included in our schema. Hierarchy in a dimension is a group of attributes logically related to each other with a defined cardinality. Creating a hierarchy is as easy as dragging and dropping attributes in the hierarchy pane of the dimension editor. We want to create a hierarchy in the Date dimension.The figure (4.9) below illustrates browsing dimension of Store after we build it.

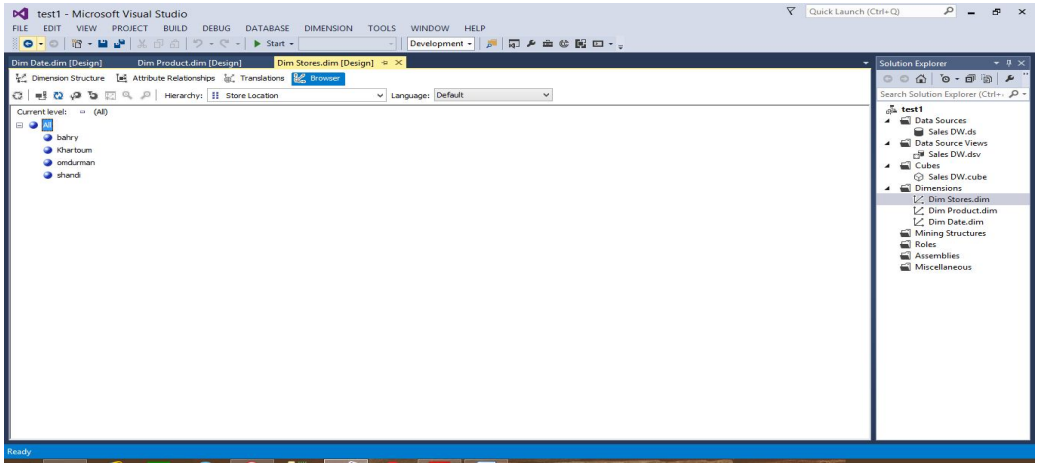


Figure (4.9): Browsing Dimension of Store

The figure (4.10) below illustrates how can drill down and drill up Date Dimension.

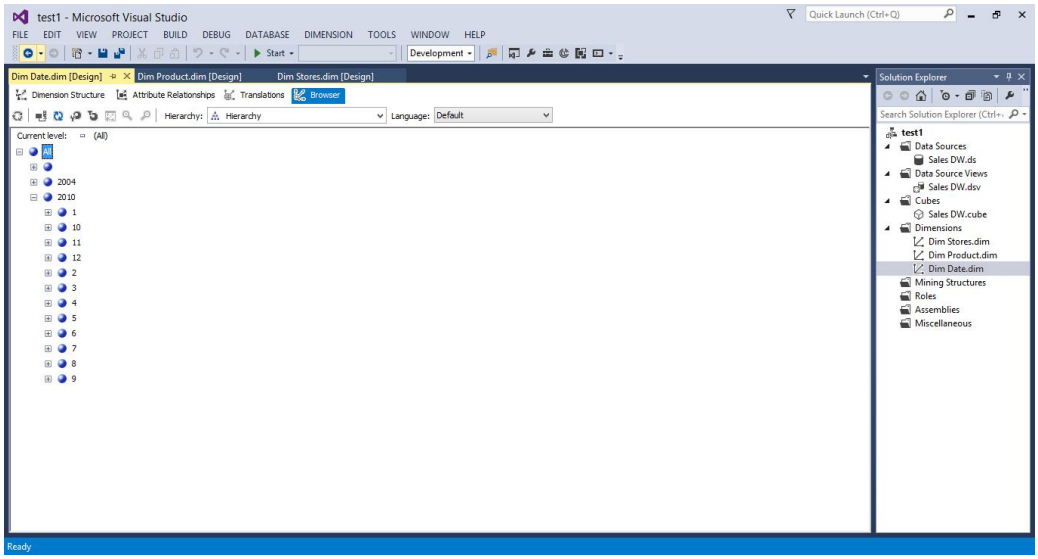


Figure (4.10): Browsing Dimension of Date

The figure (4.11) below illustrates how can drill down and drill up Product Dimension.

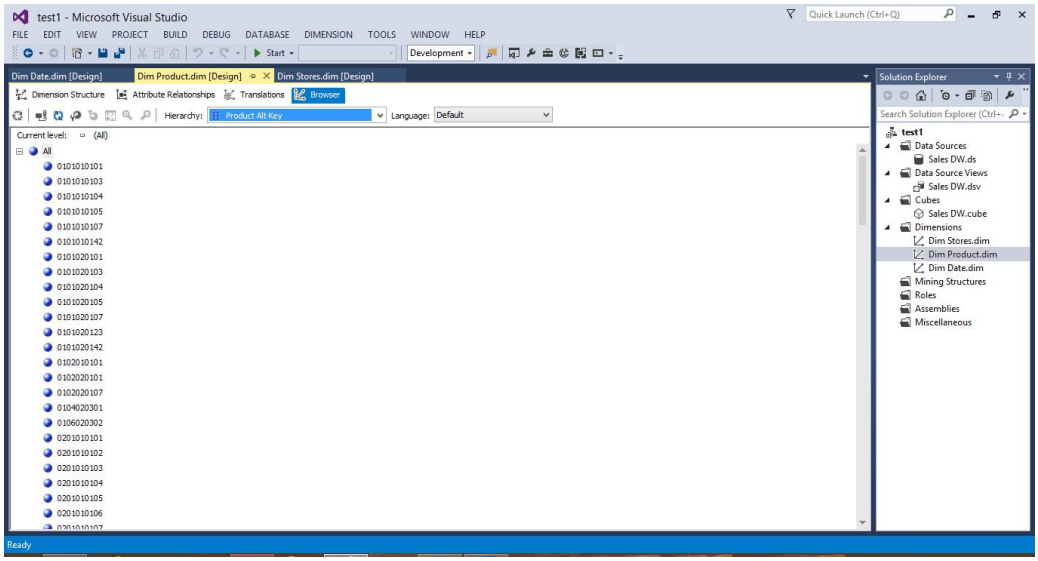


Figure (4.11): Browsing Dimension of Product

Finally we create a cube using the dimensions we just developed, and fact tables to create dimensions (cube dimensions) and measure groups (from fact tables).The figure (4.12) below illustrates the cube.

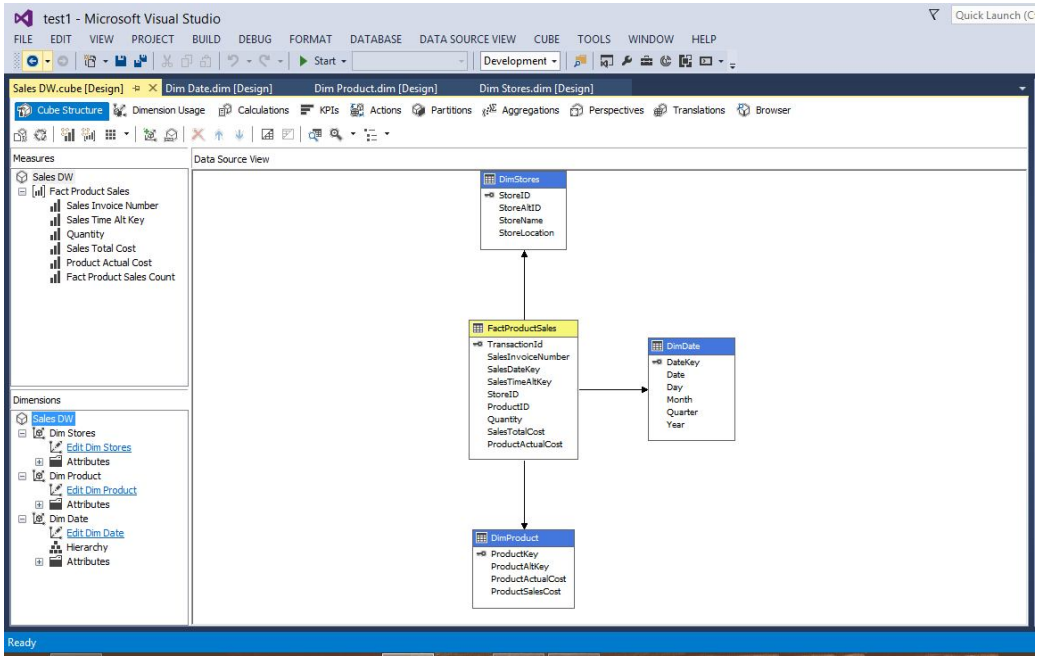


Figure (4.12): Cube Structure

We calculated measure in calculations tab in cube window. When measure groups are created in a cube, one measure group is created per fact table. Often in production systems, developing calculated measures is a regular requirement.Measures created directly from the fields of a fact table are called base measures. But often we require measures based on custom requirements, so we apply some logic and/or formula to these base measures and create calculated measures. We calculate top sale as shown in figure (4.13):

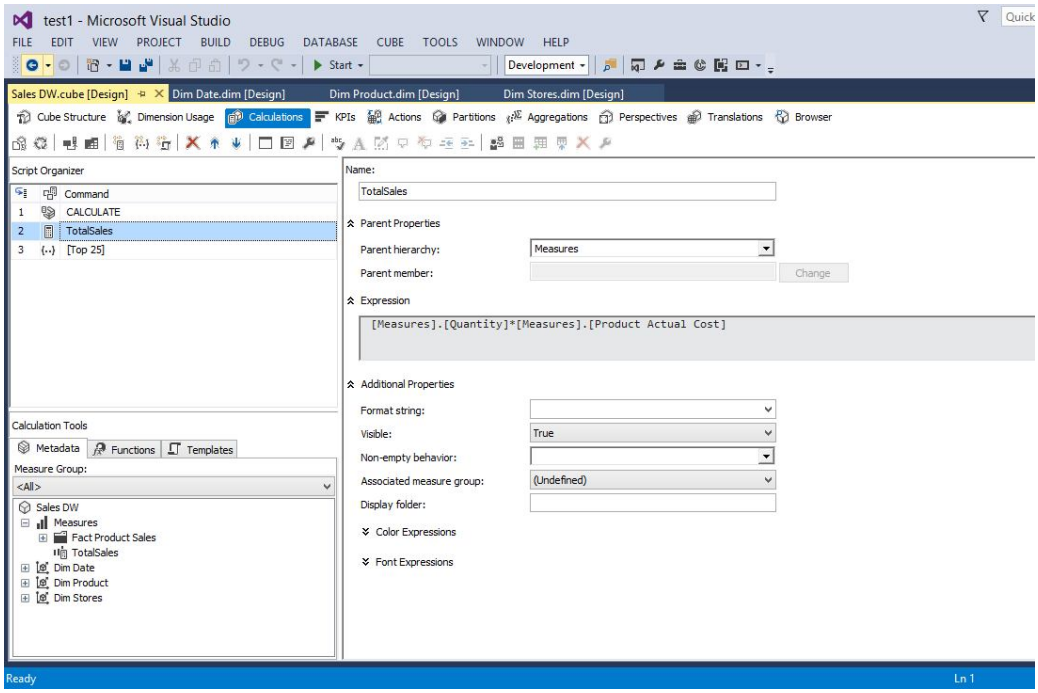


Figure (4.13): Calculation in a Cube

We used another measurement calls named sets, which can be perceived as a query already defined in the cube, similar to views in SQL Serveras shown in figure (4.14).

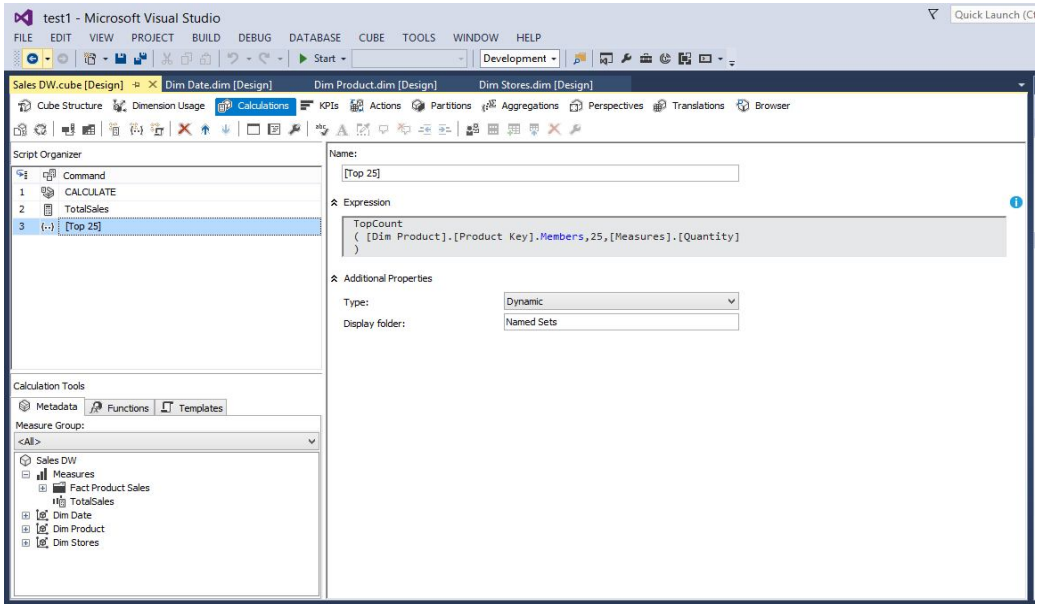


Figure (4.14): Cube Named Sets

Once the cube design and development was completed, the next step is to deploy the cube. When the cube is deployed, a database for the solution is created in the SSAS instance. Each of the dimensions and measure group definitions are read, and data is calculated and stored as per the design and configuration of these objects.After Deploying the Cube we connected to the cube and browsed data by dragging and dropping dimension attributes and measures on the Browser pane. As shown in Figure (4.15)

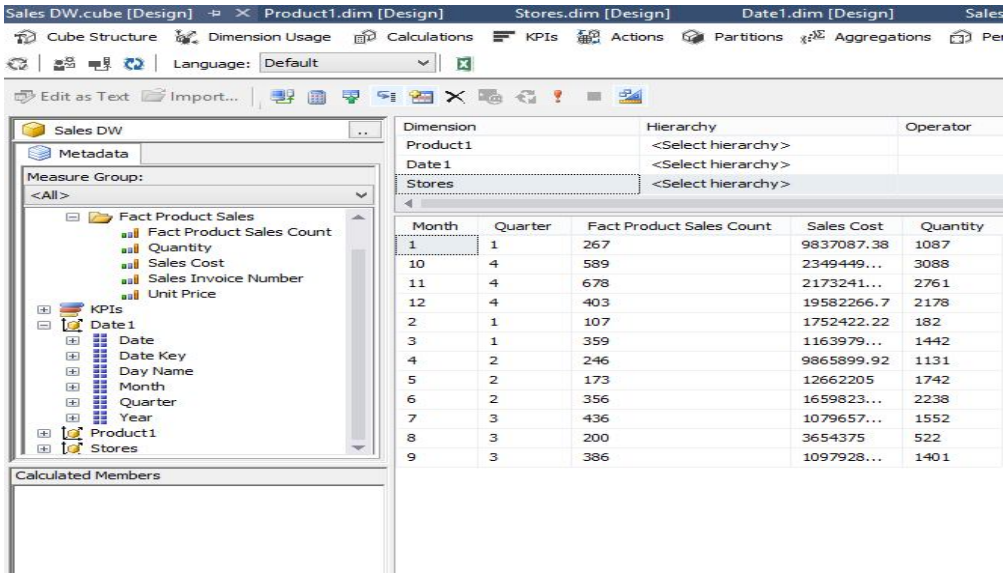
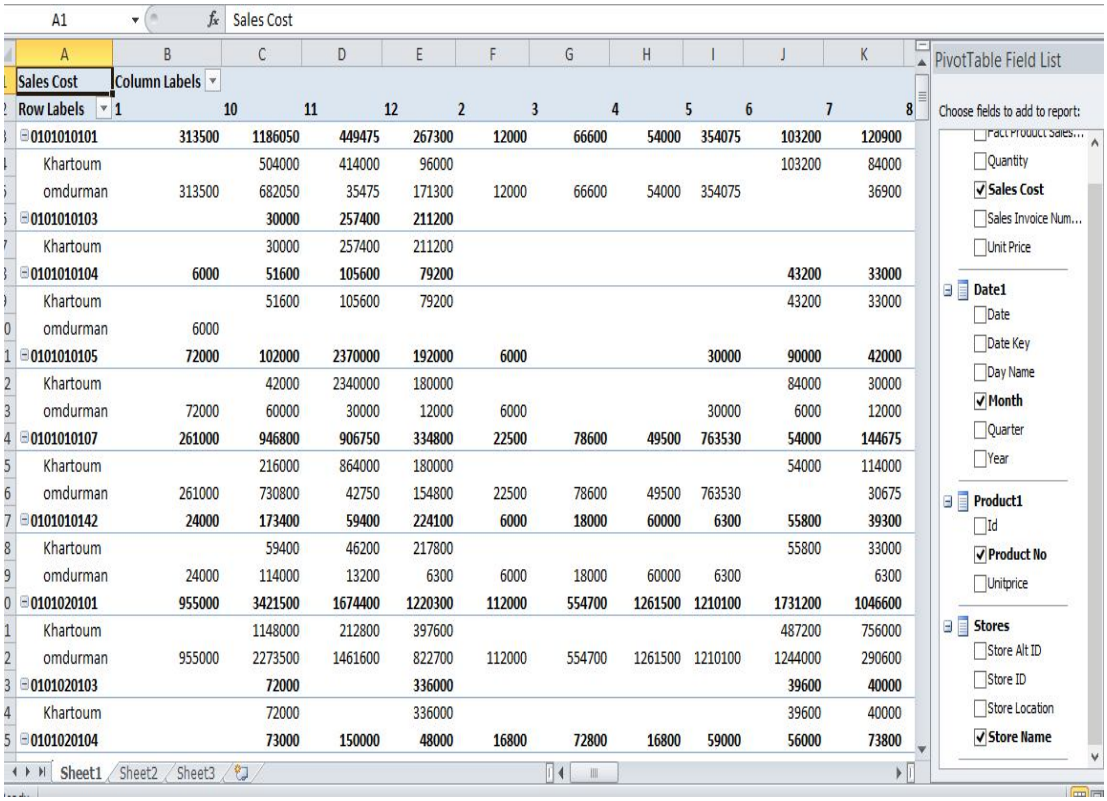


Figure (4.15): Cube Browser

As described above, According to the Manager's needs, the total sales and total quantity sold can be displayed in a month or quarter.

After that, client applications start querying the cube. One of the most user friendly client tools for business users to query a cube is Microsoft Excel. It has a built-in interface and components to support GUI based connection, querying and formatting of data sourced from a cube.

Business users used the familiar interface of Excel and create ad-hoc pivot table reports by querying the cube without any detailed knowledge about querying a multi-dimensional data source. We connect to the cube we just created using Excel and develop a very simple report using the cube data. As described in Figure (4.16).



Sales Cost	Column Labels									
Row Labels	1	10	11	12	2	3	4	5	6	7	8
0101010101	**313500**	**1186050**	**449475**	**267300**	**12000**	**66600**	**54000**	**354075**		**103200**	**120900**
Khartoum		504000	414000	96000						103200	84000
omdurman	313500	682050	35475	171300	12000	66600	54000	354075			36900
0101010103		**30000**	**257400**	**211200**							
Khartoum		30000	257400	211200							
0101010104	**6000**	**51600**	**105600**	**79200**						**43200**	**33000**
Khartoum		51600	105600	79200						43200	33000
omdurman	6000										
0101010105	**72000**	**102000**	**2370000**	**192000**	**6000**			**30000**		**90000**	**42000**
Khartoum		42000	2340000	180000						84000	30000
omdurman	72000	60000	30000	12000	6000			30000		6000	12000
0101010107	**261000**	**946800**	**906750**	**334800**	**22500**	**78600**	**49500**	**763530**		**54000**	**144675**
Khartoum		216000	864000	180000						54000	114000
omdurman	261000	730800	42750	154800	22500	78600	49500	763530			30675
0101010142	**24000**	**173400**	**59400**	**224100**	**6000**	**18000**	**60000**	**6300**		**55800**	**39300**
Khartoum		59400	46200	217800						55800	33000
omdurman	24000	114000	13200	6300	6000	18000	60000	6300			6300
0101020101	**955000**	**3421500**	**1674400**	**1220300**	**112000**	**554700**	**1261500**	**1210100**		**1731200**	**1046600**
Khartoum		1148000	212800	397600						487200	756000
omdurman	955000	2273500	1461600	822700	112000	554700	1261500	1210100		1244000	290600
0101020103		**72000**		**336000**						**39600**	**40000**
Khartoum		72000		336000						39600	40000
0101020104		**73000**	**150000**	**48000**	**16800**	**72800**	**16800**	**59000**		**56000**	**73800**

The Figure (4.16): Ad-hoc Pivot Table Reports

After the analysis, a report was extracted in Excel format for ease of use. The data can be presented in a detailed or comprehensive manner according to the needs of the executives through the pivot table. If the manager want the exact details of the product that was sold on a special day and in a particular branch, the manager can access to that by clickingthe sign (+)that mean (drill down) .If the manager want the total sales for a particular month in one branch or several branches can get the information by pressing the sign (-)that mean (drill up) and can also get the total sales in months or the year.

4.8. SQL Server Report Services (SSRS):

The report, using SQL Server Reports Services (SSRS), which is part of the Microsoft BI suite of products used as part of this solution, is the appropriate way to show an example of the ability to link amongst star schemas. By building the report to carry out the query, it is simple and fast.

4.8.1.Design report:

The Design the Query step in the Report Wizard will display the dialog as shown below figure (4.17).

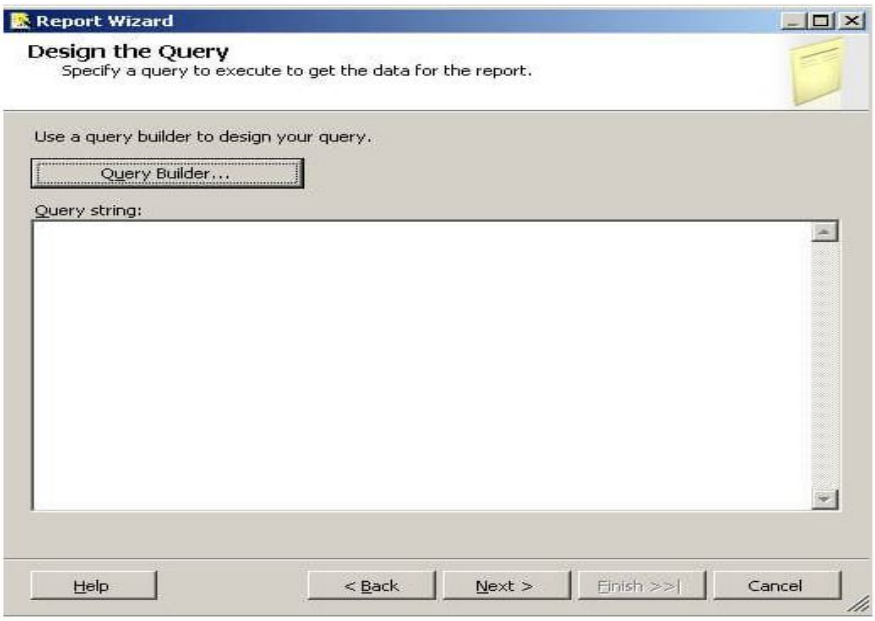
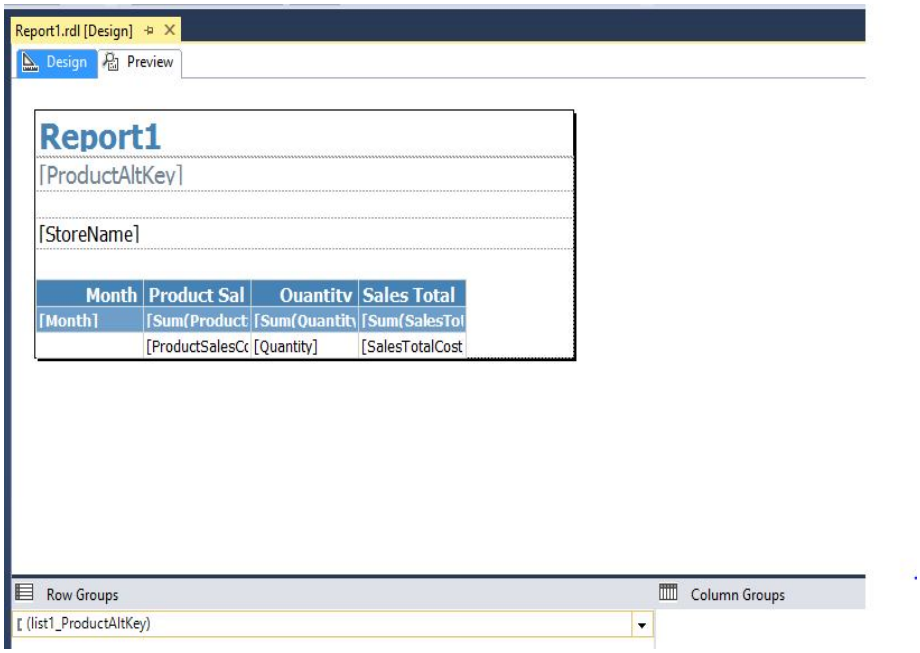


Figure (4.17): Design Report with Wizard

The next figure (4.18) illustrates the report after completed the Report Wizard.



The Figure (4.18):Report After Design

4.8.2.Perview report:

Previewing a report enables us to view the rendered report without having to first publish it to a report server.We probably want to preview our report frequently during design time. Previewing the report will also run validation on the design and data connections so we correct errors and issues before publishing the report to a report server. Enable drilldown will initially hide details and allow expanding with a click on the plus icon. As shown below in figure (4.19):

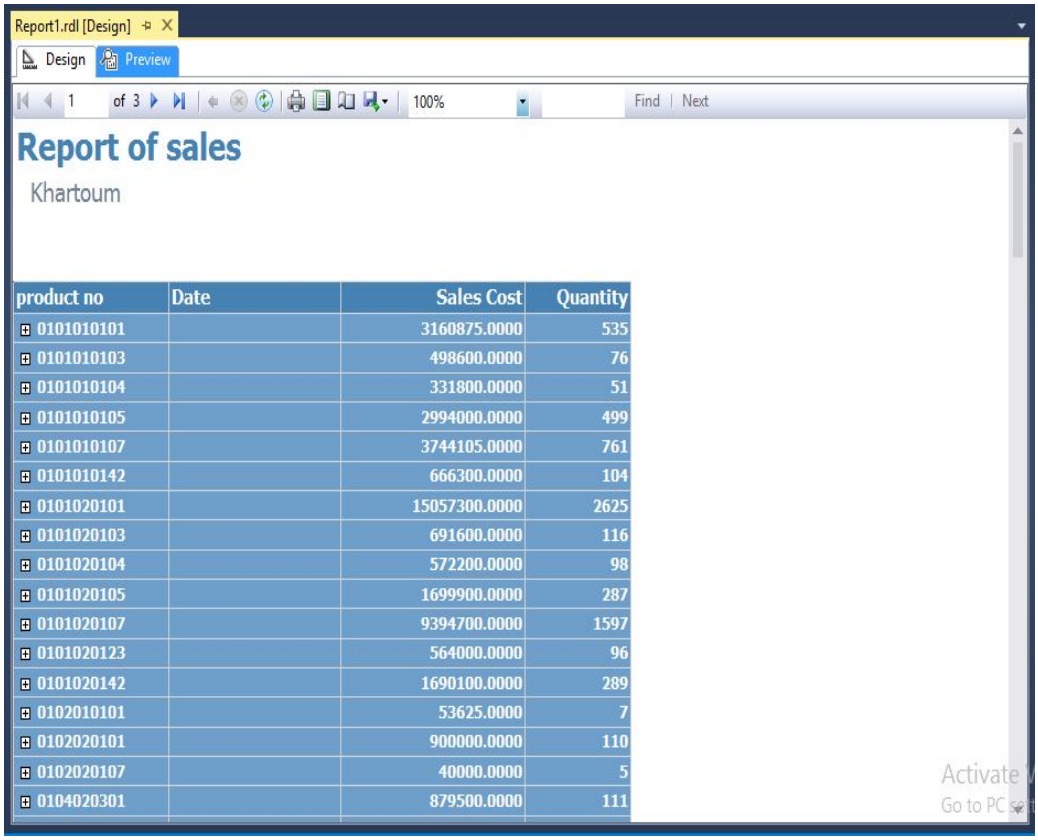
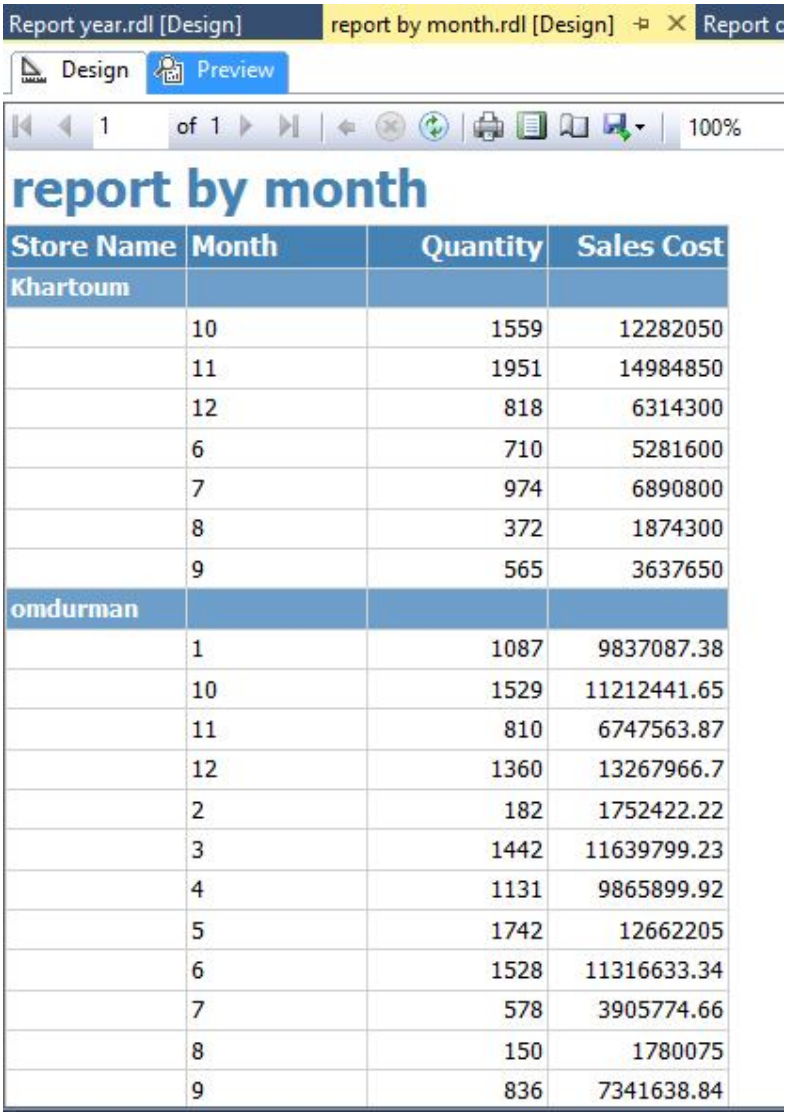


Figure (4.19): Preview Report

Now themanager hascapabilities obtain to any information that he needs it for making any decision. For instance if he wants know which is least quantity has been sold during one year, and it was on which month and in which branch in a factory. This could be through report establishing form a data warehouse as it illustrative below in figure (4.20). It was clear in the report that least quantity sales was in August in Omdurman branch and Khartoum, where in Khartoum reached (372) product and in Omdurman just only (150) product.



report by month

Store Name	Month	Quantity	Sales Cost
Khartoum			
	10	1559	12282050
	11	1951	14984850
	12	818	6314300
	6	710	5281600
	7	974	6890800
	8	372	1874300
	9	565	3637650
omdurman			
	1	1087	9837087.38
	10	1529	11212441.65
	11	810	6747563.87
	12	1360	13267966.7
	2	182	1752422.22
	3	1442	11639799.23
	4	1131	9865899.92
	5	1742	12662205
	6	1528	11316633.34
	7	578	3905774.66
	8	150	1780075
	9	836	7341638.84

Figure (4.20): The Report of Sold Quantity inyear

As well as, manager or decision maker could know what is the Best-sellingisthe product number (0101020101) was sold (65) times in the last quarter of the year in Omdurman branch.As shown below figure(4.21).

report of sales in 2 locations

65

	1	2		3		4	
	omdurman	Khartoum	omdurman	Khartoum	omdurman	Khartoum	omdurman
	Product Sales Count	Product Sales Count	Product Sales Count	Product Sales Count	Product Sales Count	Product Sales Count	Product Sales Count
0101010101	18	6	9	10	16	29	25
0101010103						14	
0101010104	1	2		2		11	
0101010105	5	5	2	6	2	40	6
0101010107	18	3	11	10	12	34	17
0101010142	5	3	2	1	1	11	6
0101020101	37	7	29	16	34	24	65
0101020103		3		7		10	
0101020104	5	1	3	3	1	11	1
0101020105	5	3	1	6	3	21	5
0101020107	23	6	19	9	8	25	36
0101020123				5		8	
0101020142	7	2	4	6	2	13	11
0102010101	1				2		3

Activate Windo
Go to PC settings to

Figure (4.21): The Report of Best Selling in Tow Location by Quarter

Also, manager could know what is the lowest-selling product isthe product number (1001010403) was sold (1) times in the first quarter of the year in Omdurman branch.As well as the product number (101020101) which was sold once in the fourth quarter of the year in Omdurman branch.as shown below figure(4.22).

	1	2		3		4	
	omdurman	Khartoum	omdurman	Khartoum	omdurman	Khartoum	omdurman
	Product Sales Count	Product Sales Count	Product Sales Count	Product Sales Count	Product Sales Count	Product Sales Count	Product Sales Count
0204011943							2
0204011944			1		1		
0205010101			2				
0205010108			2				
0206010323	7		1		5		1
0206011903	2		12		11		5
0207010102	3		1				
0301010109	2	5	6	10		1	5
0301010309	17	2	11	9	9	6	6
0301010339	3		1		1		1
0301010409	6	1	6	1	3	4	13
0301010439	5				1		1
0301010509	14	3	6	7	3	27	7
0301010539	6	4		13		15	
0302010309	3		5		2		4
0501010901	23	7	10	15	18	5	21
0501011701	25	12	20	24	6	36	
0501020201	2	11	1	47	2	61	
0601020409				8			
0601020509				7			
1001010103	8		8		9		8
1001010303	11		7	12	8	6	10
1001010403	1						
101020101							1
1101010101		1		3		1	
1101010301		1	1	2	2		
1301020101	8	2	7	12	7	16	11

Figure (4.22): The Report of Lowest-Selling by Quarter

4.9. Data Mining using Microsoft Time Series Algorithm

In the table below (4.1) sample for data used to make data mining for database.The Date column in the table contains a time identifier. Therefore, we designated this column as the key time column for the time series model. The Sales_Cost column describes the value of the sale of the product in a transaction. The Quantity column describes the quantity of the specified product had been sales. These two columns contain the data that is used to train the model. Both Sales_costandQuantity can be predictable attributes.

Table (4.1): Sample for Data in Data Warehouse Used in Data Mining

Date	Quantity	Sales_Cost
17/1/2010	50	375000
11/3/2010	24	75000
11/5/2010	15	39600
26/6/2010	259	1425000
15/8/2010	31	145700
21/9/2010	150	885000
19/12/2010	75	450000

4.9.1. Viewing a Time Series Model

After the model has been trained, the results are stored as a set of patterns, which can be explored or used to make predictions. To explore the model, the Time Series Viewer can be used. The viewer includes a chart that displays future predictions, and a tree view of the periodic structures in the data. The stored content for the model includes details such as the periodic structures detected by the ARIMA and ARTXP algorithms, the equation used to blend the algorithms, and other statistics.

The chart that is displayed in this viewer includes both historical and predicted data. Predicted data is shaded to differentiate it from historical data. Historical information appears to the left of the vertical line and represents the data that the algorithm uses to create the model. Predicted information appears to the right of the vertical line and represents the forecast that the model makes. The combination of the source data and the prediction data is called a series. We used the time series algorithm because it is suitable with data and used to predict sales in the coming period, and the result of data mining of these data by using the algorithm. The prediction of cost of sales and quantity for the first quarter will be shown in the following figure (4.20)

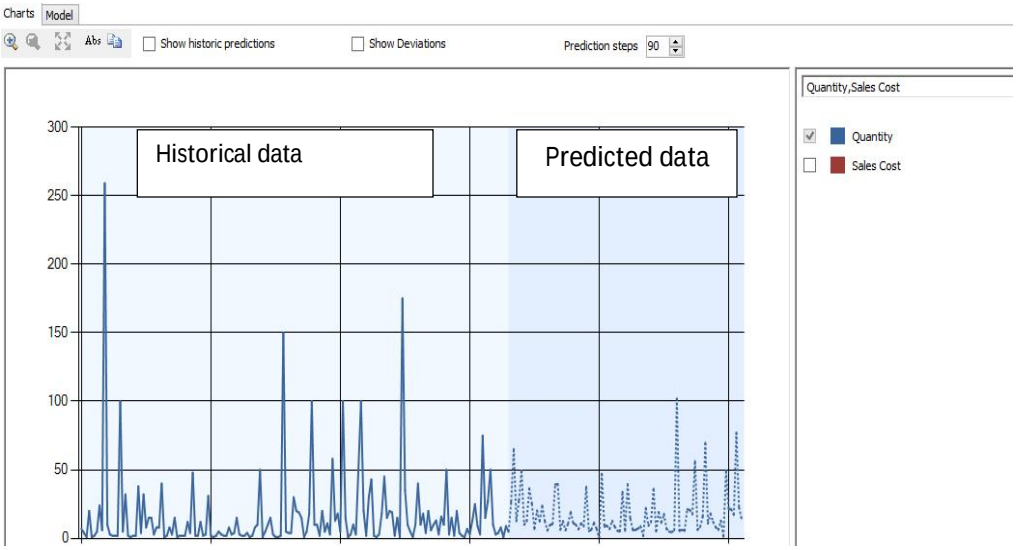


Figure (4.23): Chart of Quantity

4.9.2. CreatingTime Series Predictions

We create queries (DMX) to return a variable number of predictions, and extra columns to the predictions to return descriptive statistics.

ThePredictTimeSeriesfunction returns a prediction for the next five time steps by default. In this model, the predictable attributes are Quantity and cost of sales, we use [Quantity] and [SalesCost] in PredictTimeSeriesfunction, as shown in next figure (4.21).

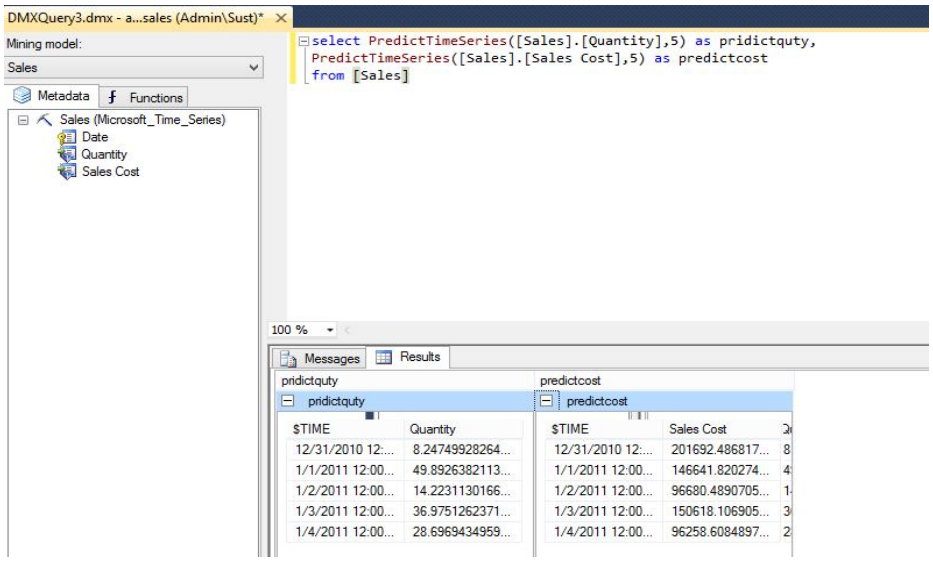


Figure (4.24): ExtractPredict Value of Quantity and SalesCost

The output of the query is display in the next table (4.2)

Table (4.2): Sample of Output of Query

$time	Quantity	SalesCost
1/1/2011	45	112530
1/18/2011	3	166117
2/4/2011	46	165280
2/25/2011	5	90393
3/5/2011	101	118983
3/28/2011	77	156768

Quantity will be lower than the end of the previous year as the beginning of the New Year in January and February, but will be increase in March, Where to support the executive managers in decisions making in this regard.

CHAPTER 5

5.1. Conclusion

The research concentrate on the design and application of the data warehouse , and get benefit from these data after its analyzing and obtaining knowledge at the end of the research to facilitate and support the executive manager's decisions , we used in this research information belong to National Company of paints Department sales as a case study . We extracted reports after analysis of the data in two forms (excel), which is simple in the presentation of the report where we can do drill down and drill up for information, as requested by the executive manager easily. The second format of the report is to extract the report from SSRS, which is a tool that presents the report in a simple, easy and beautiful way. We have also showed the results of the research through the work of Data Mining and the use of the time series algorithm to predict the quantity been sold and the sales process in the coming period, so that executive manager can do the necessary precautions and answer the question (Will this increase the productivity and profits in the coming period or not).

5.2.Recommendations

1. In this study, the focus is not on the management of the data warehouse to continue in this study, the focus should be on subsystems of management the data warehouse.
2. Build the other Data Marts to the data warehouse.
3. Developing reports to web reports.

REFERENCES

Ankur Jain, ManghatNitish, SaurabhChandra.Sales Forecasting for Retail Chains,https://docs.microsoft.com/en-us/sql/analysis-services/data_miningmicrosoft-time-series-algorithm (last retrieval:December 2017).

Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., & Paraboschi, S. (2001). Designing data marts for data warehouses. ACM Transaction Software Enineering Methodology.

Borysowich, C. (2007, 2010). Better Data Warehouse Modelling. Retrieved fromhttp://it.toolbox.com/blogs/enterprise-solutions/better-data-warehouse-modelling-20835.

Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. SIGMOD Record., 26(1), 65-74.

del Hoyo-Barbolla, E., & Lees, D. (2002). The use of data warehouses in the healthcare sector. Health Informatics Journal, 8(1), 43-46.

Dennis Wegener et al.(2010) On Integrating Data Mining into Business Processes, 13th International Conference on Business InformationSystems (BIS 2010), Berlin, Germany.

Dias, Tait, Menolli, & Pacheco, &. (2008). Data warehouse architecture through viewpoint of information system.

Edward M. Leonard.(2011). Design and Implementation of an Enterprise Data Warehouse.

Eric Johnson and Joshua Jones.(2009) .Analysis and Reporting for Business Intelligence Solutions Built on Microsoft SQL Server.

Golfarelli & Rizzi Data Warehouse Design :Modern Principles and Methodologies ,2009

Hornick, M.F. et al., "Java Data Mining: Strategy, Standard, and Practice,"Morgan Kaufmann, San Francisco (2006).

Inmon, B. (1999). Data mart does not equal data warehouse.

INMON W.H., "BUILDING THE DATA WAREHOUSE", SECOND EDITION, JWILEY AND SONS, NEW YORK, 1996

Inmon, W. H. (2005). Building the data warehouse: Wiley Publishing Inc.,Indianapolis.

Jens Lechtenbörger, Data Warehouse Schema Design, IOS Press, (2001).

Jiawei Han and MichelineKamber .Data Mining Concepts and Techniques (2006).

Kerkri, E. M., Quantin, C., Allaert, F. A., Cottin, Y., Charve, P., Jouanot, F., Yétongnon, K., (2001). An approach for integrating heterogeneous information sources in a medical data warehouse. Journal of Medical Systems, 25(3), 167-176.

Kimball, R., & Ross, M. (2002). The Data Warehouse Toolkit (2nd Edition).

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit (3rd Edition).

Lynn Langi and otger,Smart Business Intelligence Solutions SQL Server (2008).

Mahesh Kumar Yadav et al, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.1, January-2016, pg. 107-115.

March, S. T., & Hevner, A. R. (2007). Integrated decision support systems: A data warehousing perspective. Decision Support Systems, 43(3), 1031-1043.

Mohania, M., Samtani, S., Roddick, J., & Kambayashi, Y. (2007). Advances and research directions in data-warehousing technology.

Temitope Adeoye and Raufu Olalekan.Design of Data Warehouse and Business Intelligence System,Sweden (2011).

technet.microsoft.com/en-us/library/ms174923(v=sql.110).aspx
(Lastretrieval time: December 2017).

www.msdn.microsoft.com/en-us/library/ms173767.aspx (Last retrieval time: August 2017).