Designing sales data mart system for ease of analysis and developing data mining model to enhance the promotional strategies for Gamma Pizza hut

S.Shamaran

129171M

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree of Master of Science in Information Technology.

November 2016

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been duly acknowledged in the text and a list of references is given.

S.Shamaran

Name of Student

S Shamaran

Signature of Student

Date: 26/11/2016

Supervised by

Mr.Saminda Premaratne

Name of Supervisor

UOM Verified Signature

Signature of Supervisor

Date: 21/12/2016

Dedication

THIS THESIS IS DEDICATED TO MY PARENTS AND MY SISTER SHANTHIPAA FAMILY AND BROTHER SHATHULAN FAMILY

Acknowledgement

It is a great pleasure to me thanks the many people who made this thesis possible.

My sincere gratitude and thanks to Mr.Samida Premaratne, the supervisor for this project and Course coordinator MSC in IT, for the tremendous guidance, motivation and corporation given throughout the project.

Heartfelt thanks to my parents, my sister family, and my brother family for their support and encouragement during the development of this project.

Special thanks to previous working colleagues and managers (Buddhika, Chithara, Dinith) of Scienter Technologies Pte Ltd.

Grateful thanks to current working colleagues and Sujeewa Fernando, Head of Dot Net development at Mazarin.

And lastly gratitude thanks to my wife Hamshine and his family for in order to able to continue my project after wedding on this year.

Abstract

Making effective business decisions with the data is the key to succeeding in today's competitive environment. Organizations are now looking to improve their decision-making ability with their current data, but unfortunately operational systems have limited features and various ad-hoc reports for same data. This unsatisfactory & frustration lead the managers and IT industry to find new level of applications. These applications focused on ease of analysis on the single screen to make effective decisions at the time and mining techniques help to generate new business opportunities by providing prediction of trends and behaviors as well as discovery of previously unknown or hidden patterns.

The DSS/BI systems should have more analyzing features and structured data. But current OLTP data and its database design not give much more analyzing power. In order to that OLAP architecture has built from various database vendors to make to use by DSS /BI systems. The developing of a data warehouse database and Data Mart database with suitable schema and approaching with relevant architecture is make a foundation for DSS/BI systems

The Data warehouse database makes on available history data as possible of getting last update record. The fact and dimension structure are used when designing database schema for Data Warehouse. ETL process generate a data to warehouse from various data sources. The Data Marts are used for holding various subject areas like sales, purchase, production, finance, etc. But here only considering about sales and delivery data only. The Data Cube Technology (OLAP technology) is used for end user to viewing data with various dimensional and drill-down drill-up processes within the application.

Finally those data are used to mining frequent patterns, Associations and Correlations between items in menu orders by using apriori algorithm (Microsoft Association algorithm) and forecasting Predictive sales for each item by using ARIMA algorithm (Mircosoft Time Series)

The data warehouse solution can be made from by integrating various database technologies in the middle; those technologies include SQL Server Management Studio (SSMS), SQL Server Integration Services (SSIS), SQL Server Analysis Server (SSAS), SQL Server Report Service (SSRS) and SQL Server Data Tools for Visual Studio used to create Analyzing project and Data mining project. C# language, DMX and MDX queries are used to build the simple mining application.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

BI – Business Intelligence

DSS – Decision Support System

KDD – Knowledge Discovery Data

ARIMA - Autoregressive –Integrated moving average

MDX – Multi Dimensional Query

DMC – Data Mining Query

OLAP – On Line Analytical Processing

ROLAP – Relational On Line Analytical Processing

MOLAP – Multidimensional On Line Analytical Processing

HOLAP – Hybrid on Line Analytical Processing

ETL – Extract, Transform, Load

SSIS – SQL Server Integration Service

SSAS – SQL Server Analytical Server

SSRS – SQL Server Reporting Service

BIDS – Business Intelligent Development Studio

DLL – Dynamic Link Library

Chapter 1

Introduction

1.1 Introduction

This chapter describes the background and motivation for the project, the problem identified and a solution in term of users, input, output, process, technologies, features and system requirements. Furthermore, the aim and objectives of this research project and a brief description of what each of the chapters of the paper entails are briefly mentioned herewith

1.2 Back ground and Motivation

I'm working on hospitability software solutions and I found that, the customers have a problem and difficulties with using current software and its database applications. The Inventory and Restaurant Management System and Call Center systems have go through huge amount of data when processing reports, It's t take too many minutes to getting output of results and it's frustrate the user to using the system.

The users complain are now getting increases every day regards system slow, system does not respond and we have not needed information from reports from back office staff mostly senior and middle managers. I'm mostly experienced with Pizza Hut management, their database have 10 years data in one single database to use by ERP System. Increasing of outlets definitely increase transactional data and then that data come as bottle neck for back office operations. And also these data are not able to give enough information to make effective business decisions and identify the profitable promotional strategies to retention the customers. This will lead to operation inefficiency and losing sales of the organization and also lead to hating business applications from their staff.

This will create a huge problem to our clients and our self also to move the business very strong manner. There should be huge gap increase between them and us with the current applications that, what are now we providing. They have huge new requirements that are not able sorted by transactional applications and also it's take more time us for writing queries and testing results and them more time needed for view the results in the report.

In today context and infrastructure there are various Business Intelligent /Data ware house/Data mining solutions carried by various vendors in international market and those are used by many organizations, because they have big data in their database and they really need this type of application for make various analysis. In the mean time in our country this should not needed by looking their database data capacity. But now days the databases are grown well quickly and they also felt in big data problem and converting data into useful information to run the competitive business

1.3 The Problem Domain

The usages of data are gradually increasing very largely. The increasing of outlets and operations is increase the data growth rate very frequently. Those data's are growing very speedily and that creates lot of drawbacks and bottlenecks to use current business applications. And current business applications do not satisfy the user need especially for decision makers and marketing peoples. And existing reports are not adequate to make effective business decisions and format of report changed time to time that be a headache for a service providers and IT staff within the organization.

The users of the system always complain of unresponsiveness of the system and huge time to take to deliver reports. The getting demand of new reports and existing reports with different formats with responsible time is the now looked by the clients for make various analyzing in order to run the business in current competitive fast food business operations. These features not satisfy with current operational systems, because they designed based on the capturing transaction data without losing data and reach high consistently.

The marketing department have now more responsible for their various marketing campaign to develop their business and acquiring new customers and retaining existing customer base and also make sure to gain the more profit by selling more products to huge customers. But now these activities looked as a very competitive in order to high competition from various fast food franchises. They required various analyses and getting knowledge data from existing data is a main business operation by marketing people.

1.4 The Solution Addressed By Others

Since the late 1960, researchers have been developing and implementing computerized systems to support management decision makers. The expanded DSS framework developed in detail helps decision makers and DSS developers explain and categorized potential decision support projects [1].Decision support systems are the core of business IT infrastructure because they give companies a way to translate a wealth of business information into tangible and lucrative results [2]. Many terms are used for specific types of DSS. For example "business intelligence", "collaborative systems", "computationally oriented DSS", "data warehousing", "model-based DSS" and online analytical processing (OLAP) software to label decision support software[1]

These problems are addressed on many organization those solved by designing and developing Data warehouse solutions. The data warehouse solution is foundation for Decision support systems. Decision Support systems are the core of business IT infrastructures because they give companies a way to translate a wealth of business information into tangible and lucrative results.[2]

1.5 Aim & Objective

Aim

The aim of this project is to be developing a Data warehouse database for effective decision making and mining customer order/purchase pattern to define and develop promotional strategies for fast food restaurants.

Objectives

1) Design a suitable data warehouse database by considering data needed for marketing department.
2) Process ETL to Generate data into warehouse
3) Generate suitable data sources for sales and deliveries
4) Build an OLAP cubes to apply a MDX queries
5) Identify business parameters to choose mining model
6) Ready the data for using into relevant mining models
7) Build a Mining model to test data and apply DMX queries
8) Develop a data mining application for end user

1.6 Solution in brief

Finding an associate item sets, which are sale most of the time with another set of items and finding predictive sales of most frequent sales items with period wise to preplan the promotional strategies by considering the associate item sets.

Users

Data Analyst

Input

- Enter a menu name,
- Minimum items in an order
- Customer Telephone Number

Processes

- Connected to data source
- Using relevant mining model
- Processing DMX query in Analyzing Server

Output

- Return predictive result set

Technology

- Used SSIS package for migrate data into different data sources
- Used Data Analyzing Project template to generate Cubes and mining models
- Used Data Cube Technology to drill down/up summarized data
- Used Microsoft Association Rules algorithm to identify most associate items
- Used Microsoft Time Series (ARIMA) to predict sales
- Used MDX, DMX query to get result from Analysis Server
- Used C# language to connect to backend

System Requirement

- Microsoft .NET framework 4.5 or higher
- Microsoft SQL Server 2014 or higher

1.7 Structure of the Dissertation

The first part of this 'Introduction' chapter spells out the reason for choosing this project. This is followed by explaining the problems which this project will address. Next we move on to discuss the aims and objectives set out and there after continue to put down the solution in brief for the project. In the 'Solution in brief' section give the main tools and mining techniques used into drive the mining application. The Introduction chapter ends with describing the structure of the dissertation.

The next chapter is the 'Review of Others' work', and it will give the background information about the project. It will include the findings from the literature survey covering similar and related websites. The Chapter 3, 4 and 5 describes the theoretical foundation of several technologies that are used in this project. The Chapter 6 describes the development methodologies and approach take into develop the project. The Chapter 7 will specify the analysis part of the project and its requirement to carry out the project. The Chapter 8 will specify design details of the project. The database design will be explained and the system overview diagram will be included to give a good understanding about the design decisions. The Chapter 9 is the Implementation chapter, which spells out the implementation details and include list of tables and process diagrams used in the project. Then we move on to the Testing chapter. Here we discuss the how the test results validated with predicted results. The last chapter gives the conclusions about the project; the future work which can be done is shown here. Finally, the references are given in accordance with the Project Guidelines. This includes websites and books which were referred in order to do the project.

Chapter 2

Review of others' work.

2.1 Introduction

In the previous chapter we discussed on brief introduction on the project. It was describe the background and motivation of the project, aim and objectives and main solutions in terms of Users, input, processes, output, technologies and system requirement.

This chapter discusses on other's review works on the literature survey. Also state about others' approaches to solve similar problems and give a what are the parameters they had used for they approach.

2.2 Background

At a time when really knowing customer preference is crucial to successful restaurants operations, data mining should be at the forefront of a manager technology toolbox. Customer attraction, retention and prediction are important marketing concepts in the restaurant industry and central components of data mining [3]. Foodservice operations have long known there is a need to exceed customer expectations in order to stimulate current sales while creating the opportunity for repeat business. Classification is one type of information that could be obtained from data-mining to help decision makers make predictions and forecasts as well as devise campaigns and promotions [4]

2.3 Decision Support Systems concept

Here the author discussed about building DSS in terms of four major components: those named as, the user interface, the database, the models and analysis tools, and the DSS architecture and network [1]. He generically categorized the DS systems with five areas based on the major characteristic. The five categorized are communication-driven, data-driven, document-driven, knowledge- driven, and model-driven DSS.

In Data-driven DSS, he mention on file drawer and management reporting systems, data warehousing and analysis systems, executive information systems(EIS), some spatial DSS(SDSS), and Business Intelligent systems. Data warehouse systems that allow the manipulation of data by computerized tools tailored to a specific task and setting by more general tools and operators provide additional functionality. Data-driven DSS with OLAP, drill-down and scorecards provide the highest level of functionality and decision support that are linked to the analysis of a large collection of historical data.

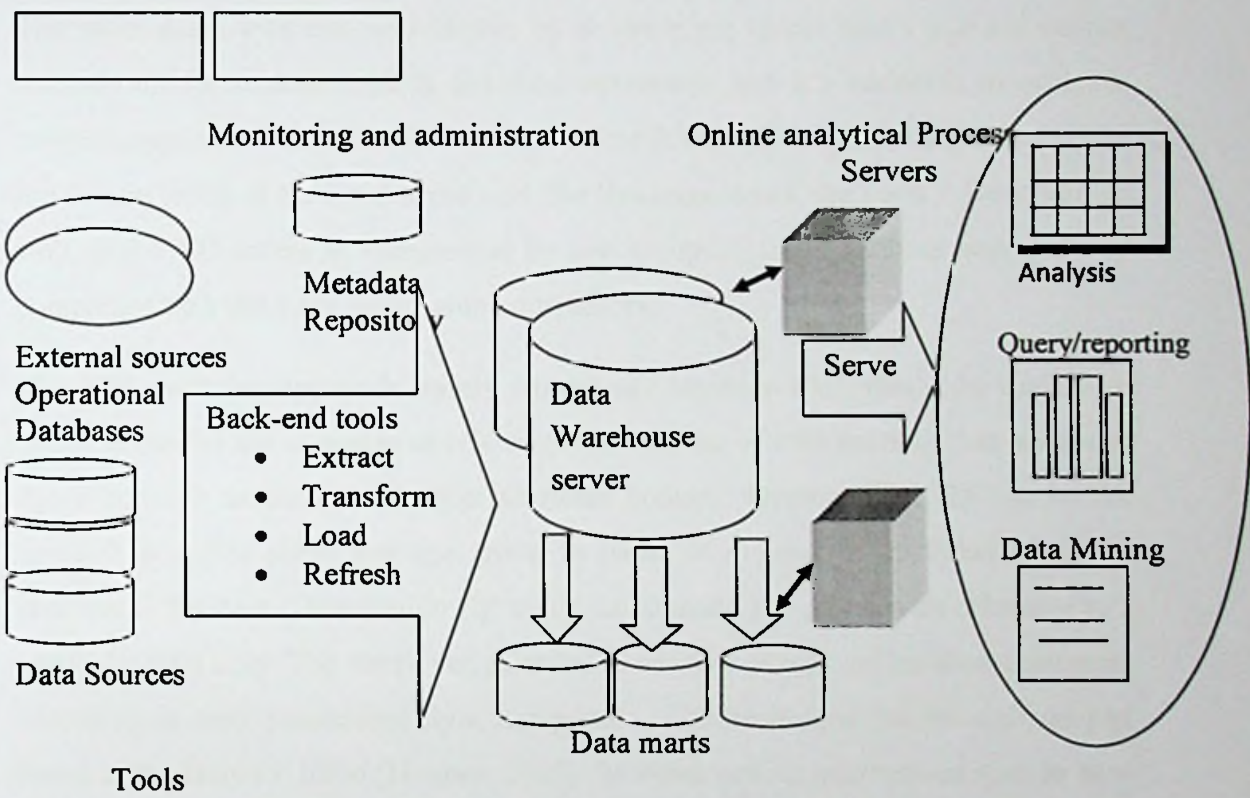


Figure 2.1: Decision Support system architecture

2.4 Database and Data Mining for Coffee shops in Egypt

The paper discussed about terms such as Business Intelligence (BI), data-mining, data warehousing, knowledge discovery, knowledge Discovery in Databases (KDD), and market basket analysis [4].

This paper studied about five coffee shops in Egypt and tries to introduce business intelligent systems and the use of database and data mining techniques and also present methods that support decision making process by using data mining techniques with MS Excel tool.

The research carrying with the data collect on standard format by data collectors by using a laptop. The data collection made on various cities by watching customer consuming. The data are based on customer demography (Gender, Age, Duration, segment, nationality) and Customer order pattern (What did they order? what are they doing? duration).And also shift based (Evening, Morning, Afternoon) study is used to collect these data with choosing days. This study analyzed the top most important three variables by analyzing the way of customers who came and the number of items sold.

The paper came with various statistics by showing pie charts, that's provide various information for readers regards fast-food operations and it's variables to analysis, business segments ,customer outliers. The research used the ADI technique to analyze the data in terms of the Food items sold, the Beverage items, the Food + Beverage, as well as the NO-orders as categorized by age groups. Further analysis was done by comparing with the other coffee ship competitors.

The ADI modeling approach simply termed as 'Attribute (A)', that's the quality or characteristic of the customer as related to the number of customers in that particular age group not an absolute number of items bought, 'Demographic (D)' to be the demand in coffee shops per age group in terms of number of customers and last attribute is the Item 'Distribution (I)' that's the demand per group with reference to a particular item only. The above variables led to analyze or rank coffee shop customers according to their purchasing style and pattern. This technique has been developed based of the Hughes' RFM (Hughes, 2010). Its mean getting information such as how recent, frequent, or the amount of money paid. But researcher said RFM is not a practical approach by looking each coffee shop customer.

The analysis conclude that ADI technique ,it supports decision making by simply delivering a summary of customer behavior and purchasing patterns and further conclude the data mining techniques help to generate new business opportunities by providing prediction of trends and behaviors as well as discovery of previously unknown or hidden patterns.

2.5Data mining on time series: an illustration using fast-food restaurant franchise data [5]

This paper discussed around how mining of time series data from large number of data to improvement in forecast performance, better inventory management and identify potential sales opportunities in a fast-food restaurant franchise. They discussed on business operation and parameters associated with time series such as hourly, daily, weekly, etc. and its granularity level of the aggregation when building data warehouse. And also they mentioned on useful records on external events, special promotions, local events holidays, daily in the database.

They apply data mining in exploration and knowledge discovery when a large number of time series are available for business applications. KDD is considered to be more encompassing process that includes data warehousing, target data selection, data cleaning, preprocessing, transformation and reduction, data mining, model selection evaluation and interpretation

The researchers presented an approach on time series data mining in which automatic time series model identification and automatic outlier detection and adjustment procedures and the use of automatic procedures how to modeling and forecasting of a large number of time series in an efficient manner. And further employs ARIMA (autoregressive –integrated moving average) models for time series data mining.

2.6 Summary

This chapter discussed the background information of the project. As well as mentioned on Decisions support system and available techniques and models to available to make decision on the data.

Next chapter discusses on the theme "Technology adapted". It elaborates on the different technology that can be adapted to solve the problem.

Chapter 3

Technology Adapted

3.1 Introduction

In the previous chapter we discussed on other similar approaches that can be solved a similar problems. In their discussed about Decision Support System/Business Intelligent and data warehouse terms and concepts.

In this chapter we talks about benefits and how/why data warehouse technology that can help out for the project.

3.2 Data Warehouse Technology

The Data warehouse technology is adapted for making decision support/BI/ systems and data mining models. This technology now widely used industries for business decision making, analyzing business data movement pattern and predicting business by doing various forecasting models. Designing and developing data warehouse for decision support system is a different approach rather than designing database for operational system.

3.3 Data Warehouse

Data warehouses contain data consolidated from several operational databases and tend to be orders of magnitude larger than operational databases. Typically the data warehouse is maintained separately from the organization's operational databases because analytical application's functional and performance requirements are quite different from those of operational databases.

Ralph states that a data warehouse is" a copy of transaction data specially structures for query and analysis. ". There are many definitions and arguments related to definition of data warehouse. Most data warehouses use relational database technology because it offers a robust, reliable, and efficient approach for storing and managing large volumes of data. The most significant issue associated with data warehouse construction is database design, both logical and physical. Building a logical schema for an enterprise data warehouse requires extensive business modeling. Data from various online transaction processing (OLTP) applications and other sources is selectively and organized on the data warehouse database for use by analytical applications and user queries.

3.3.1 Basic Elements of Data Warehouse

The table that follows presents the main components of one data warehouse in an interrelated vision of Kimball et al. (1998)

PHASE	BASIC ELEMENT	DEFINITION
Data Sources	**Source systems**	Operational system that the function is to capture business transactions
Staging Area	**Staging area**	Storage area and set of processes that clear out, transform, combine, remove duplications, archive and prepare the sources data to be used in the data warehouse
Integration	**Data Mart**	
	Operational Data Storage	Integration point of operational systems of the organization. Create to integrate the different systems of the organization at operational level.
Data Warehouse	**Presentation server**	Physical target machine where data are organized and stored for end users access, reports generators and applications queries.
	Dimensional model	Specific subject matter for data modeling as alternative for Entity-Relationship model.
	Relational Online Analytical Process (ROLAP)	Set of user and application interfaces that gave multidimensional characteristics to relational databases
	Multidimensional Online Analytical Process(MOLAP)	Set of user interfaces, applications with a proprietary database that are strongly
	Metadata	All information in the data warehouse environment that is not real data
Dimension Constructio n	**Online Analytical Process/ Online Analytical Process Cubes**	Generic activity of querying and presenting textual or numeric data from data warehouse, as well as a specific dimensional way to query and present. It is a non-relational technology and usually based on data multidimensional cubes

Data analysis Tools and Applications/Users Exploration of Information	**Business process**	Coherent set of organizational business activities that make sense to business users of data warehouse.
	Application for end user	Collection of tools that query, analyze and present desired information
	Data Access Control Tool for End-Users	Data warehouse client. Can be simple as ad-hoc queries systems or complex and sophisticated as data mining or modeling applications
	Tools for Ad-Hoc queries	Specific type of data access tool that includes the end user to create his/her own queries, manipulating relational tables and its functions directly.
	Modeling Applications	Type of sophisticated tool with analytical capacity to transform or understand the data warehouse outputs (e.g. Data Mining, Forecast Models, Behavior Models)

Table 3.1: Definitions of basic elements of the data warehouse

3.3.2 Data Warehouse Management

Software tools are very significant parts of the infrastructure in a warehouse environment. Software tools are available for every architectural component of the data warehouse. Figure 3.7 shows the tool group that support the various functions and services in a data warehouse[6]

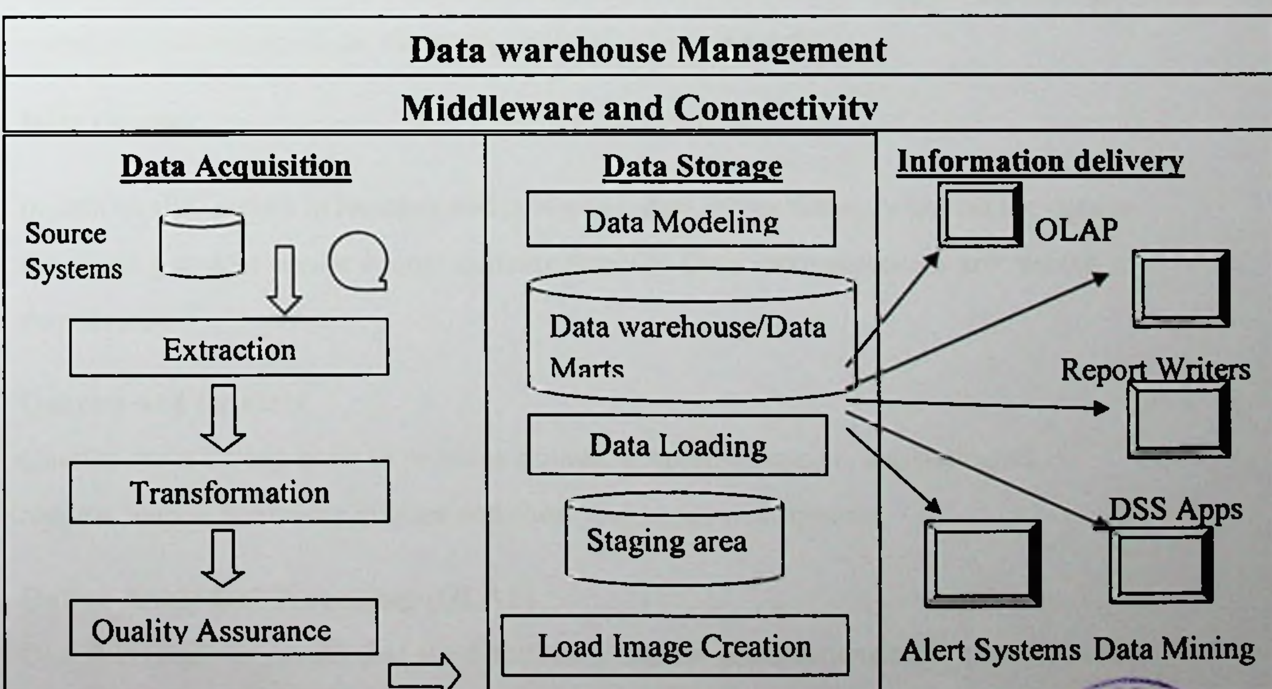


Figure 3.1: Collection of Data Warehouse Tools

Data Modeling

Data models are database schema that was create and maintain data models for the source systems and the data warehouse target database, If necessary , data models may be created for the staging area. The dimensional modeling capabilities are used on when designing schemas. Specially for STAR schemas when designing Sales fact table with multiple dimension

Data Extraction

Two primary extraction methods are available; those are bulk extraction for full refreshes and changed-based replication for incremental loads. The both technique are used for data extraction to making comfortable operation and easiest maintenance.

Data Transformation

Transform extracted data into appropriate formats and data structures, transformation can include field splitting, consolidation, and standardization and provide default value when data is not applicable

Data Loading

In the Data loading the transformed and consolidated data load into data warehouse repository. Some loaders generate primary keys for the tables being loaded and pre coded procedures stored on the database itself to load the data.

Data Quality

In data quality assists in locating and correcting data errors that may use on the data in the staging area or on the source systems directly. Data inconsistencies are resolve in this activity.

Queries and Reports

Queries are allowing users to produce canned, graphic-intensive, sophisticated reports, help to formulate queries and show end result with reports.

Online Analytical Processing (OLAP)

OLAP is a server model that used for run complex multidimensional queries and it enabled users to generate canned queries.

Middleware and Connectivity

Provide transparent access to source systems in heterogeneous environments and access to databases of different types on multiple platforms. Tools are used for providing interoperability among the various data warehouse components

Data Warehouse Management

Assist data warehouse administrators in day-to-day management. Some tools focus on the load process and track load histories and other tools track types and number of user queries.

Meta Data

Metadata is as one of the major building blocks for a data warehouse and that contain the answers to questions about the data in the data warehouse. That can be grouped in to three types, namely, operational, extraction and transformation, and end-user metadata. The metadata is absolutely necessary for using, building, and administrating data warehouse.

There are many database products are available with these features for developing data warehouse databases. i.e.: Microsoft SQL Server, Oracle, and SAP. Each product has advantages and disadvantages and they compare with requirement and other features required by clients.

3.4DatabaseDesign

The database design is a core part of data warehouse project and it's has more than one design schemas and got various assumptions when designing physical database. Here logical and physical designs are described those are specifically design for Data Warehouse

3.4.1 Logical Design

There are different schemas available for designing database for data warehouse. Those are star schema, snowflake schema and constellation schema. The schema design depends on business requirements of decision makers. These schemas build dimensional modeling concepts such include Facts, Dimensions and Measures (variables).And data can be categories' in hierarchies and levels.

In the stat schema design, the database consists of a fact table that describes all transactions and a dimensional table for each entity. Snowflake schemas are a refinement of star schemas in which a dimensional hierarchy is explicated represented by normalizing the dimension tables. A Constellation Schema contains more than one fact table sharing one or more dimension tables.

Fact tables are the large tables in warehouse schema that store business measurement. That represent data, usually numeric and additive, that can be analyzed and examined. (i.e.: sales, Purchase, Production, etc.). A dimension is a structure, often composed of one or more hierarchies that categorize data, these attributes help to describe the dimensional value normally descriptive textual values. They are combined with facts to able with answer the business questions (i.e.: Menu, menu Category, Customer, Outlet, Date, etc.). A measure is a numeric attribute of a fact, representing the performance of a behavior of the business relative to the dimensions (i.e.: quantity, rate, etc.).

Hierarchies are logical structures that use ordered levels as a means of organizing data. This can be used to define a data aggregation. For example; in a time dimension, a hierarchy might aggregate data from the month level to the quarter level to the year level. A hierarchy can also be used to define a navigational drill path and to establish a family structure. A level represents a position in a hierarchy that represents data at the month, quarter, and year levels.

3.4.2 Physical design

Physical design is the creation of the database with SQL statements. During the physical design process, the gathered during the logical design phase data convert into physical database. Here defined the model of data warehouse consisting of entities, attributes and relationships. The entities are linked together using relationship. Attributes are used to describe the entities.

Database systems use redundant structures such as indexes and materialized views to efficiently process complex queries. Determining the most appropriate set of indexes and views is a complex design problem. Improving the efficiency of table scans and exploiting parallelism to reduce query response times are important design considerations.

3.5 OLAP Technology

On-Line Analytical Processing (OALP) is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access in a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user.

While relational databases are considered to be two-dimensional, OLAP data is multidimensional, meaning the information can be compared in many different ways. This analysis tool need for serious analysis, because of limitation on traditional tools of report writers, query products, spreadsheets and OLTP and basic data warehouse environment. The data warehouse provides the best opportunity for analysis and OLAP is the vehicle for carrying out involved analysis. And it is another layer in the data warehouse, providing interface between the data and the user.

The term OLAP was introduced in a paper entitled "Providing On-Line Analytical Processing to User Analysts," by Dr.E.F.Codd, the acknowledged "father" of the relational database model. The paper, published in 1993, defined 12 rules or guidelines for an OLAP system. Later, in 1995, six additional rules were included.

Basic features

Multidimensional Analysis	Consistent Performance	Fast response times for interactive queries
Drill-down and roll-up	Navigation in and out of details	Slice-and-dice or rotation
Multiple view modes	Easy scalability	Time intelligence(year-to-date, fiscal period)

Advanced Features

Powerful calculations	Cross-dimensional calculations	Pre-calculation or pre-consolidation
Drill-through across dimensions or details	Sophisticated presentation & displays	Collaborative decision making
Derived data values through formulas	Application of alert technology	Report generation with agent technology

Figure 3.2: General features of OLAP

3.6 OALP Servers/Models

There are three models exist for store data in the OLAP Servers. Those are ROLAP, MOLAP and HOLAP servers. The variations of these models relates to the way is stored for OLAP server that's be storage methodology. The HOLAP model is a combination of ROLAP and MOLAP

3.6.1The ROLAP Model

In the ROLAP model, data is stored s rows and columns in relational form. This model presents data to the users in the form of business dimensions. In order to hide the storage structure to the user and present data multi dimensionally, a semantic layer of metadata is created. The metadata layer supports the mapping of dimensions to the relational tables. Additional metadata supports summarizations and aggregations

3.6.2 The MLOAP Model

In the MOLAP model, data for analysis is stored in specialized multidimensional databases. Large multidimensional arrays form the storage structures. The array values indicate of the cells. These cells are intersections of the values of dimension attributes. Pre calculated and prefabricated multidimensional data cubes are stored in multidimensional databases. The MOLAP engine in the application layer pushes a multidimensional view of the data from the MDDBs to the users

3.6.3 The HOLAP Model

HOLAP technologies attempt to combine the advantages of MOLAP and ROLAP. For summary-type information, HOLAP leverages cube technology for faster performance. When detail information is needed, HOLAP can "drill through" from the cube into the underlying relational data.

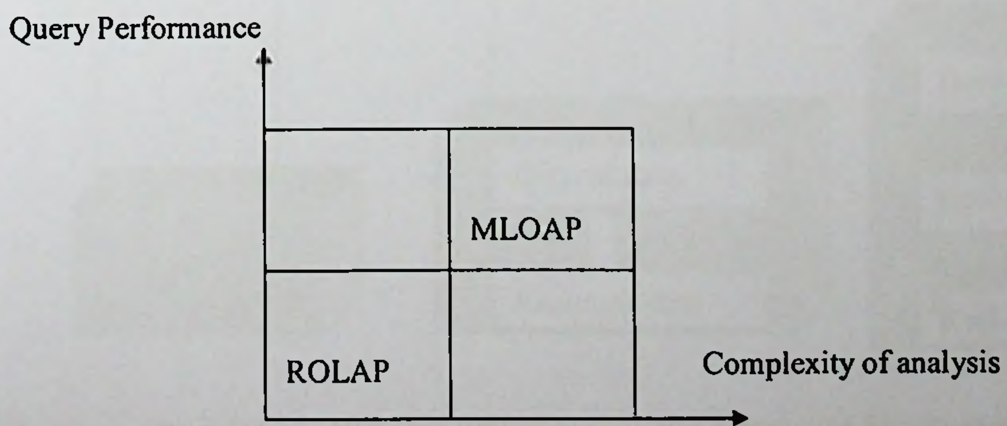


Figure 3.3: MOALP vs. RLOAP

Each model have advantages and disadvantages, the MOLAP is used for excellent query performance and perform complex calculations but limited on amount of data and need additional investment cost for human and capital resources. ROLAP model can handle large amounts of data and can leverage functionalities in the relational database, but performance is slow and limited by SQL functionalities.

Here the HOLAP model was taken to implement the solution; The HOLAP model aims at mixing the advantages of both basic models. It takes advantage of the standardization level and ability to manage large amount of data from ROLAP implementations, and the query speed typical of MOLAP systems. HOLAP implies that the large amount of data should be stored in an RDBMS to avoid the problems caused by sparsely, and that a multidimensional system stores only the information users most frequently need to access. If the information is not enough to solve queries, the system will transparently access the part of the data managed by the relational system.

3.7 Data Warehouse Architecture

There are three type of Data warehouse Architectures are available for information systems to derived data from existing data. Conceptually those named as Single - Layer, Two – Layer and Three – Layer.

3.7.1 Conceptual View

Is a conceptual view mentioned three layers

<u>Single-Layer</u> <u>Two-Layer</u> <u>Three-Layer</u>

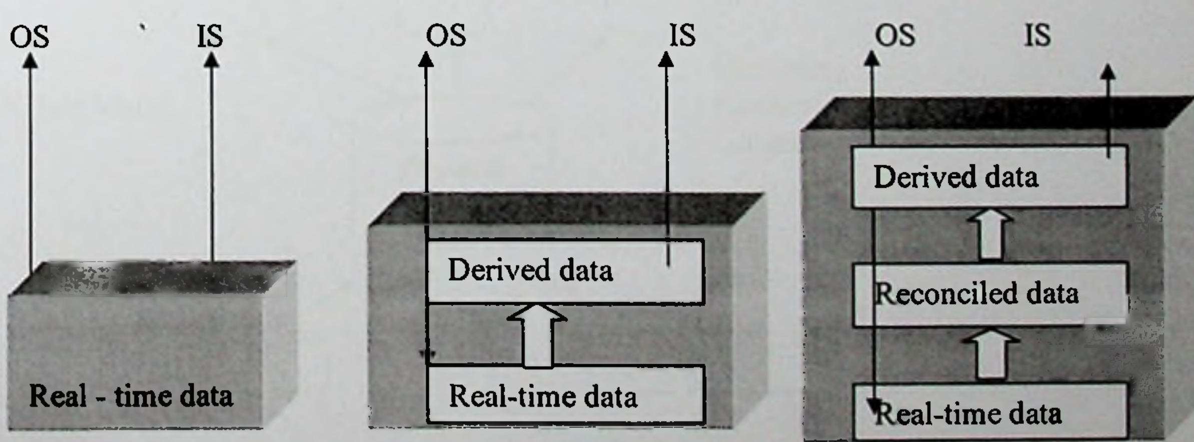


Figure 3.4: Conceptually Comparison of Data Warehouse Architecture

3.7.2 Physical view

There are three basic architectures for constructing a data warehouse, those named Centralized, Federated and Tiered. The data warehouse is distributed for load balancing, scalability and higher availability.

3.7.2.1 Centralized Architecture

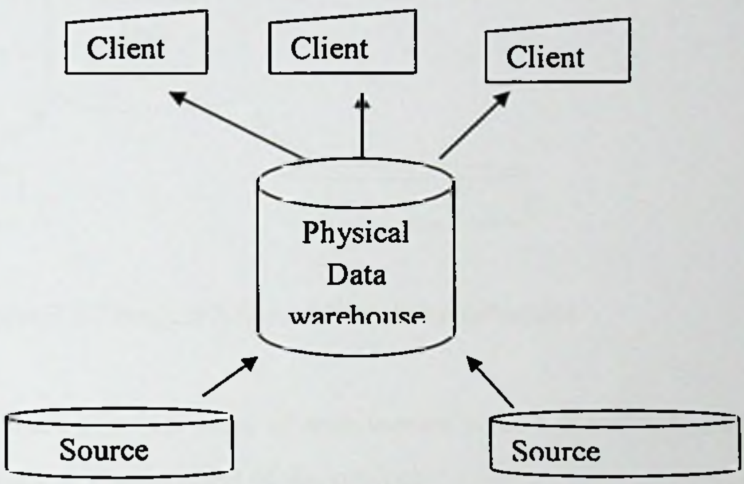


Figure 3.5 Physical View of Centralized Architecture

3.7.2.2 Federated Architecture

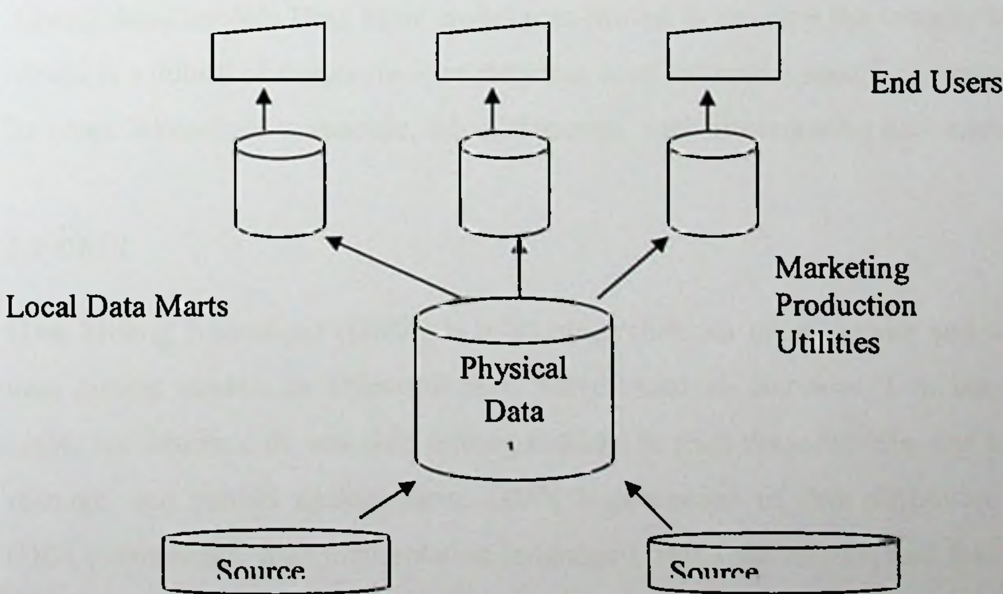


Figure 3.6 Physical View of Federated Architecture

3.7.2.3 Tiered Architecture

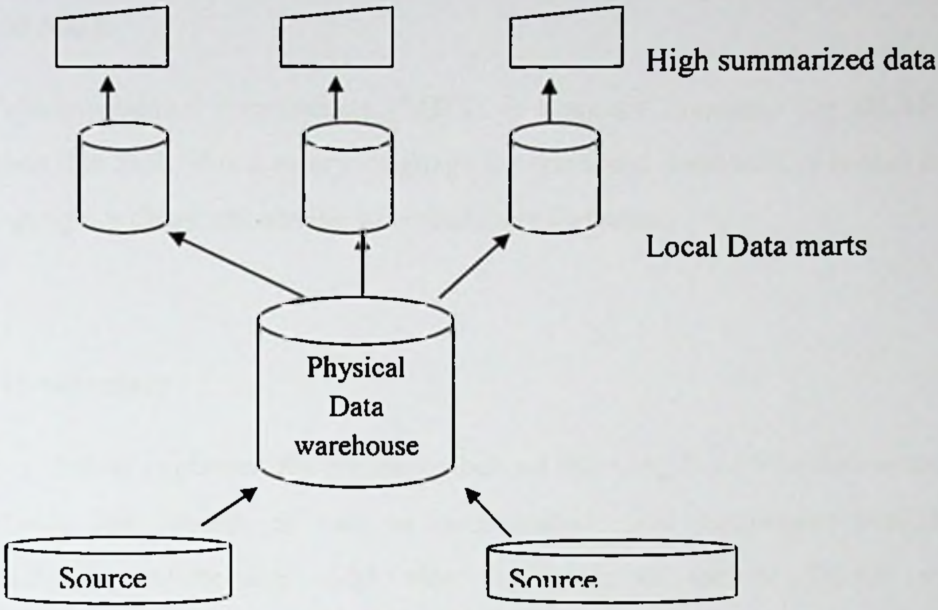


Figure 3.7 Physical View of Tiered Architecture

Among the above three physical view of architecture's the Centralized architecture is most suitable for current requirement of the project

3.8 Data warehouse Models

There are three data warehouse models exist when designing the data warehouse architecture, those are Enterprise warehouse, Data Mart and Virtual warehouse. Among these models Data Mart model was chosen to develop the system. Data Mart Model is a subset of corporate-wise data that is of value to a specific groups of users. Its scope is confined to specific, selected groups, such as marketing data mart.

3.9 DMX

Data Mining Extensions (DMX) is a language that can use to create and work with data mining models in Microsoft SQL Server Analysis Services. Can use DMX 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.

3.10 MDX

Multidimensional Expressions (**MDX**) is a **query** language for OLAP databases. Much like SQL, it is a **query** language for relational databases. It is also a calculation language, with syntax similar to spreadsheet formulas.

3.11 Summary

This chapter explained the reasoning behind choosing Data Warehouse technology to develop this project, as well as main features and component available in data warehouse architecture. And also discussed on various OLAP models and architectural design, how they different from the other architecture.

The following chapter concentrates on the selection of the data mining algorithm

Chapter 4

Data Mining Algorithms

4.1 Introduction

In the previous chapter we discussed on one of the decision support system technology and its capabilities, how is support to for this project. This chapter talks about two Data mining algorithms, which are related to project and its benefits of each of them and how/why choose those algorithms for our project.

4.2 Frequent itemset mining

Frequent itemset mining leads to the discovery of association and correlations among items in large transactional or relational data sets. With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining such pattern from their databases. The discovery of interesting correlation relationships among huge amounts of business transaction records can help in many business decision-making processes such as catalog design, cross-marketing, and customer shopping behavior analysis.

4.2.1 Association rules

Association rules are statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk.

An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent.

Association rules are created by analyzing data for frequent if/then patterns and using the criteria *support* and *confidence* to identify the most important relationships.*Support* is an indication of how frequently the items appear in the database. *Confidence* indicates the number of times the if/then statements have been found to be true [15].

4.2.2 Apriori algorithm

An association rule mining algorithm, **Apriori** has been developed for rule mining in large transaction databases. A*itemset* is a non-empty set of items.

They have decomposed the problem of mining association rules into two parts

- Find all combinations of items that have transaction support above minimum support. Call those combinations frequent itemsets.
- Use the frequent itemsets to generate the desired rules. The general idea is that if, say, ABCD and AB are frequent itemsets, then we can determine if the rule AB CD holds by computing the ratio r = support(ABCD)/support(AB). The rule holds only if r >= minimum confidence. Note that the rule will have minimum support because ABCD is frequent. The Apriori algorithm used in Quest for finding all frequent itemsets is given below

```
procedureAprioriAlg()
begin
    L1 := {frequent 1-itemsets};
    for ( k := 2; Lk-1 0; k++ ) do {
       Ck= apriori-gen(Lk-1) ; // new candidates
    for all transactions t in the dataset do {
      for all candidates c Ck contained in t do
            c:count++
      }
      Lk = { c Ck | c:count >= min-support}
    }
    Answer := kLk
End
```

It makes multiple passes over the database. In the first pass, the algorithm simply counts item occurrences to determine the frequent 1-itemsets (itemsets with 1 item). A subsequent pass, say pass k, consists of two phases. First, the frequent itemsets L_{k-1} (the set of all frequent (k-1)-itemsets) found in the (k-1)th pass are used to generate the candidate itemsetsC_k, using the apriori-gen() function. This function first joins L_{k-1} with L_{k-1}, the joining condition being that the lexicographically ordered first k-2 items are the same. Next, it deletes all those itemsets from the join result that have some (k-1)-subset that is not in L_{k-1} yielding C_k.

The algorithm now scans the database. For each transaction, it determines which of the candidates in C_k are contained in the transaction using a hash-tree data structure and increments the count of those candidates. At the end of the pass, C_k is examined to determine which of the candidates are frequent, yielding L_k . The algorithm terminates when L_k becomes empty [16].

4.3 Sales Trend Analysis

Performing sales trend analysis can give you valuable insight into the inner-workings of business. Then use data to make informed decisions, like when to raise or lower prices on your products. These decisions shouldn't always be just a "gut" feeling. While we gut can sometimes be reliable, it shouldn't be our decision making tool.

When look for trends, or patterns in sales data, then can determine both opportunities, and potential problems. Then we can track if a particular product is increasing or decreasing in sales. If it's declining, we can make timely decisions such as to cut prices, market more, or discontinue the product [17].

4.3.1 Autoregressive Integrated Moving Average (ARIMA) model

Time series analysis and forecasting have their roots originated from classical statistics. In the current context, these methods are rarely used alone as computational intelligence has become the trend. However almost all the intelligent algorithms take advantage of these statistical methodologies. So establishing a general idea about these methods is substantial in understanding modern ways of time series prediction.

ARIMA is a general statistical model which is widely used in the field of time series analysis. General ARIMA model is denoted as the ARIMA (p, d, q) where p, d and q are non-negative integers. In the above notation p parameter basically refers to the autoregressive part, d parameter refers to integrated part and the last parameter q refers to the moving average part. Based on these parameter values there are several child models which we can derived from the ARIMA model. For example if d=0 ARIMA (p, 0, q) model also referred as ARIMA. If parameter d and q both equal to zero the ARIMA (p, 0, 0) is referred as the AR model [18].

4.3.1.1Autoregressive model

ARIMA methodology attempts to describe the movements in a stationary time series as a function of what are called "autoregressive and moving average" parameters. These are referred to as AR parameters (autoregressive) and MA parameters (moving averages). An AR model with only 1 parameter may be written as...

X(t) = A(1) * X(t-1) + E(t)

where X(t) = time series under investigation

A(1) = the autoregressive parameter of order 1

X(t-1) = the time series lagged 1 period

E(t) = the error term of the model

This simply means that any given value X (t) can be explained by some function of its previous value, X (t-1), plus some unexplainable random error, E (t). If the estimated value of A (1) was .30, then the current value of the series would be related to 30% of its value 1 period ago. Of course, the series could be related to more than just one past value. For example,

X (t) = A(1) * X(t-1) + A(2) * X(t-2) + E(t)

This indicates that the current value of the series is a combination of the two immediately preceding values, X (t-1) and X(t-2), plus some random error E(t). Our model is now an autoregressive model of order 2 [19].

4.3.1.2 Moving Average Models

A second type of Box-Jenkins model is called a "moving average" model. Although these models look very similar to the AR model, the concept behind them is quite different. Moving average parameters relate what happens in period t only to the random errors that occurred in past time periods, i.e. E(t-1), E(t-2), etc. rather than to X(t-1), X(t-2), (Xt-3) as in the autoregressive approaches. A moving average model with one MA term may be written as follows...

$X(t) = -B(1) * E(t-1) + E(t)$

The term B (1) is called an MA of order 1. The negative sign in front of the parameter is used for convention only and is usually printed out auto- manically by most computer programs. The above model simply says that any given value of X (t) is directly related only to the random error in the previous period, E(t-1), and to the current error term, E(t). As in the case of autoregressive models, the moving average models can be extended to higher order structures covering different combinations and moving average lengths [19].

4.4 How to apply

The DMX (Data Mining Queries) are applying to get result from analysis server of SQL. The Analyze project have separate feature called mining models, there we can apply many model algorithms and apply dataset for tuning the algorithm.

4.5 Summary

This chapter explained the detail description of the model and its algorithm. These models used to fine pattern and predictive results in order to achieve project goal.
The next chapter discusses on approach what we are taken to do the project.

Chapter 5

Theoretical Foundation of .NET

5.1 Introduction

In the previous chapter we discussed on one of the major technology in Decision Support System that how we apply for this project. Also identify the benefits of the Data warehouse technology.

This chapter talks about benefits and how/why .net technology that can be help out for this project

5.2 Benefits of .NET Framework

Microsoft .NET (pronounced "dot net") is a software component that runs on the Windows operating system. .NET provides tools and libraries those enable developers to create Windows software must faster and easier. .NET benefits end-users by providing applications of higher capability, quality and security. The .NET Framework must be installed on a user's PC to run .NET applications [21].

.NET provides the best platform available today for delivering Windows software. It helps make software better, faster, cheaper, and more secure. .NET is not the only solution for developing Web software – Java on Linux is a serious alternative.

For developers, .NET provides an integrated set of tools for building Web software and services and Windows desktop applications. .NET supports multiple programming languages and Service Oriented Architectures (SOA).

For companies, .NET provides a stable, scalable and secure environment for software development. .NET can lower costs by speeding development and connecting systems, increase sales by giving employees access to the tools and information they need, and connect your business to customers, suppliers and partners.

For end-users, .NET results in software that's more reliable and secure and works on multiple devices including laptops, smart phones and Pocket PCs [21].

5.3 Integration with .NET

In software industry there are plenty of computer programming languages available today. In their .net is most popular software development framework as well as its support many different languages such as C#, VB, j# etc. Most of these languages can communicate with the different DBMS's

There are various features available on .net [20, 17]

- Assembles

  - An assembly is either a .DLL or .EXE that forms a part of an application. It contains MSIL code that is executed by CLR.

- Common type System

  - Common Type System (CTS) specifies the rules related to data types that languages must follow. As programs written in all languages are ultimately converted to MSIL, data types in all languages must be convertible to certain standard data types.
  - CTS are a part of cross-language integration, which allows classes written in one language to be used and extended by another language.

- Cross language Interoperability

  - .NET provides support for language interoperability. It means every program written in a language can be used by another language.

There are different types of applications that can be developed in .NET:

- Windows applications

  Typical Client/Server applications.

- Web applications

  Web sites and Intranet applications.

- Web services

  Programs those are accessible from anywhere using universal protocols like HTTP and SOAP.

- Console Applications

  Simple console applications without any GUI. Run from command prompt. Best suited to learn fundamentals and also for applications such as server sockets.

- Mobile Applications

  Contain web pages that run in mobile devices such as PDAs (Personal Digital Assistant) and Cell phones.

Because of above features and different application was help to develop a mining application using .NET framework 4.5 or higher. This is mainly supported for 100% C# (C Sharp) based implementation, Main factors to choose .Net as an implementation technology is, it is the most popular software development platform and it supports several database management systems to work with it

5.4 How to apply

The data mining application implement through C# language and need to connect analysis server and able to connect data sources and process DMX queries. There is need to add separate DLL file for connect to Analysis server.

5.5 Summary

This chapter explained the reasoning behind choosing .net framework to develop this project, as well as the main features available in .net. Subsequently it has discussed about different types of application that can be developed by using .net technology.

The Next chapter discusses on approach that present to design Data warehouse database.

Chapter 6

Development Methodologies

6.1 Introduction

In the previous chapter we discussed on one of the program technology platform .NET and language C# and its capabilities, how is support to for this project. This chapter talks about current Data warehouse solution development approaches and methodologies and benefits of each them and how/why choose one of approach to develop the this project

6.2 Overview of Approaches

There are four different approaches/methodologies available to develop data warehouse solutions. On those there are two mostly popular and are defined by defined by Ralph Kimball (Bottom-Up methodology) and Bill Inmon (Top-Down methodology). The other approaches are Hybrid approach and federated approach. We can compare these methodologies with different data warehouse architecture. Enterprise data warehouse architecture falls on Top-down methodology and Data Mart architecture falls on bottom-up methodology.

6.3The Bottom-Up Approach

Ralph Kimball designed the data warehouse with the data marts connected to it with a bus structure. The bus structure contained all the common elements that are used by data marts such as conformed dimensions, measures etc. defined for the enterprise as a whole. He felt that by using these conformed elements, users can query all data marts together. This architecture makes the data warehouse more of a virtual reality than a physical reality. All data marts could be located in one server or could be located on different servers across the enterprise while the data warehouse would be a virtual entity being nothing more than a sum total of all the data marts

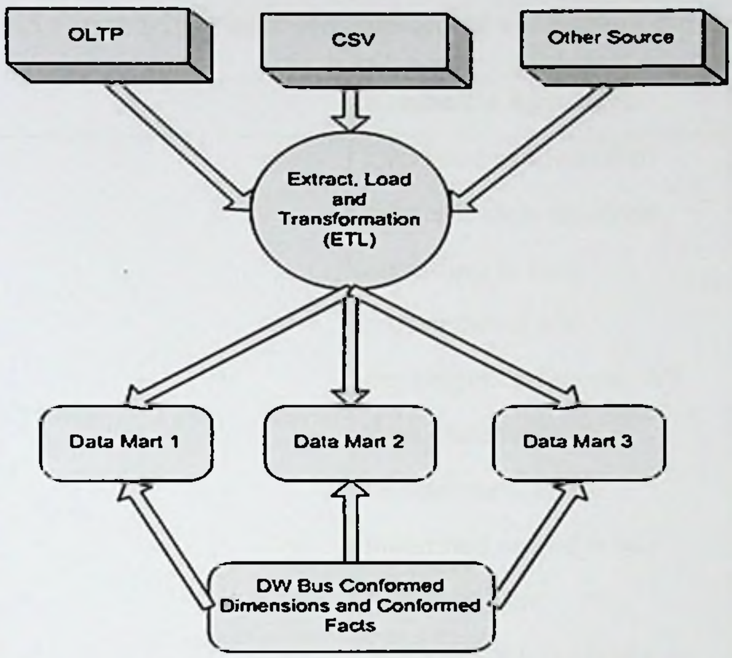


Figure 6.1: Bottom-Up Approach

6.4The Top-Down approach

This data warehouse architecture introduced by Inmon, it is the first data warehouse architecture. The first step is the extraction, transformation, migration and load of data coming from legacy systems or external sources. In the ETL process, data are collected from different sources and stored in the data staging area. After that, data and necessary metadata are loaded into data warehouse.

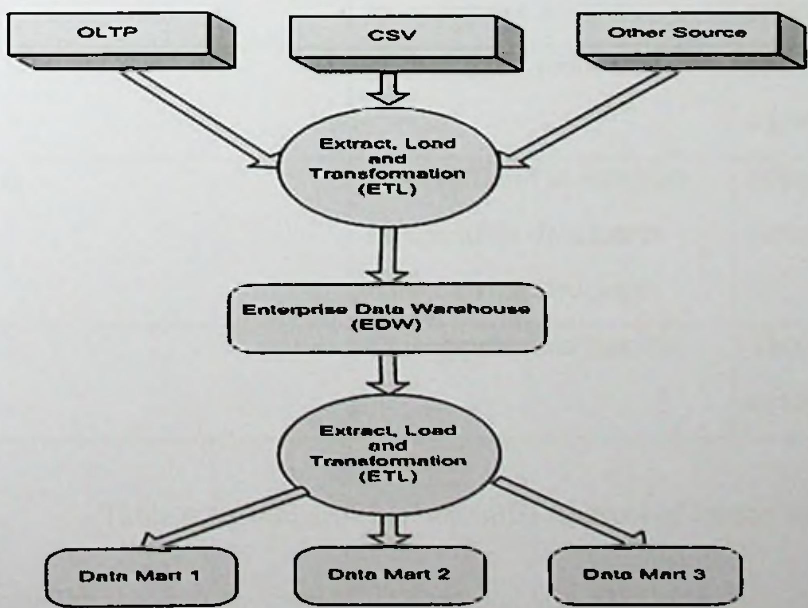


Figure 6.2: Top-down data Approach

6.5 Comparison between Top-Down vs. Bottom-Up

	Kimball's Approach	Inmon's Approach
1	Everyone is allowed to fabricate their database according to their requirements and department structure. All these independent repositories can be integrated as and when required. This methodology is known as bottom up approach	Supports a top down approach. Here no one is allowed to develop any database independently. The database for an organization should be planned and designed centrally. Every department within the organization will follow the centrally designed schema to fabricate their database
2	This structure is easier to build	The structure proposed is very typical one to craft
3	It is a nimble approach	Rigorous analysis and designing is required
4	Problematic to maintain as an enterprise resource	Easier to maintain as an enterprise resource
5	Data is often redundant	Redundancy is regulated to a great extent
6	Very difficult to integrate independent data marts with varying structure	Integration of data mart is comparatively easier
7	This approach is flexible	This approach is comparatively rigid

Table 6.1 Comparison of essential features of Inmon'sand Kimbal's

6.6 Software Process Model

A process model is a description of the sequence of the activities carried out in a Software Engineering Project, and the relative order of the activities. It provides a fixed generic framework that can be tailored to a specific project. There are many models exist, but in order to top-down approach, waterfall model is match with that approach. The Spiral model is used when Bottom-up approach is used. But here waterfall model is used; because of the final objective of the project is to making standard BI application for fast food a restaurant.

6.7 My Approach

According to client requirement they need specific applications for their divisions. Mostly they need decision support systems and Business Intelligent solutions to marketing division. There are many other divisions such as finance, production, cost control and HR. Those divisions have special requirement other than marketing division. E.g.: Marketing division need sales and customer data, but cost controller need sales and inventory and production data. Their need impose to data mart architecture that was bottom-up methodology.

6.8Data warehousing Development Process

The data presently available in various data sources is put into the target database through the data Extraction, Transformation and Loading (ETL) process. During the process, the metadata is also created. Using the front-end tools, the end-users access the data warehouse for obtaining the business intelligence. Administration tools are required for carrying out the regular admin activities on the data warehouse.

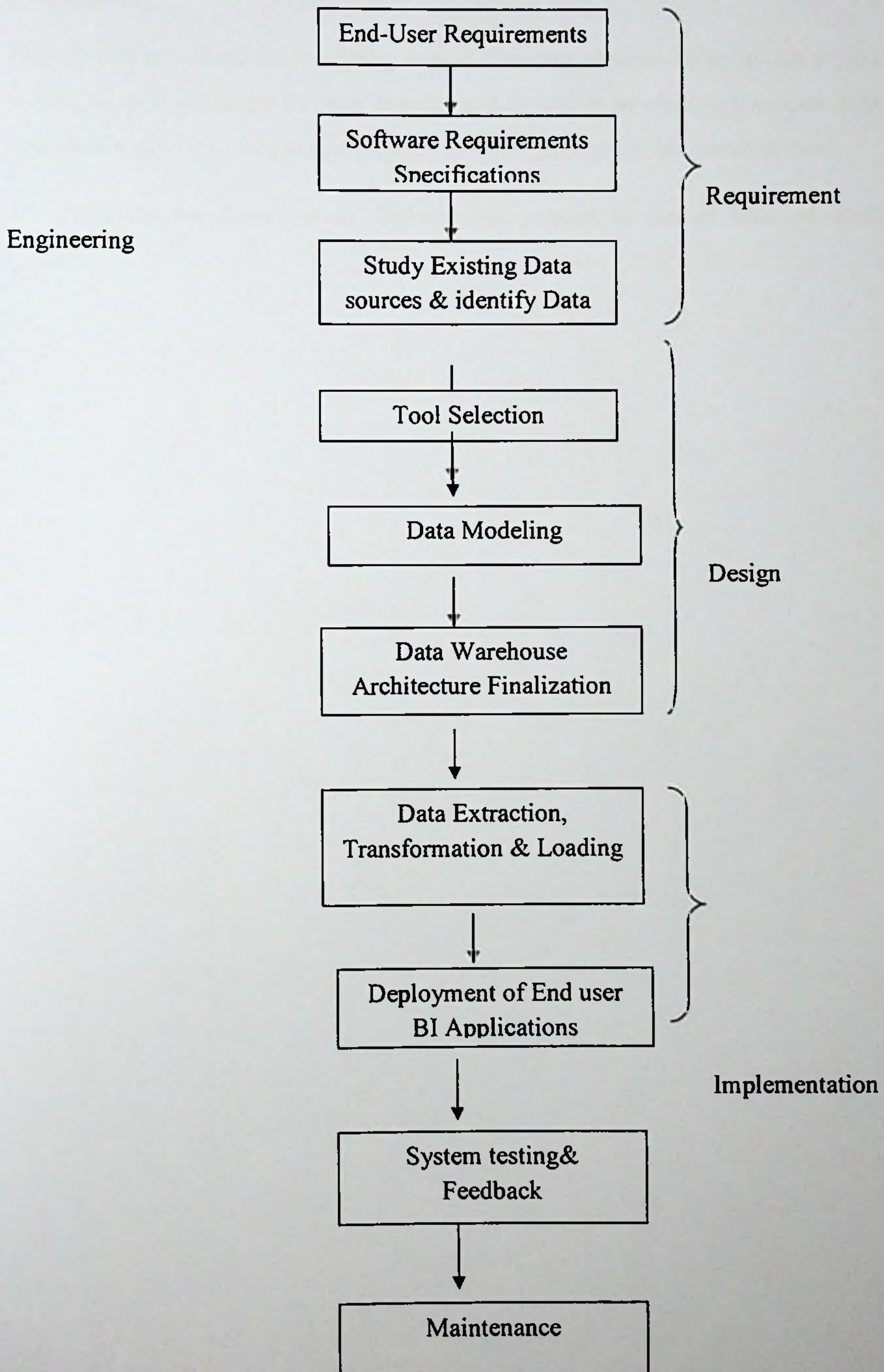


Figure 6.3: Data warehouse Development process

6.9 Summary

This chapter explained the reasoning behind choosing .Bottom-up approach to do the project, as well as discuss on main benefits and drawback in which are happen in both approaches. And also looking on project life cycle, how is affect overall project.

The Next chapter discusses on analysis that present to design Data warehouse database.

Chapter 7

Analysis

7.1 Introduction

In the previous chapter we discussed about development approaches and their benefits and drawback. We choose bottom-up approach, because the requirement of the project is concentrated into sales and deliveries data only for the analyzing purpose of the marketing department. And also look on to the Microsoft tools that are selected to use in this project.

Here we spells out the, how we analyze the existing data and what are the parameters most influenced to perform cube operation and making hierarchy level of aggregation.

7.2Requirement Analysis

The requirements are focused between marketing divisions and IT divisions. The both departments are the most valued division in the organization to improve the business functions and operations. The IT staff is a central person between service provider and their organization staff.

The marketing people they need analysis report according to their input parameters. The parameters are discussed through hourly sales, cover analysis and various sales reports. They need drill down and drill up results without going for additional reports. And they deeply concentrating in customer data and promotional items for analyzing their order pattern during specific period and identified the promotion that how effect customer order pattern and identify no of items ordered by with different quantities for specific period. And also they need annual sales and monthly sales by comparing history years by Menu Category, Main Menu, Sub Menu, order type and outlet wise.

7.2.1Existing architecture analysis

Architecture analysis describe how existing system work with current OLTP databases and servers

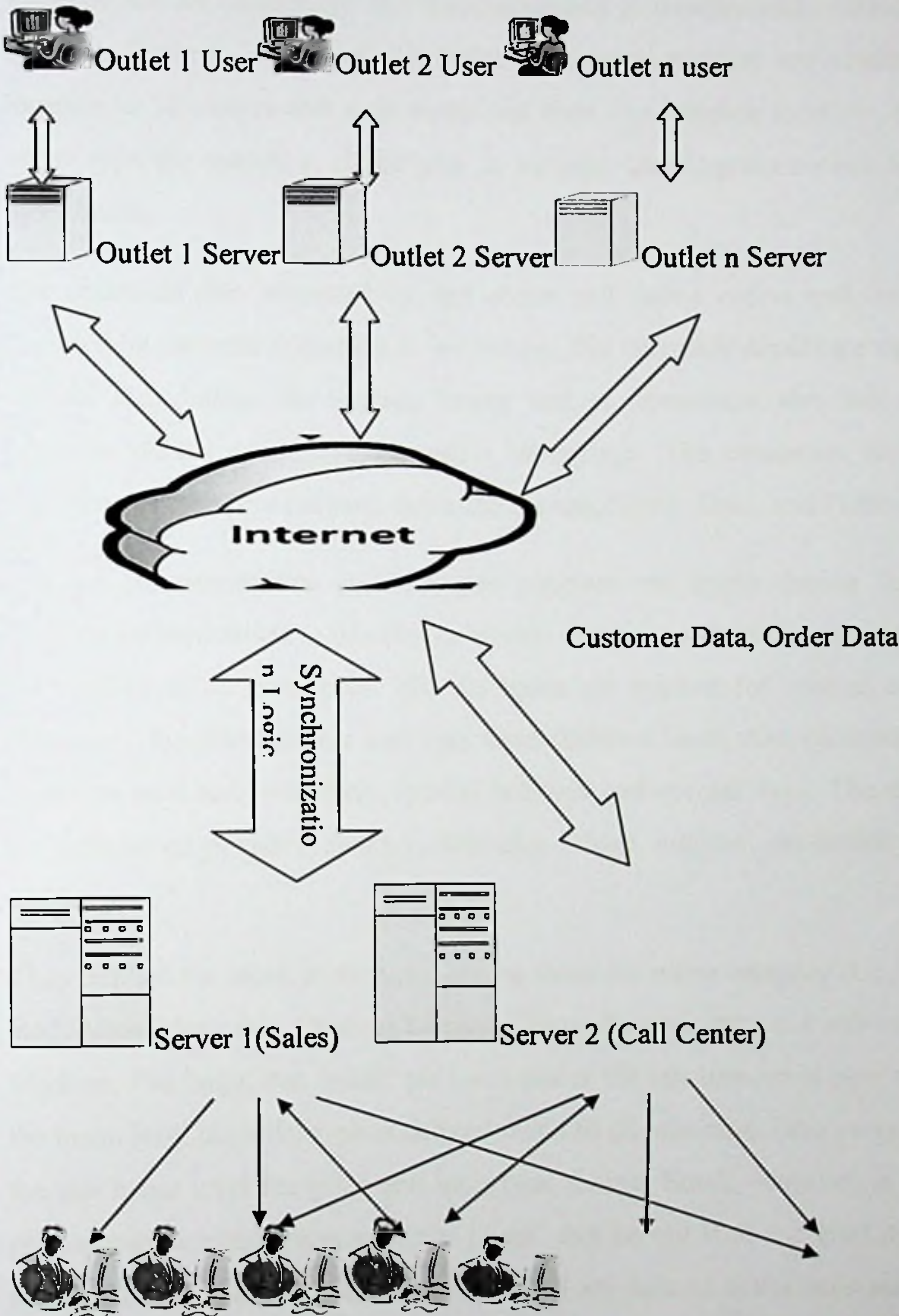


Figure 7.1: Existing System Architecture

7.2.2 Business parameters

As discuss above introduction their main business function is producing dough and selling pizzas to customers. The customers have corporate customers and individual customers. The operation time is between mornings 10.00 AM to night 11.00 PM and they serve food in different channels, those are Dine In, Take away and Delivery. There outlets are categorized in 4 ways according to standard and operation, those are Dine In, Express, Red Roof and Delco. The semi products are sending from one location to 30 outlets and each outlet has their own kitchen location. According to outlet type the operation of the time is varying, i.e.: Express outlets have to work more hours.

The operation also expanded by call center and online orders and those data also captured by database according to order type. The customer details are entered by call centers staff before the making orders and at restaurants also has entering the customer details are before the select order type. The customers are categorized according to their no of orders; those are Bronze, Silver, Gold, and Platinum.

The various promotions and discount program can apply during the year. The promotions are limited to one day to several days depend on the promotion type and it's applied to all customers. The discounts are applied for special corporate and customers, the discount rate also vary from different level. And promotion also done based on weekend, weekdays, special holidays and special days. The different days are defined via poyaday, Sunday, Saturday, Friday, mayday, pre-holiday and special holiday

They defined the menu at the three levels, those are menu category (i.e., Pizza, Pasta, etc.), Main Menu (i.e. Chicken Lovers, Cheese Lovers, etc.) and sub menu (i.e. Pan Medium, Pan large, Pan Small, etc.) and this is the last item level now they have. At the menu level the order type is defined that's be (Restaurant, Take away, both) and at the sub menu level the price and item type (Large, Small, Regular) is defined. The product can be a combination of two items , that be call Half and Half item and set of items that's be called meal deal item .They all are defined in the main menu level. But in both ways item are individually records in order tables.

The waiters and bike riders are also part of operation but they are not directly participating on the sales figures. But they also consider to make analysis for performance decisions and their standard and commitment of work task The sales figures mostly relevant on menu category, main menu, sub menu Outlet, Discount type, Day, time, quantity, rate, discount value and order type.

7.3 Source Data Analysis

The source data are available on two databases in two different servers. The sales data exist on database called "StandardInventory" and Deliveries data exist on database called "CallCenter". The sales data are exist from 2007 and deliveries data are exist from 2009. The volume of sales data are over than 30 GB and deliveries data volume are over 10 GB.

7.3.1 Source Conceptual Schema (SCS)

The Source Conceptual Schema describes the various entities and parameters that they need to run the operation within the OLTP databases. This is a data source for the data warehouse Database. The below two class diagrams are specified the data related to customer order from restaurant and through call center. And also including order cancel and delivery order cancel also.

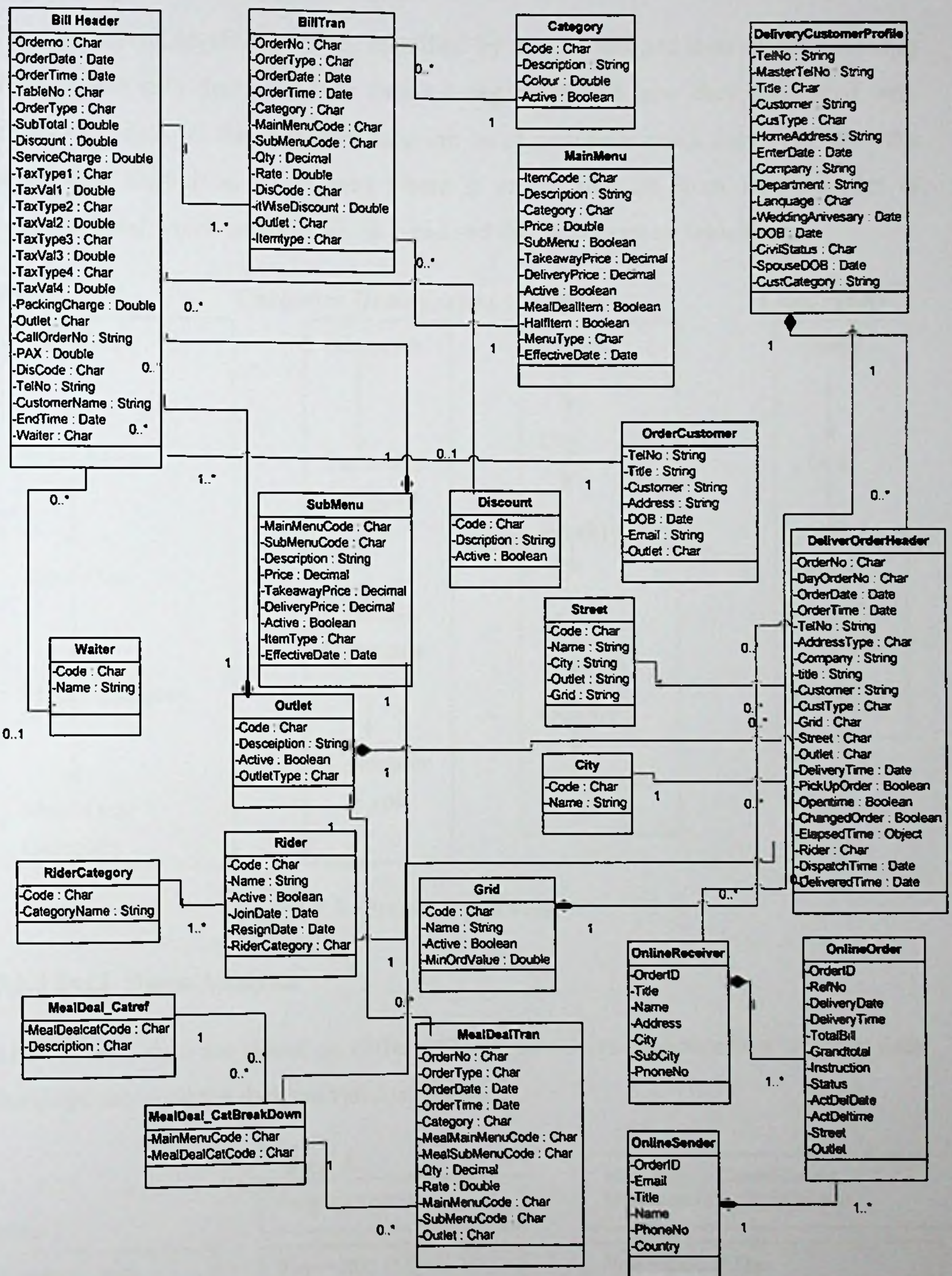


Figure 7.2: Source Conceptual Schema (SCS)

7.3.2 Hierarchical Levels

The hierarchical levels of data are identified by master data and their relations among them. These data described how data are organized and how they integrated with transaction data.i.e. the customer data can be stored with street data only. But the street data should be maintained where it exists, that be from which outlet is connected and which exists in which grid and that grid exist on which city.

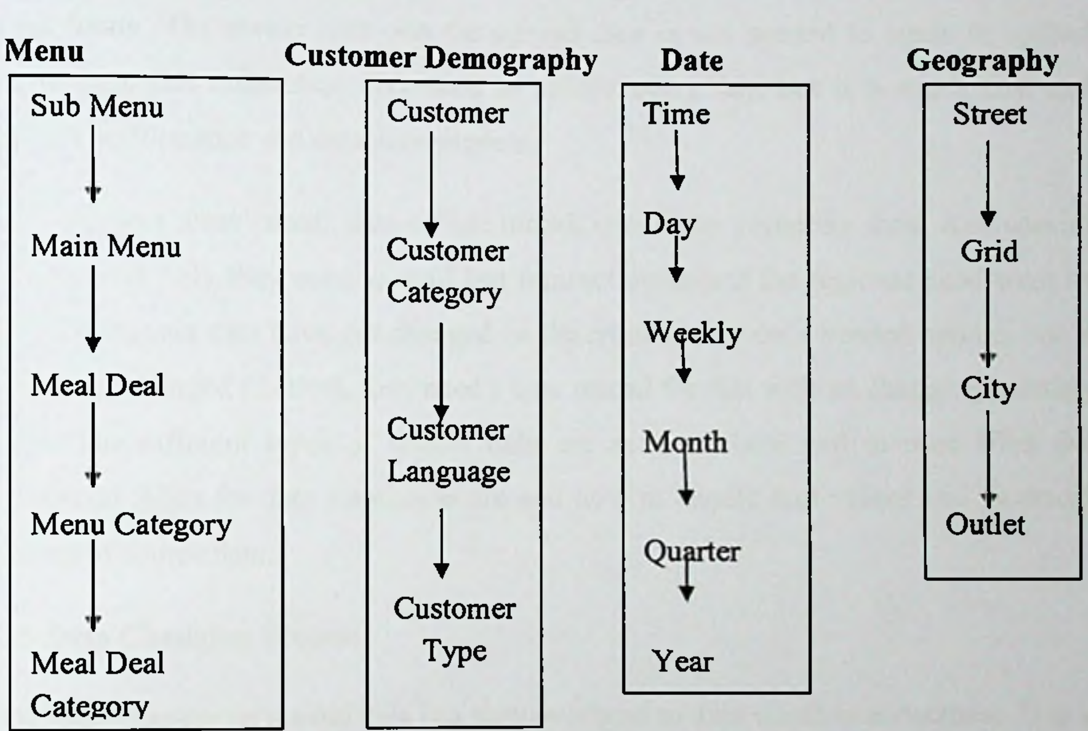


Figure 7.3 Hierarchical levels

7.3.3 Drill- Down Analyses

The business data are stored on different hierarchical level. According to those data the users can view the data via various reports.

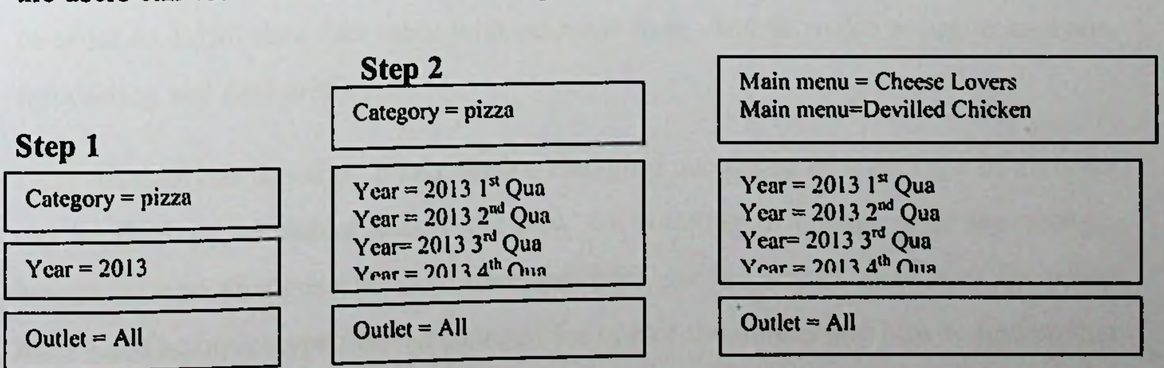


Figure 7.4DrillsDownAnalysis Steps

7.4 ETL Process

The ETL process defined how to extract, transform and load to data warehouse database through various sources. The sources are with many different formative .i.e. SQL Database, oracle Data base, csv files, XML files, Excel files and etc. According to this system the most of the data are available on Microsoft SQL database. The source is existed, so then we need to analysis on how to upload the data with what time frame. The master data one the upload then is not needed to again to upload same data. But transaction data need to upload every day, but it is much cost and affects performance and data accurateness.

The business users' needs data of last month regards to yesterday data. Any special case arrives only they need to until last transaction data if the regional head want to see. The master data have got changed on description they only needed update, but if the status changed (Active), they need a new record for that without changing existing data. The different types of master data are analyses here and marked what the important fields for data warehouse are and how to handle null values and incorrect values of source data.

7.5 Data Changing Process

The data changes on master data is a very overhead to data warehouse database. It is a very crucial part of ETL process and got many challenge to develop the ETL process. Because one of the master data is newly added or update existing data with different name, those need to reflect on data warehouse database also. Most of the master data are not change frequently I.e. Menu category. But Customer data are changed day by day via new customer. So those data must be extracted from data warehouse database in order to fulfill data fact table with relevant data. And to make accurate analysis, forecasting and data mining

Here the analysis based on how data are changing according to what type of changes can be done on to master data a and how far is taking to change after the change. Nature of type changes discussed here and what are the action taken also discussed here. I.e. The outlet type may be changed for one of the outlets and how to update that on data warehouse. The Customer data identified with need fullness of tracking history data and discuss on how to update the customer on data warehouse database according these kind of requirement. I.e. the customer city is change

7.6 Tool Selection

There are enough database vendors providing OLAP technology in their product. Choosing one of them is bit a hard by analyzing each one advantages and disadvantages and it is a very time consuming activity. The existing operational systems run under the Microsoft SQL database technology. Microsoft is a one of the market leader to providing business demanded database architectures. [9]

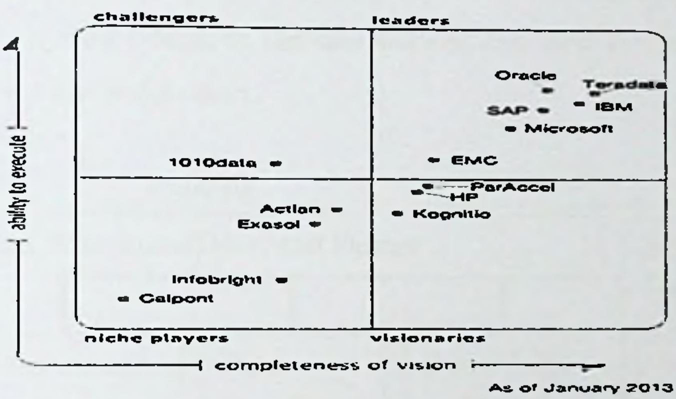


Figure 7.5: Data warehouse Database vendors

7.6.1Microsoft Corporation

Microsoft Corporation is contributing to the rapid growth of data warehousing for decision support. A combination of products from Microsoft, and from an alliance of independent software and service provides, enabled customers to operate powerful, yet affordable data warehousing systems. Microsoft SQL Server ™ database, implemented as a data warehouse storage engine, provides the price/performance, ease of management and Windows NT ® and BackOffice ® integration that makes SQL Server the preferred solution platform for many data warehouse and data mart systems.

7.6.2 Microsoft Data Warehousing Framework

The Microsoft Data warehousing Framework is to simplify the design, implementation, and management of data warehousing solutions. The designed to provide

- Open architecture that is easily integrated with and extended by third-party vendors
- Heterogeneous data import, export, validation, and cleansing services with optimal data lineage.
- Integrated metadata for data warehouse design, data extraction/transformation, server management, and end-user analysis tools
- Core management services for scheduling, storage management, performance monitoring, alerts/events, and notification

7.6.3 Framework Components

Building the data warehouse requires a set of components for describing the logical and physical design of the data sources and their destinations in the enterprise data warehouse or data mart.

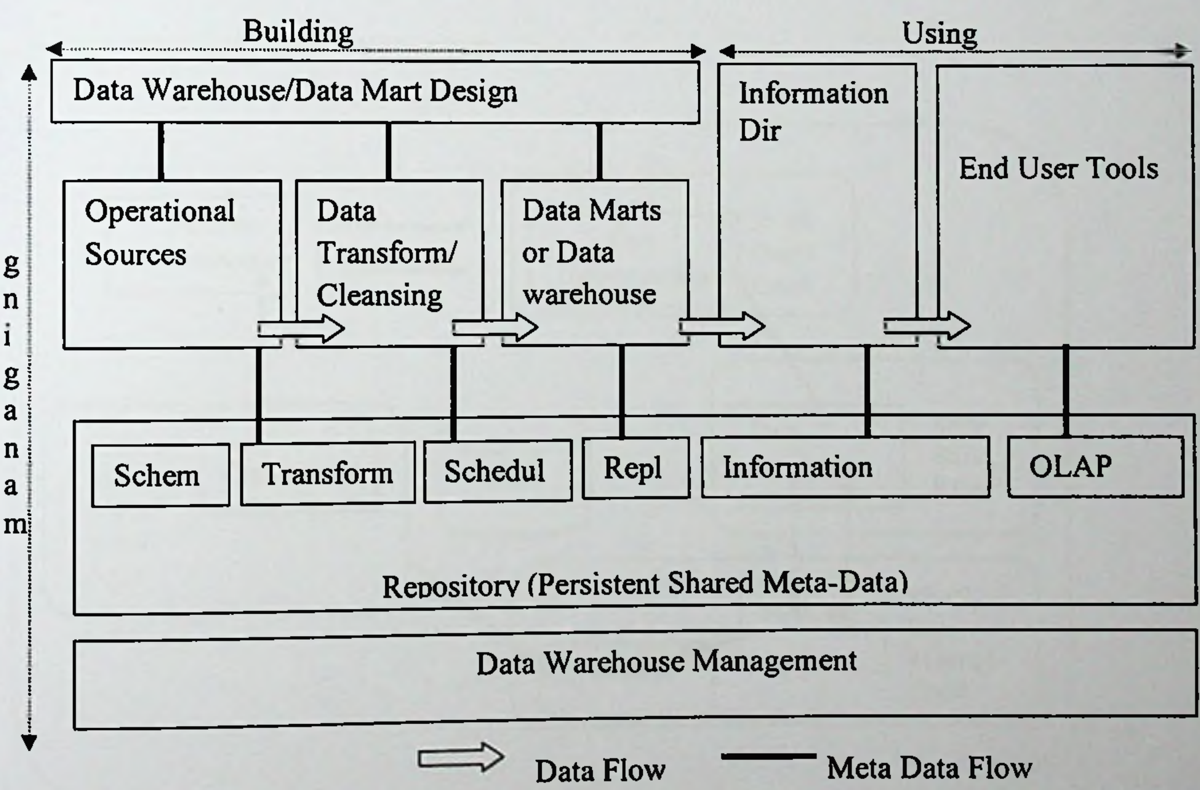


Figure 7.6: Data staging Process

7.6.4 BIDS (Business Intelligent Development studio)

Business Intelligence Development Studio is SQL Server Data Tools for Visual Studio 2013. Business Intelligence Development Studio is the primary environment that will use to develop business solutions that include Analysis Services, Integration

Services, and Reporting Services projects. Each project type supplies templates for creating the objects required for business intelligence solutions, and provides a variety of designers, tools, and wizards to work with the objects.

7.6.5 SSIS (SQL Server Integration Service)

Microsoft Integration Services is a platform for building enterprise-level data integration and data transformations solutions. That uses Integration Services to solve complex business problems by copying or downloading files, sending e-mail messages in response to events, updating data warehouses, cleaning and mining data, and managing SQL Server objects and data. The packages can work alone or in concert with other packages to address complex business needs. Integration Services can extract and transform data from a wide

7.6.6 Microsoft Data Mining Life Cycle

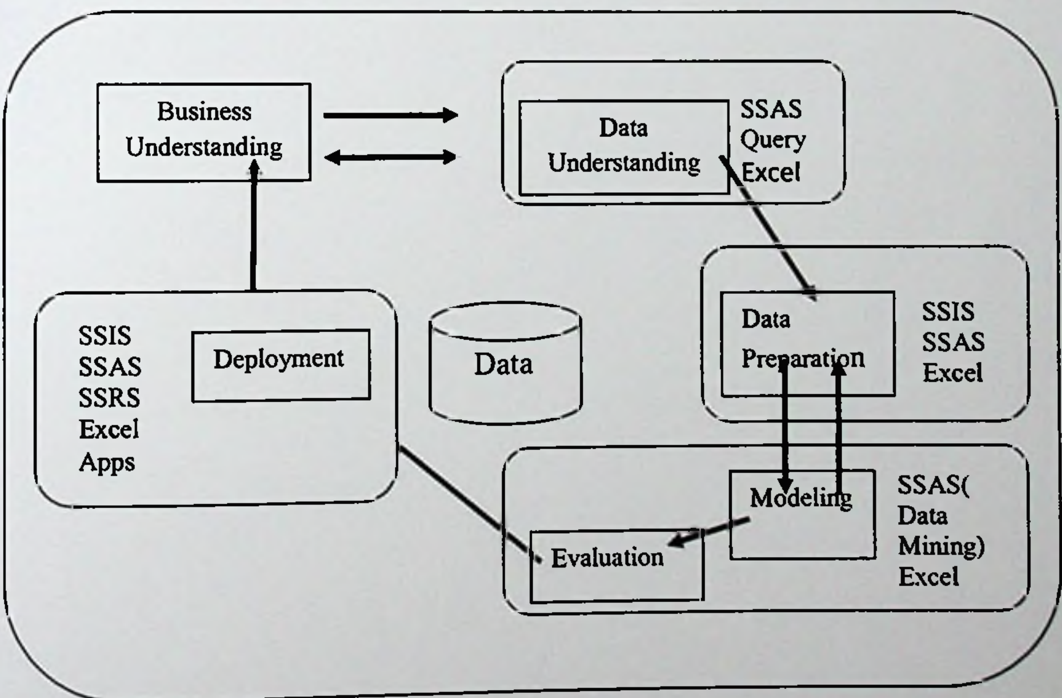


Figure: 7.7 Microsoft Data mining Life Cycle

7.7 Summary

This chapter explained the business requirement of the project and how the data are organized in the existing data sources to run the operation. And also help to identify the hierarchy level of data and atomic level of data when migrate data into data warehouse database.

The Next chapter discusses on design part of data warehouse database and overall proposed system architecture.

Chapter 8

DESIGN

8.1 Introduction

In the previous chapter we discussed about needed data to build the database for data warehouse and how to maintain history data. And also discuss on hierarchy level and drill down/up scenarios.

This chapter spells out the design those includes the use case diagram for elaborating the architecture design of data warehouse solution and designing dimensional model with lists all the facts and associated dimensions.

8.2Use Case diagram

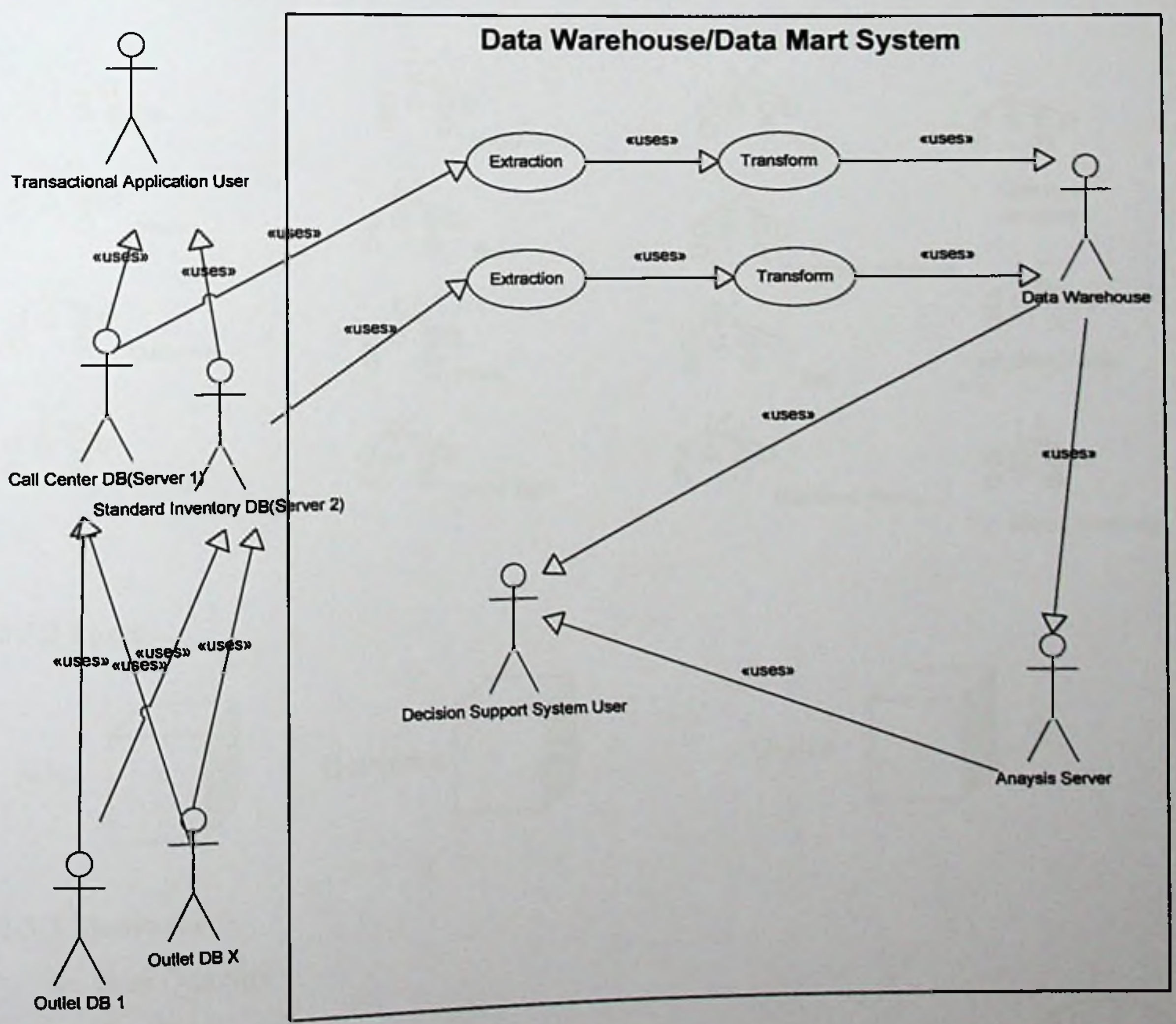
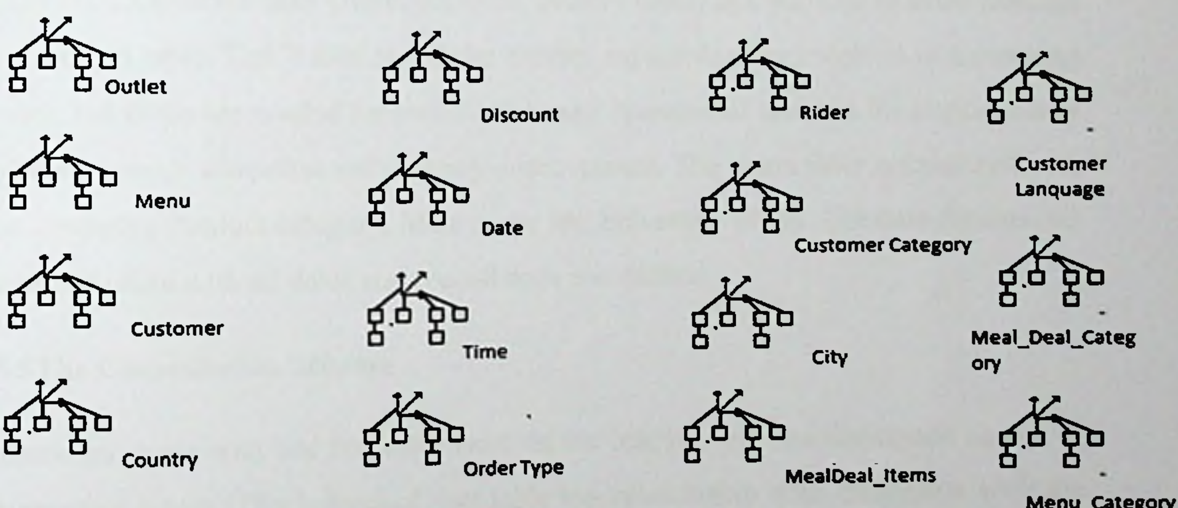


Figure 8.1 Use Case Diagram

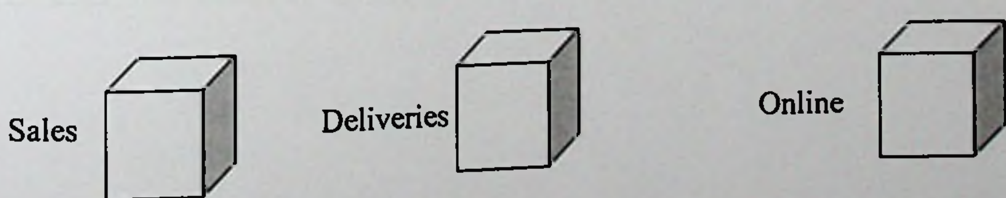
8.3Matrix Design

From the analyzing of the existing source conceptual schema the proposed system identified the requird dimensions , facts and measures and hierarchical levels of data . These are the key component part of designing a data warehouse database. The dimensions have include entities such as customer,menu,outlet,waiter, Rider and also including additional dimensions such as date and timetype for system purpose. The facts have include sales, deliveries and void sales and cancel orders. The need of void sales and cancel order data are the more valuable facts for reducing operation harshness and identifying factors related with that why those was happening. The measures is a actual outcome of the historic data, that data is a "king" for various decision making process. (Appendix A)

8.3.1 Dimensions



8.3.2 Facts



8.3.3 Measures

- ❖ Item Quantity
- ❖ Item Price
- ❖ Orderwise PAX

Figure 8.3 Dimension, facts, matrix

8.4 Design Decisions

The subject of the process chosen from sales related parameters and the grain level is chosen as time of the order for the data warehouse data base. The time type is created manually in the database and those be categorized on select query on source database data are uploaded to data warehouse database once a week. Any master level data can be changed that should be captured through CDC (Change Data Capture) and updated on to destination database. The Date, time and Country are external dimensions those are not in operational database. These dimensions are created for analysis purpose of business users to make business decisions and arrange marketing campaigns.

The order table has all orders but some parameters do not exist. Those data are available on delivery order and online order. So those tables also considered as fact table for various analyzing purposes. i.e. the city based analysis can be done by delivery facts rather than grid/street facts, country based analysis can be done through online fact table. The Waiter and Rider entities are not directly involved in measuring sales, but those are needed for performance and operational analysis for improvement of the resource allocation and delivery improvement. The menu table normalized with by emerging Product category, Main menu and Sub menu tables. The date dimensions accommodate with all dates and special days are marked.

8.5The Constellation Schema

There are more than one fact table exist on the matrix. And one dimension shared to many fact tables. The individual fact table has relationship with dimension with the Star schema, but the total design includes many fact table and dimensions so Constellation Schema design approach made for designing database.

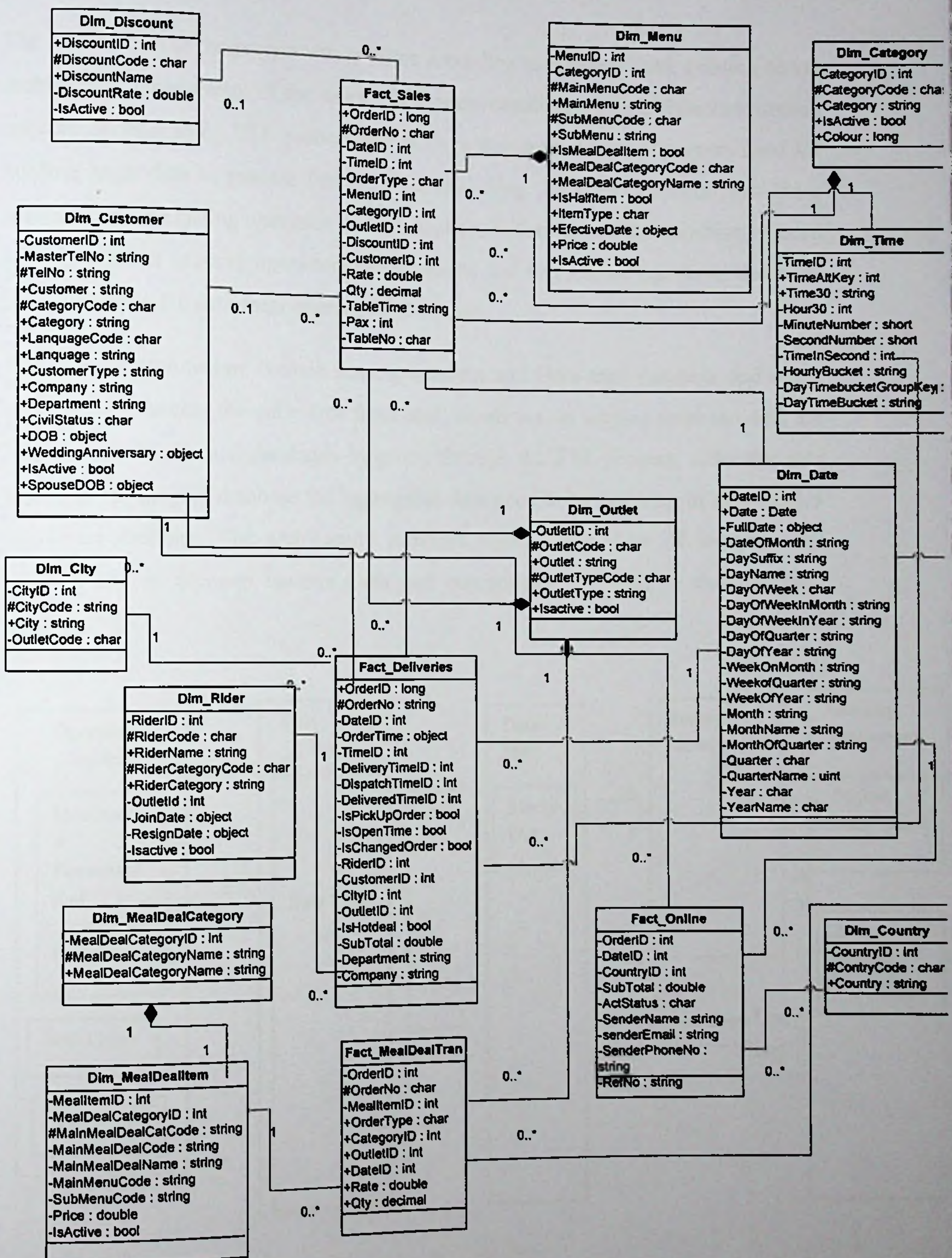
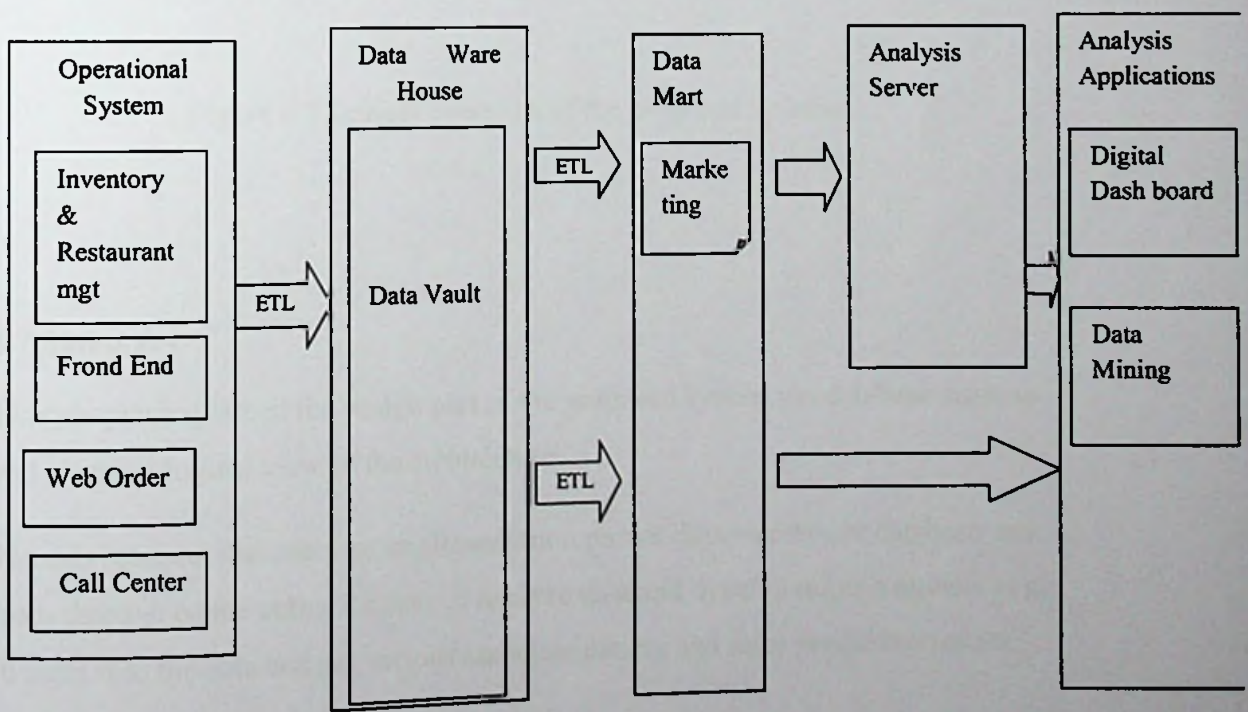


Figure 8.3 Data Warehouse Conceptual Schema (DWCS)

8.6 Proposed Architecture

The architecture of the project was analysis according to business data, existing server architecture, requirement of the users and process capability. The architecture mainly concern on how to do ETL process and is there any special type of servers need for holding huge data to process that data for analyzing. And also concern how these solutions affect existing operation and application performance. The architecture need without conflict existing operational applications and to avoid or cut down the users from accessing the data from operational database.

The propose architecture contain staging database and Data mart database and client application to access the cube data from analysis server. In staging database data are loaded from operational databases by going through the ETL process. After the data loaded to the staging database the aggregated data need to be inserting in to the Sales data mart database. The architecture provides logical over view of the proposed solution and in between various tools and components are used to develop final solution.



8.4 Proposed Data warehouse Architecture

8.6.1 The Logical Overview

The Figure 5.3 shows the logical overview of the proposed solution and that can be contained source databases, staging database, destination databases and end user's need of the solution.

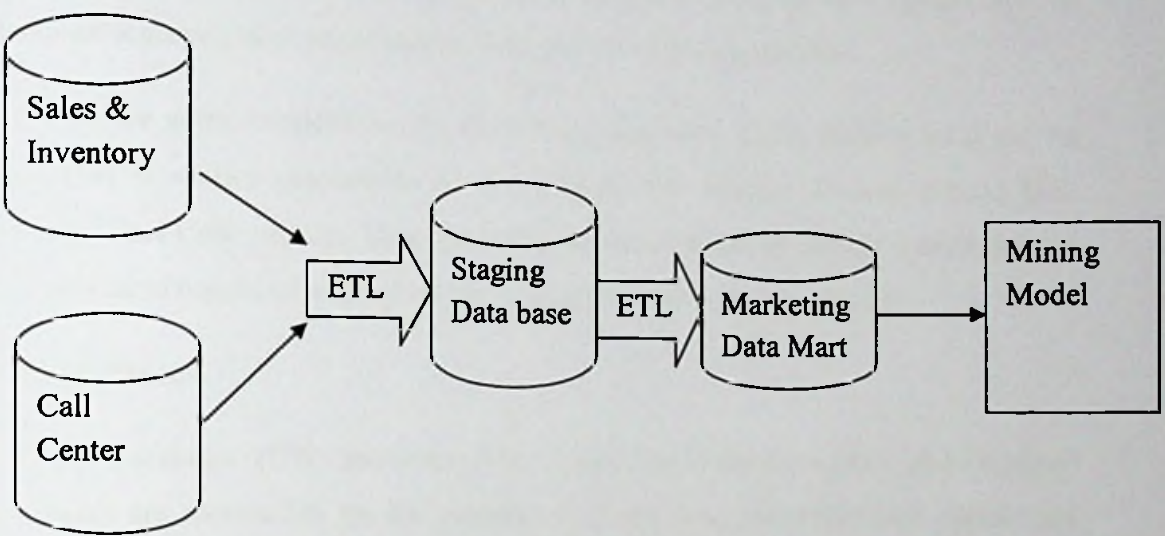


Figure 8.5 Logical overview of the proposed Solution

8.7 Summary

This chapter explained the design part of the proposed system via database schema and physical/logical view of the architecture

The next chapter discusses on implementation part of data warehouse database and from through create cubes for easy of analyze data and develop mining models to go in sight into the data and get various associate pattern and sales predictive results. .

Chapter 9

Implementation

9.1 Introduction

In the previous chapter we discussed about design of the proposed system and its database schema and physical/logical components of the architecture.

Here we are going to spells out the implementation steps of the project, what are we discussed in earlier chapters to do the project. The chapter discuss around ETL process, Data Cube process, Data mining model creation and prototype system for test the predictive results of mining models related to promotional programs.

9.2ETL Process

In Data Warehouse (DW) scenarios, ETL (Extraction, Transformation, and Loading) processes are responsible for the extraction of data from heterogeneous operational data sources, their transformation (conversion, cleaning, normalization, etc.) and their loading into the DW.

9.2.1 Develop SSIS Package

The data are extracted by developing Integration service Connection project on the SQL Server Business Intelligent Development Studio (BIDS) .The data sources are transformed to destination via the SSIS packages. Each SSIS package has their connection managers to identify the data coming by which data flow sources. The each package has steps, those named Control Flow, Data Flow, Event handlers and package Explorer.

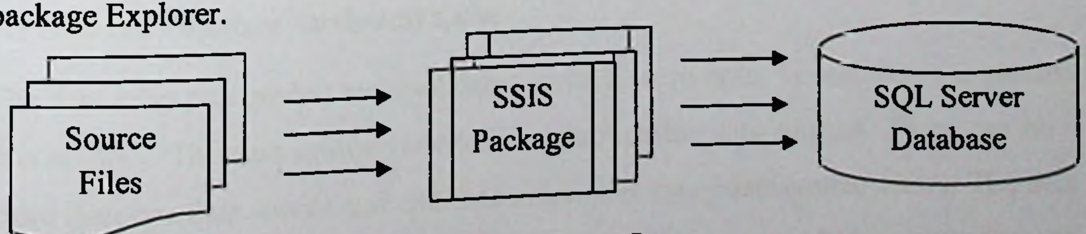


Figure 9.1 Conceptual data flow

The Control flow has Data Flow Task, the Data Flow Task have various steps in order to identify what are the source tables, SQL statements, Data conversion techniques and destination source. The bellowing screen shows the in detail level of data transformation process in data flow control within SSIS package.

The sample of Data flow diagram, which executed on SSIS package show below

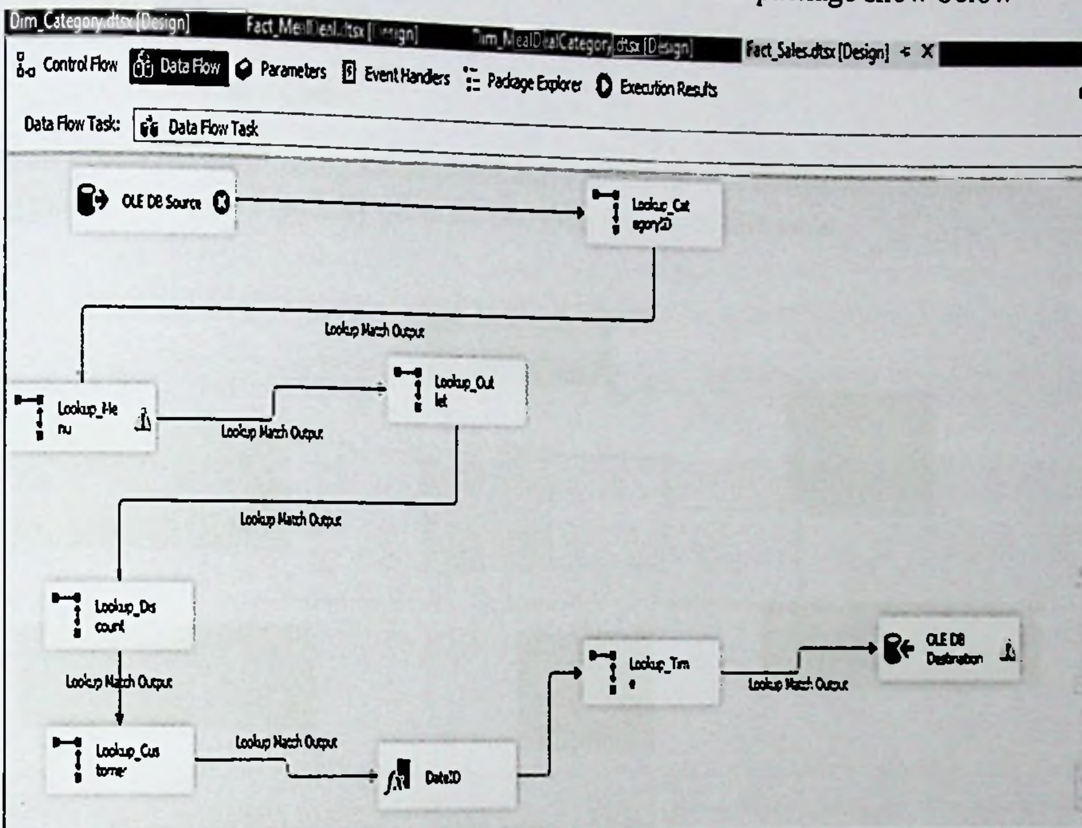


Figure 9.2 SSIS- Data Flow Task for migrating sales data

There are several packages created for transform the data from external sources. The creating of package varies package to package according to historical requirement and its data. The creations of package steps are mentioned on Appendix A.

9.3 Build & Populate Analytical Cube

The data cube process has required data source view in order to complete the process successfully. The data source view follows through the data sources. There can be a more than one data source and one data source has many data source views. The data source view generate from required fact and dimensional tables according to requirement of the aggregation data and hierarchy level.

9.3.1 Develop SSAS Project

Analysis Services Multidimensional and Data Mining Project help to create data cubes in SQL Server Data Tools for Visual Studio 2013. The detail steps and screen shots available in Appendix.

9.3.1.1Create Sales Cube

Sales cube have following Dimensions and Fact table. The dimensions are default generated/ can we manually generate and then map with fact table.

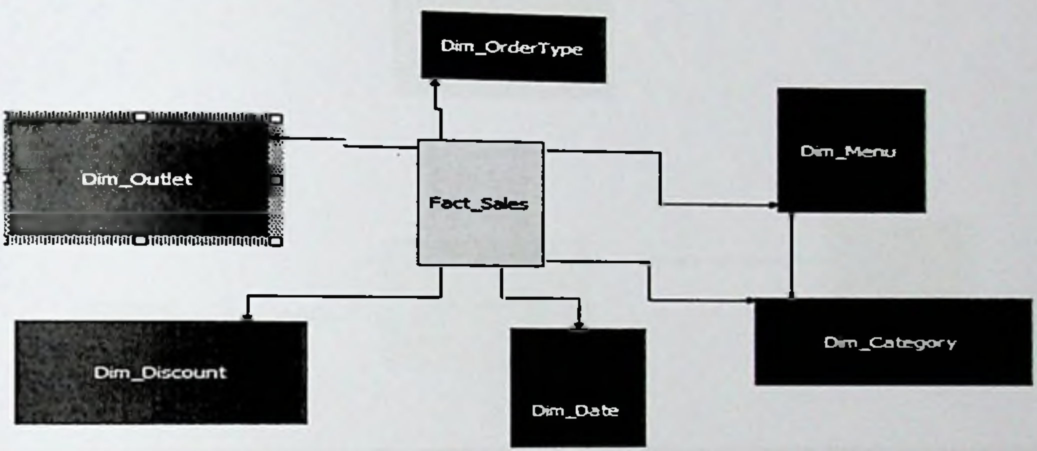


Figure 9.3 Sales Cubes – Data Source View

Date Hierarchy Level

The Hierarchies levels are important to make sum by particular attribute. Here we show the Date hierarchy level

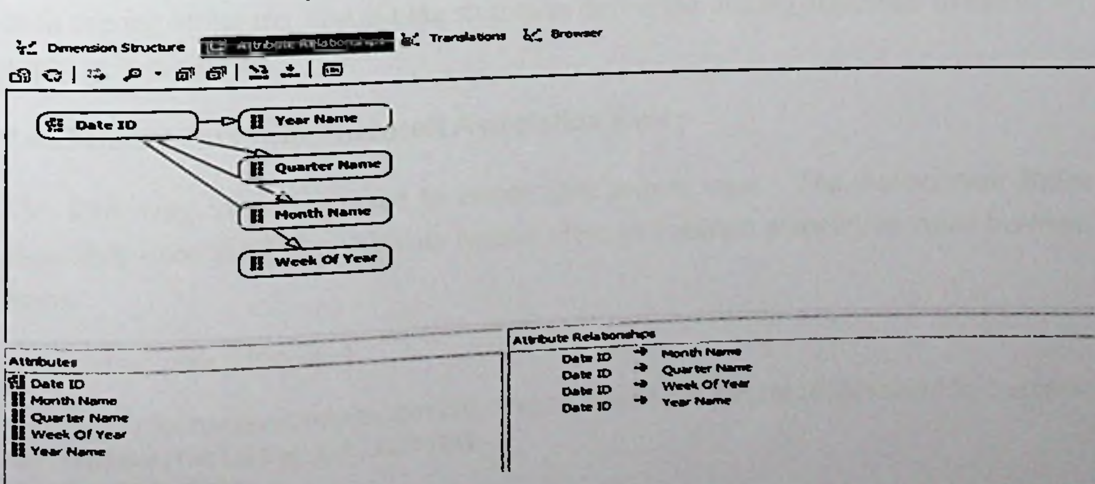


Figure 9.4 Date Hierarchy

Deploy and Browse Cube

After generating of Sales cube need to deploy the cube on analysis server. Then we can able to browsing the cube by giving relevant dimension parameters with hierarchy. Then drag and drop needed measure variables and dimension parameters from the measure group to browse the data.

9.3.1.2Create Delivery Cube

Delivery cube have following Dimensions and Fact table. The dimensions are default generated/ can we manually generate and then map with fact table.

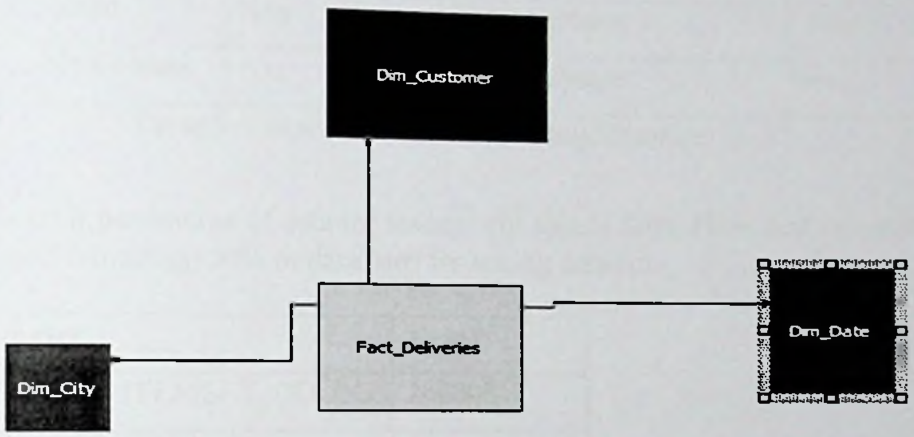


Figure 9.5 Delivery Cubes – Data Source View

9.4 Data Mining Project

There be must to a Data Source View exists to create data mining project in SSAS. The data source view can create more than one mining structure for same data set by different mining algorithm or same algorithm in different way. Mining models are created based on the Mining Structure. There can be more than one mining model for each mining structure. The mining structures define the mining algorithm to use to apply in to a data set.

9.4.1Mining data with Microsoft Association Rule

The following views are used to create data source view. The Association Rules algorithm used the following data source view to generate association rules between items.

```
CREATEVIEW [dbo].[VOrder]
AS
SELECTDISTINCTUniqueOrderNo,DateID,TimeID,OrderType,OutletID,DiscountID,Custome
rID,TableNo,TableTime,Pax,SubTotal
FROMdbo.Fact_Sales


CREATEVIEW [dbo].[VOrderMenu]
```

```
AS
SELECTdbo.Fact_Sales.UniqueOrderNo,dbo.Fact_Sales.Quantity,dbo.Dim_Menu.MenuNam
e
FROMdbo.Fact_SalesINNERJOIN
dbo.Dim_MenuONdbo.Fact_Sales.MenuID=dbo.Dim_Menu.MenuID
```

Input parameters	**Content Type**	**Name**	**Data Type**
Input table	View	VOrder	
Nested table	View	VOrderMenu	
Input Column	Unique Key	UniqueOrderNo	Text
Input Column	Key	MenuName	Text
Predictable Column	Ket	MenuName	text

Table 9.1 Association Rule - Mining Structure

The default percentage of data for testing will appear 30%, Here that value change to 50 %.and remaining 50% of data used for testing data set.

Parameter	**Value**
MAXIMUM_ITEMSET_COUNT	200000
MAXIMUM_ITEMSIZE	3
MAXIMUM_SUPPORT	1.0
MINIMUM_IMPORTANCE	-99999999
MINIMUM_ITEMSIZE	1
MINIMUM_PROBABILITY	0.1
MINIMUM_SUPPORT	0.01

Table 9.2 Association Rule - Algorithm Parameters

Algorithm parameters are set with below values. MINIMUM_PROBABILITY and MINIMUM_SUPPORT have got changed by overriding they default value.

9.4.1.1 Mining Results

The algorithm produces 22 rules for above parameters with ascending of probability.

1	Probability	Importance	Rule
2	0.322	0.538	Garlic Bread = Existing -> Pet Coca Cola = Existing
3	0.317	0.632	Sausage Delight - Personal Pan = Existing -> Devilled Chicken - Personal Pan = Existing
4	0.253	0.520	Add Cheese - Pan Small = Existing -> Devilled Chicken - Personal Pan = Existing
5	0.241	0.386	Devilled Chicken - Pan Mediam = Existing -> Pet Coca Cola = Existing
6	0.219	0.697	Devilled Chicken - Pan Large = Existing -> Jumbo Coca Cola = Existing
7	0.210	0.586	Pet Coca Cola = Existing -> Garlic Bread = Existing
8	0.207	0.702	Jumbo Coca Cola = Existing -> Devilled Chicken - Pan Large = Existing
9	0.200	0.480	Garlic Bread = Existing -> Devilled Chicken - Pan Mediam = Existing
10	0.192	0.470	Jumbo Coca Cola = Existing -> Garlic Bread = Existing
11	0.191	0.399	Jumbo Coca Cola = Existing -> Cheesy Garlic Bread Supreme = Existing
12	0.189	0.436	Add Cheese - Pan Regular = Existing -> Devilled Chicken - Pan Mediam = Existing
13	0.189	0.484	Devilled Chicken - Pan Mediam = Existing -> Garlic Bread = Existing
14	0.166	0.411	Pet Coca Cola = Existing -> Devilled Chicken - Pan Mediam = Existing
15	0.166	0.700	Devilled Chicken - Personal Pan = Existing -> Sausage Delight - Personal Pan = Existing
16	0.163	0.192	Devilled Chicken - Personal Pan = Existing -> Pet Coca Cola = Existing
17	0.142	0.563	Devilled Chicken - Personal Pan = Existing -> Add Cheese - Pan Small = Existing
18	0.141	0.488	Garlic Bread = Existing -> Jumbo Coca Cola = Existing
19	0.140	0.118	Cheesy Garlic Bread Supreme = Existing -> Pet Coca Cola = Existing
20	0.135	0.454	Devilled Chicken - Pan Mediam = Existing -> Add Cheese - Pan Regular = Existing
21	0.126	0.198	Pet Coca Cola = Existing -> Devilled Chicken - Personal Pan = Existing
22	0.122	0.420	Cheesy Garlic Bread Supreme = Existing -> Jumbo Coca Cola = Existing
23	0.105	0.122	Pet Coca Cola = Existing -> Cheesy Garlic Bread Supreme = Existing

Table 9.3 Mining rules from Association algorithm

Mining model viewer produced 11 Itemsets for Minimum itemset size 2

1	Support	Size	Item Set
2	79532	2	Garlic Bread = Existing, Pet Coca Cola = Existing
3	62901	2	Devilled Chicken - Pan Mediam = Existing, Pet Coca Cola = Existing
4	49306	2	Garlic Bread = Existing, Devilled Chicken - Pan Mediam = Existing
5	48468	2	Sausage Delight - Personal Pan = Existing, Devilled Chicken - Personal Pan = Existing
6	47582	2	Devilled Chicken - Personal Pan = Existing, Pet Coca Cola = Existing
7	41479	2	Add Cheese - Pan Small = Existing, Devilled Chicken - Personal Pan = Existing
8	39747	2	Cheesy Garlic Bread Supreme = Existing, Pet Coca Cola = Existing
9	37579	2	Devilled Chicken - Pan Large = Existing, Jumbo Coca Cola = Existing
10	35217	2	Add Cheese - Pan Regular = Existing, Devilled Chicken - Pan Mediam = Existing
11	34815	2	Jumbo Coca Cola = Existing, Garlic Bread = Existing
12	34638	2	Jumbo Coca Cola = Existing, Cheesy Garlic Bread Supreme = Existing

Table 9.4Mining Item sets

9.4.1.2 Dependency Network

Dependency Network shows the links between menus. The strongest links is starting with Deviled Chicken-Pan Large with Jumbo Coca Cola. After that Devilled Chicken-

Personal Pan with Sausage Delight-Personal Pan Then Devilled Chicken-personal Pan with Add Cheese-Pan Small and Garlic Bread with Pet Coca Cola.

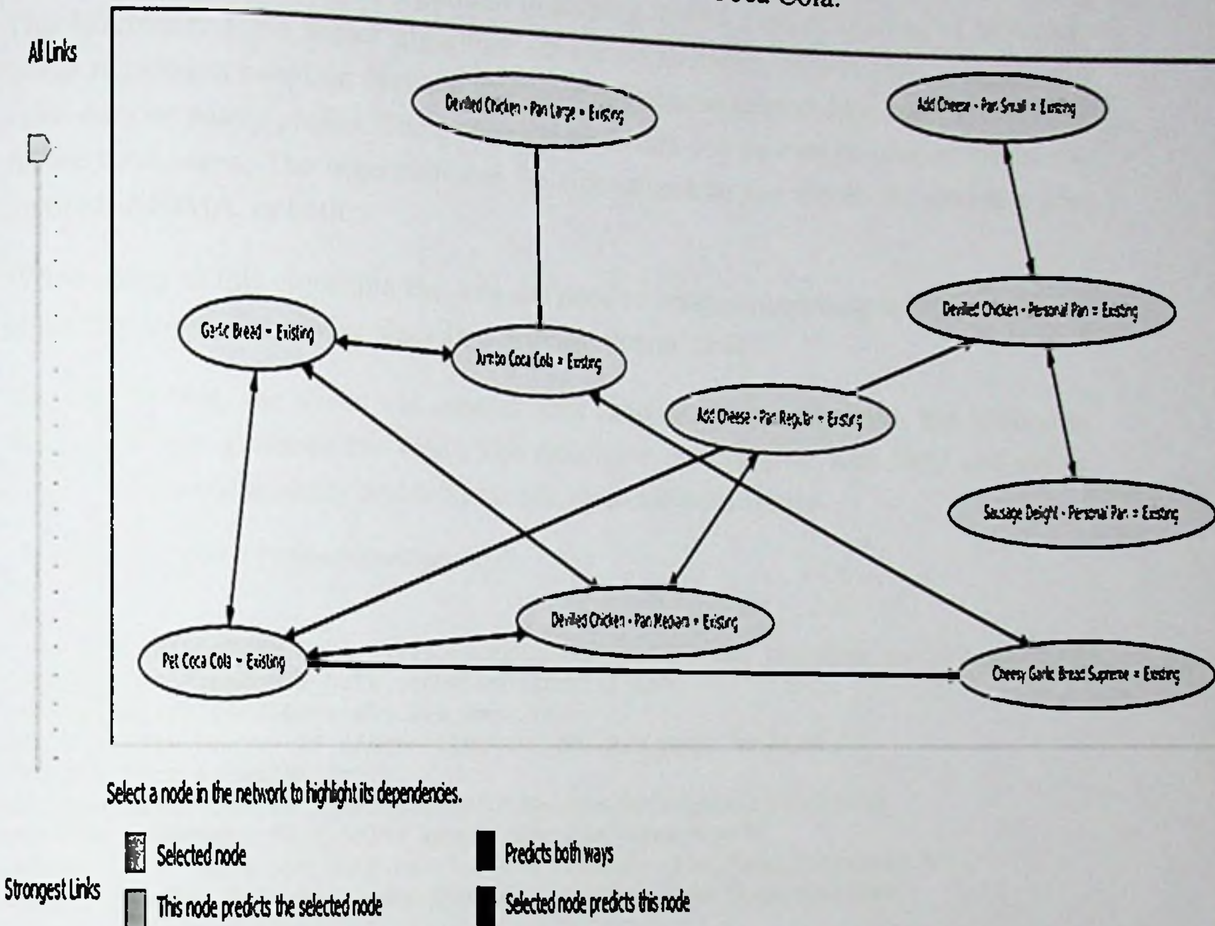


Figure 9.6 Dependency Network

So as from Association rule we are able to find which the items are mostly sold with another item. Minimum probability, minimum support and maximum item set variables/parameters used to find those items.

From the results of item set we can identify the order pattern by customers. So when make promotional strategies consider those items to adding one of them as free item with another selling menu or can be any else of discount or cross selling . And also we can the item, which have more support in a dataset. That we can identify with minimum item set 1. From that we can design effective promotional program for valuable customers. And also need consider on those item quality and perception level of new customer.

Identifying the purchase pattern is only show the support of the item with other item within order but it's not able to mention frequent of sales in particular period. Identifying or predicting the sales of item period wise is help to make plan of inventory and other wise is able to give idea about planning promotional strategies for increase of sales in particular less sales period for those items.

9.4.2 Mining data with Microsoft Time Series

The Microsoft Times Series algorithm is helping to find predicting of sales in future. The Microsoft Time Series algorithm uses a combination of ARIMA analysis and linear regression based on decision trees to analyze time-related data, such as monthly sales data or yearly profits. The patterns it discovers can be used to predict values for future time steps. The algorithm can be customized to use either the decision tree method, ARIMA, or both.

When using of this algorithm the data are need to arrange according to expectation of algorithm input, Otherwise algorithm provide wrong result.

In order to that, the view was created with amount of quantity sold, the following SQL statement generate the view. The data have on and after year 2012 and active items, the sum of quantity group by month, year and menu name.

```
CREATEVIEW [dbo].[VMenuSalesByMonth]
AS
SELECTTOP
(100)PERCENTSUM(dbo.Fact_Sales.Quantity)ASQty,CAST(dbo.Dim_Date.MonthASsmallint
)ASMonth,CAST(dbo.Dim_Date.YearASsmallint)ASYear,
dbo.udfBuildISO8601Date(dbo.Dim_Date.Year,
dbo.Dim_Date.Month, 28)ASReportingDate,dbo.Dim_Menu.MenuName
FROMdbo.Fact_SalesINNERJOIN
dbo.Dim_DateONdbo.Fact_Sales.DateID=dbo.Dim_Date.DateIDINNERJOIN
dbo.Dim_MenuONdbo.Fact_Sales.MenuID=dbo.Dim_Menu.MenuID
WHERE        (dbo.Dim_Date.Year>='2012')AND(dbo.Dim_Menu.IsActive= 1)
GROUPBYdbo.Dim_Date.Month,dbo.Dim_Date.Year,dbo.Dim_Menu.MenuName
```

	Qty	Month	Year	ReportingDate	MenuName
1	27.00	6	2014	2014-06-28 00:00:00.000	Chocolate Cone
2	61.00	11	2012	2012-11-28 00:00:00.000	Add Pineapple - Pan Small
3	53.00	7	2014	2014-07-28 00:00:00.000	4 Coke Glass for Promo
4	120.50	7	2014	2014-07-28 00:00:00.000	Hot & Spicy Chicken Mayo - Pan Large
5	1.00	12	2012	2012-12-28 00:00:00.000	Add Capsicum - Sausage Crust Large
6	323.00	4	2012	2012-04-28 00:00:00.000	Add Prawns - Pan Regular
7	306.00	11	2013	2013-11-28 00:00:00.000	Prawn in Cre Sau Pasta TW - Large
8	169.00	1	2012	2012-01-28 00:00:00.000	GB Sup Sup - Chicken - Large
9	14.50	8	2013	2013-08-28 00:00:00.000	Hot & Spicy Chicken Mayo - Sausage C
10	1.00	6	2012	2012-06-28 00:00:00.000	Tandoori Chicken Mayo - Pan Mediam
11	125.00	2	2014	2014-02-28 00:00:00.000	Tre F. S. & Sweet Corn - Tradizionale Ita
12	733.00	3	2012	2012-03-28 00:00:00.000	Pet Fanta Portello
13	1102...	7	2014	2014-07-28 00:00:00.000	Spicy Seafood - Pan Large
14	41.00	4	2014	2014-04-28 00:00:00.000	Devilled Fish Mayo - Personal Pan
15	105.00	6	2014	2014-06-28 00:00:00.000	Add Pineapple - Pan Regular
16	50.50	7	2013	2013-07-28 00:00:00.000	Sausage Delight Mayo - Pan Mediam
17	63.00	12	2013	2013-12-28 00:00:00.000	Large Sausage Delight
18	783.00	3	2014	2014-03-28 00:00:00.000	Add BBQ Chicken - Large

Figure 9.7 Sum of Item sales by Month

Input parameters	Content Type	Name	Data Type
Input table	View	VMenuSalesByMonth	
Input Column	key	Menu Name	
Input Column	Unique Key	Reporting Date	Text
Input Column	Key	Month	numeric
Input Column	Key	Qty	Numeric
Predictable column	key	Qty	Numeric

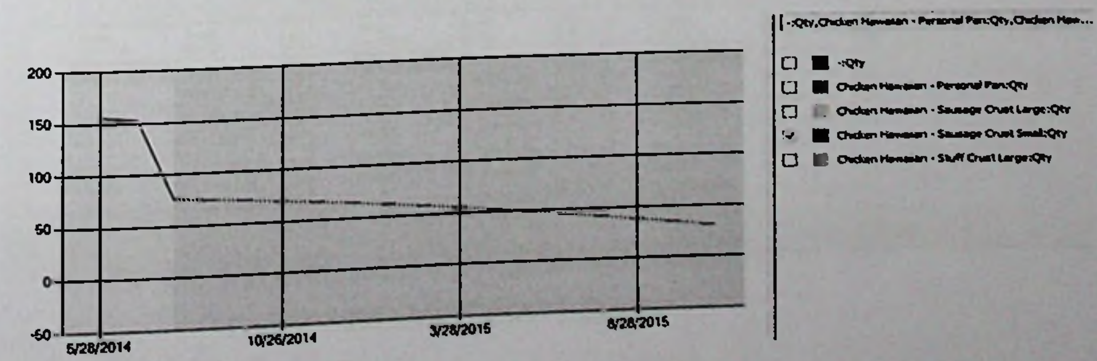
Table 9.5 Time Series - Mining Structure

The following parameters are used when data process, for the missing value, the Mean is used to fill the value.

Parameter	Value
AUTO_DETECT_PERIODICITY	0.6
COMPLEXITY_PENALTY	0.1
FORECAST_METHOD	MIXED
HISTORIC_MODEL_COUNT	1
HISTORIC_MODEL_GAP	10
MINIMUM_PROBABILITY	0.1
INSTABILITY_SENSITIVITY	1.0
MAXIMUM_SERIES_VALUE	+1E08
MINIMUM_SERIES_VALUE	-1E08
MINIMUM_SUPPORT	10
MISSING_VALUE_SUBSTITUTION	Mean
PERIODICITY_HINT	{12}
PREDICTION_SMOOTHING	0.5

Table 9.6 Time Series - Mining Algorithms

Below graph show the sum of items sales prediction until 2016



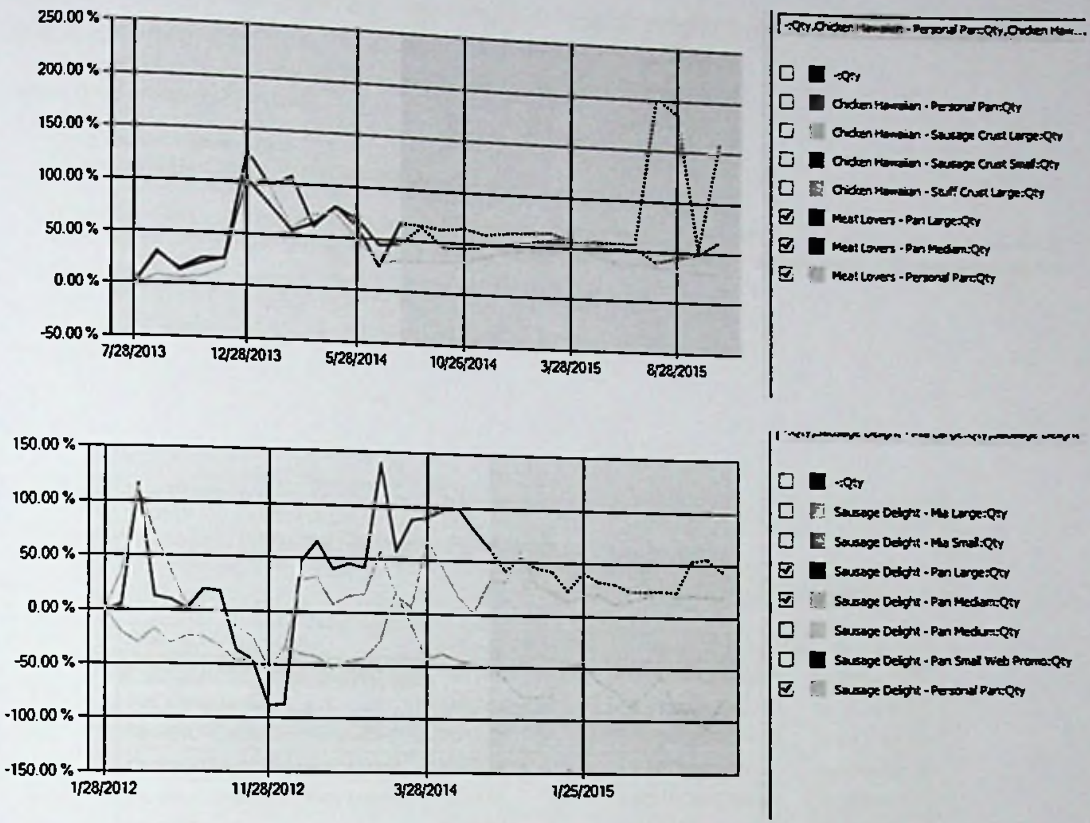


Figure 9.8 Time Series Graph

All this results are getting from mining model viewer. But business user has not permission to access SSAS. So there is need to build the application to business users able to see prediction results or test results with actual values from transactional database. The simple application builds as a prototype to view results.

9.6 User Interface Development for Data Mining

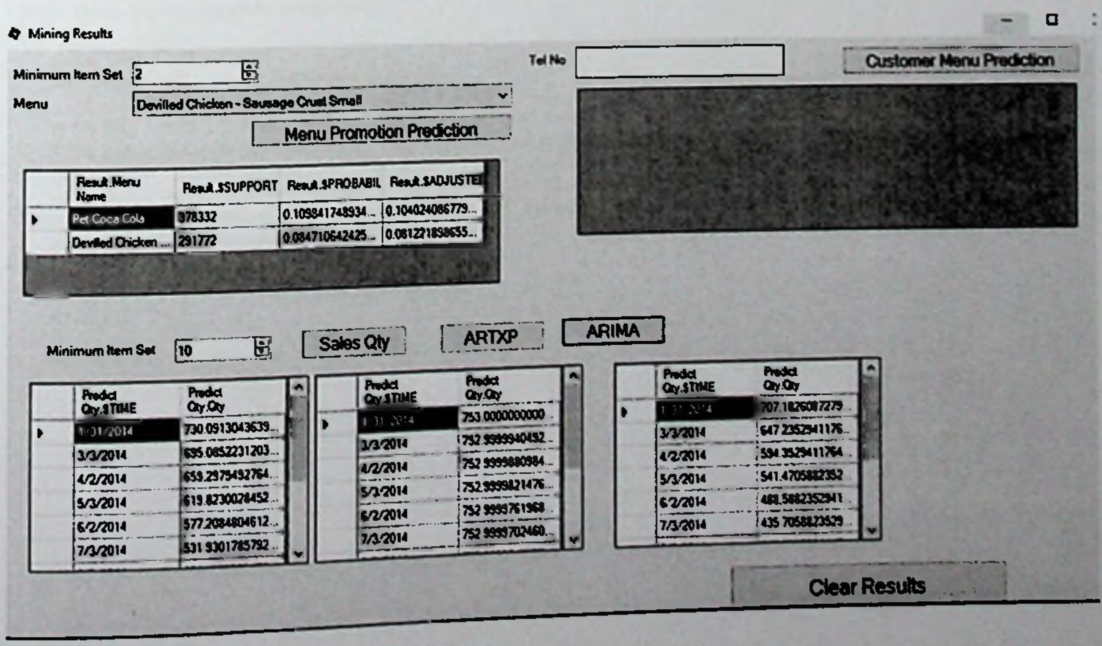
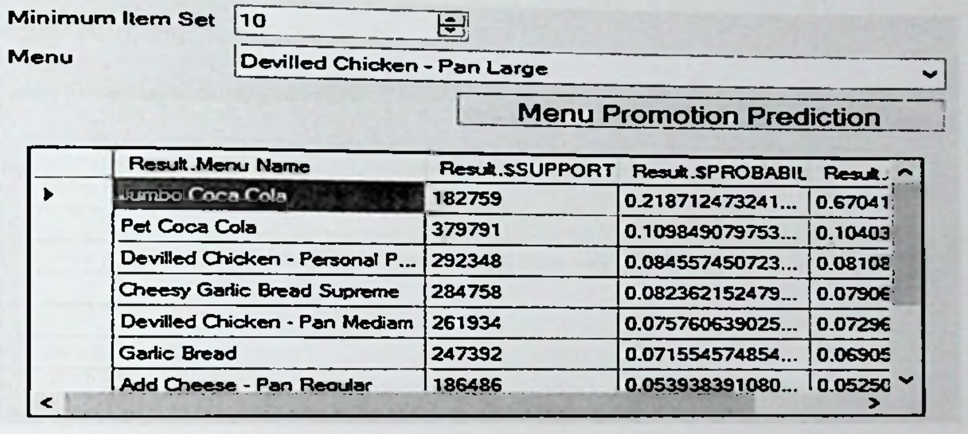


Figure 9.9Mining Results Interface

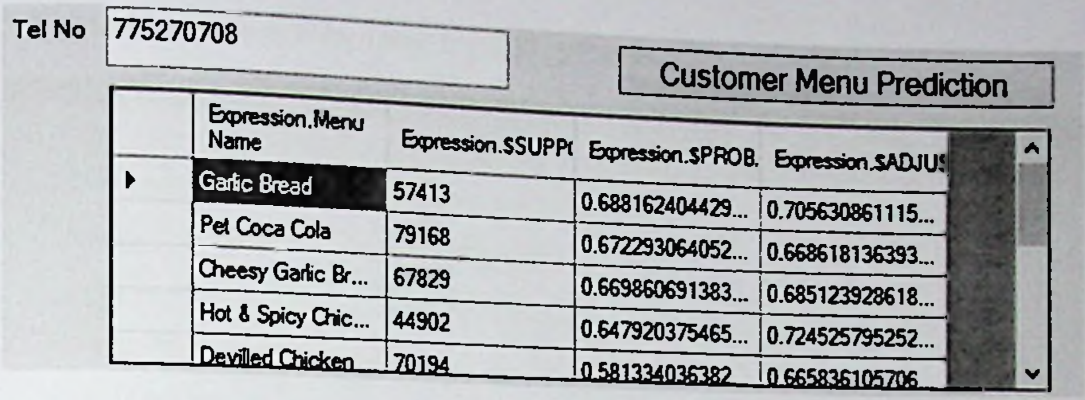
The bellowing code generates the results of menu prediction for selected menu with minimum item set.

```
public string GenerateMenuPrediction()
{
    string menuName = cmbMenu.Text;
    int minimumItemSet = Convert.ToInt16(minItemSet.Value);

    string DMX1 = "SELECT flattened PredictAssociation([OrderMenu_AR].[V Order Menu],INCLUDE_STATISTICS, " + minimumItemSet + ")  as Result ";
    string DMX2 = " From[OrderMenu_AR] NATURAL PREDICTION JOIN (SELECT (SELECT '" + menuName + "' AS [Menu Name]) AS [V Order Menu]) AS t";

    string DMX = "";

    DMX = DMX1  + DMX2;
    return DMX;

}
```



The bellowing code generates the results of menu prediction for entered telephone number for particular customer with minimum item set.

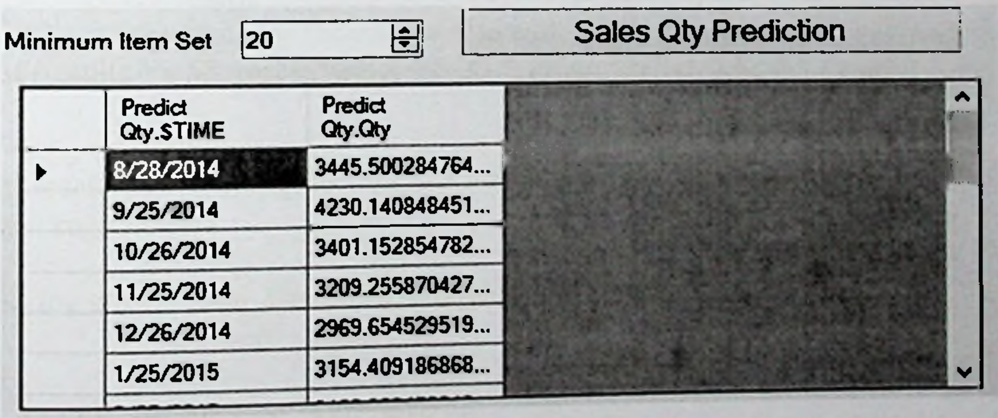
```
public string GenerateCustometrMenuPrediction()
{
    string telno = txtTelNo.Text;
    int minimumItemSet = Convert.ToInt16(minItemSet.Value);

    string DMX1 = "SELECT flattened  PredictAssociation([CustomerOrderMenu_AR].[V Customer Order], INCLUDE_STATISTICS, " + minimumItemSet + ") ";
    string DMX2 = " From [CustomerOrderMenu_AR] PREDICTION JOIN SHAPE {OPENQUERY([Pizza_DatawareHouse],'SELECT[MasterTelNo], ";
    string DMX3 = " [CustomerID] FROM [dbo].[Dim_Customer] ORDER BY [CustomerID]')} APPEND ({OPENQUERY([Pizza_DatawareHouse], ";
    string DMX4 = "'SELECT[MenuName],[CustomerID] FROM [dbo].[VCustomerOrder] ORDER BY [CustomerID]')} RELATE[CustomerID] ";
    string DMX5 = " TO [CustomerID]) AS [VCustomerOrder] AS t ON [CustomerOrderMenu_AR].[Master Tel No] = t.[MasterTelNo] AND ";
    string DMX6 = " [CustomerOrderMenu_AR].[V Customer Order].[Menu Name] = t.[VCustomerOrder].[MenuName] WHERE ";
    string DMX7 = " t.[MasterTelNo] ='" + telno + "'";

    string DMX = "";

    DMX = DMX1 + DMX2 + DMX3 + DMX4 + DMX5 + DMX6 + DMX7;
    return DMX;

}
```



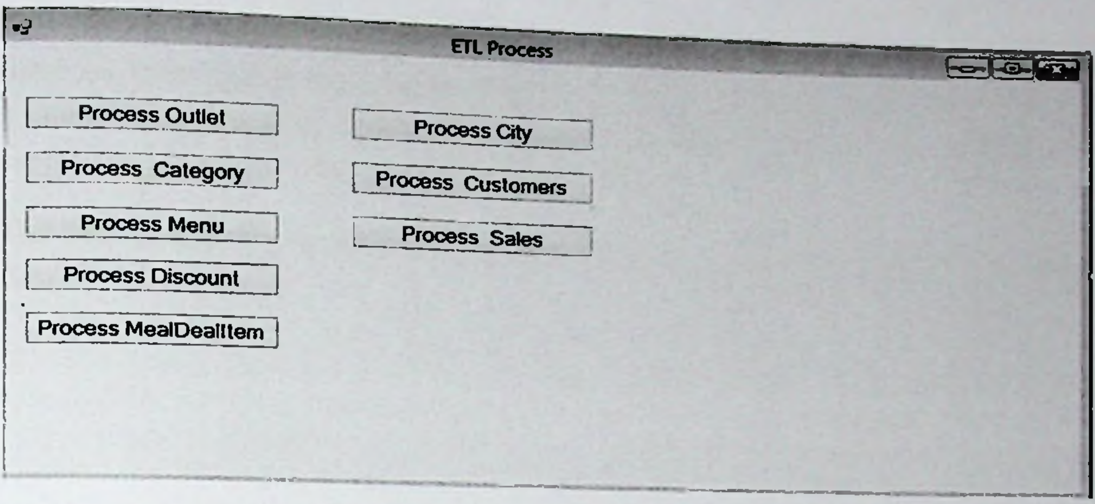
The bellowing code generates the results of quantity prediction with time period for selected menu item.

```
public string GenerateSalesQtyPrediction()
{

    int minimumItemPrediction = Convert.ToInt16(minimumPrediction.Value);
    string menuName = cmbMenu.Text;

    string DMX1 = " SELECT Flattened (PredictTimeSeries([MenuSales_TS].[Qty], " + minimumItemPrediction + ")) as [Predict Qty] ";
    string DMX2 = " From [MenuSales_TS] WHERE [MenuSales_TS].[Menu Name] ='" + menuName + "'";

    string DMX = "";

    DMX = DMX1 + DMX2;
    return DMX;

}
```

NumericUpDown btnGridclear.minimumPrediction



9.6 User Interface development for ETL process




```
Public void RunPackage(string packagepath)
      {
          Microsoft.SqlServer.Dts.Runtime.Application app = new
Microsoft.SqlServer.Dts.Runtime.Application();
Package package = null;

//Load the SSIS Package which will be executed
          package = app.LoadPackage(packagepath, null);

          Microsoft.SqlServer.Dts.Runtime.DTSExecResult results =
package.Execute();
//Check the results for Failure and Success
if (results == Microsoft.SqlServer.Dts.Runtime.DTSExecResult.Failure)
          {
string err = "";
foreach (Microsoft.SqlServer.Dts.Runtime.DtsError local_DtsError in
package.Errors)
              {
string error = local_DtsError.Description.ToString();
                  err = err + error;
              }
              lblMessage.Text = err;
              lblMessage.ForeColor = Color.Red;
          }
if (results == Microsoft.SqlServer.Dts.Runtime.DTSExecResult.Success)
          {
string message = "Package Executed Successfully....";
              lblMessage.Text = message;
              lblMessage.ForeColor = Color.Green;
          }

      }
```

9.7 Summary

This chapter explained the implementing part of migrating data into data warehouse database, creating data cube and building mining model for finding associate items on a sale and predicting sales quantity. There is a user interface to connect to analyze server and get predict results and execute the SSIS package.

The Next chapter discusses on testing, there we going to test predict result with actual data

Chapter 10

Testing

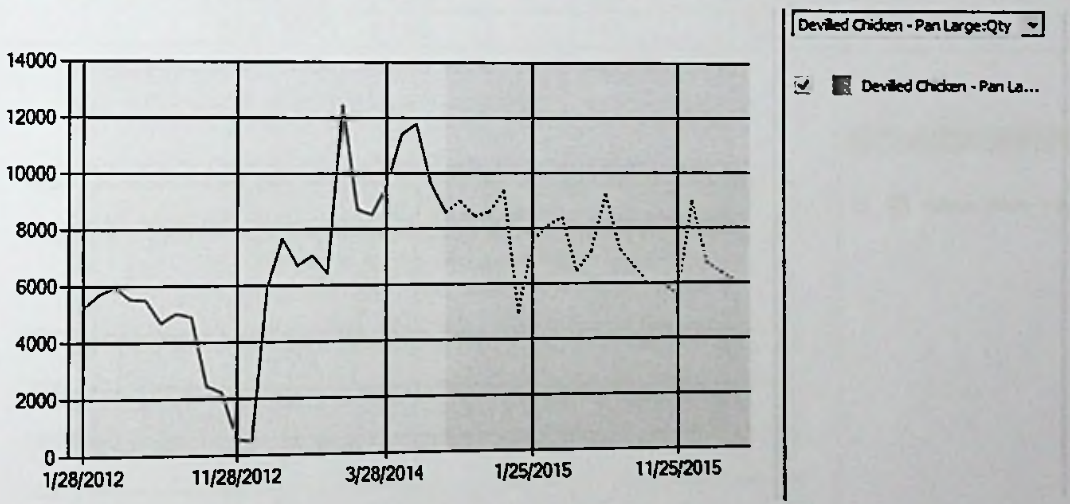
10.1 Introduction

The earlier chapter elaborates about implementation details of the project. The implementations have creating SSIS package, Develop SSAS project, Creating Mining models and user interface for see the predict results and process ETL.

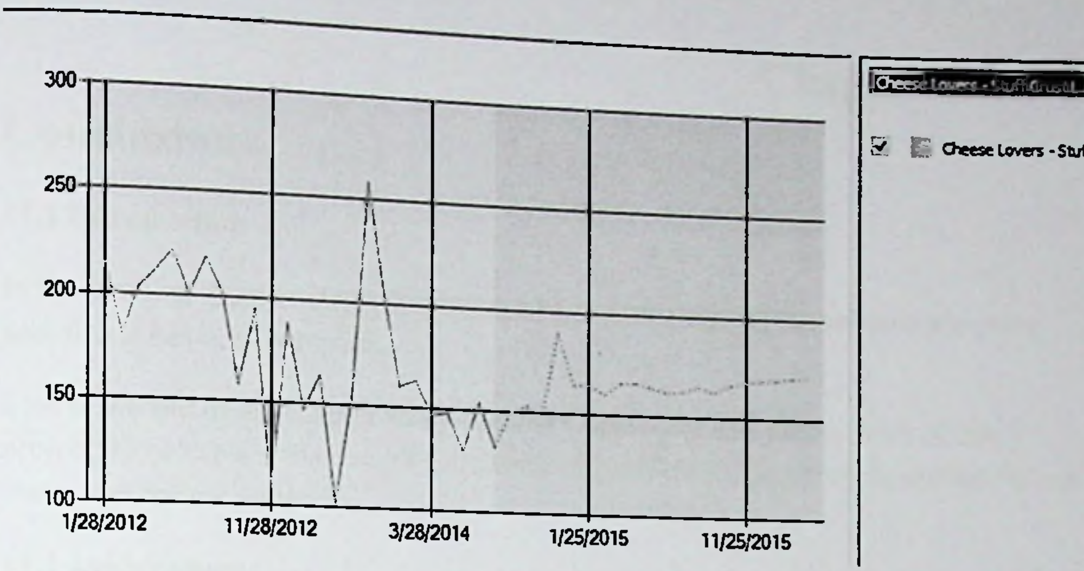
This chapter spells out on testing results of mining, through the actual sales value. From via we can identify the accuracy of algorithm and its results for future use. The mining app used for get the predict result and actual values get from current operational database

10.2 Testing Sales value with predicted values

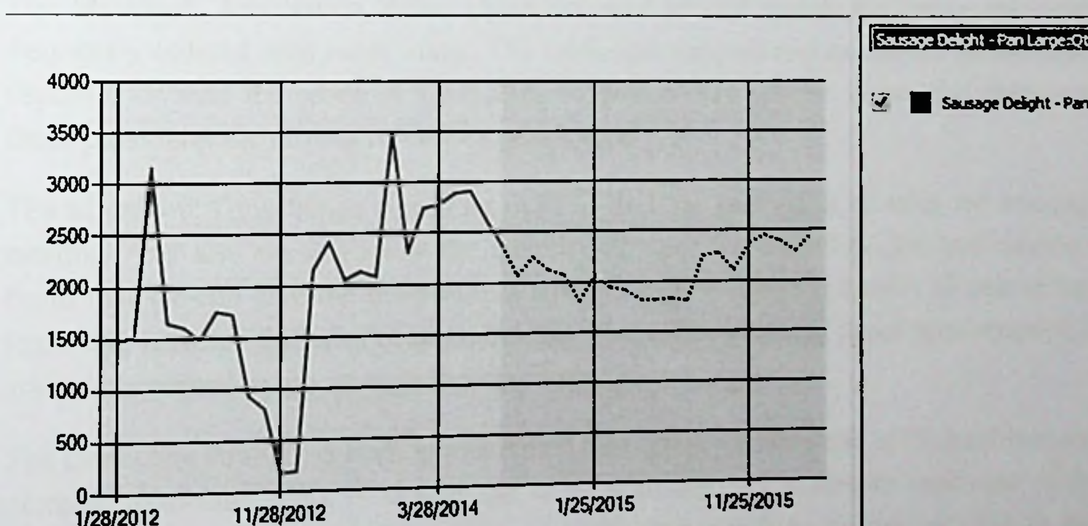
The predicted results are getting tested with actual sales values. There is a less differentiate between actual sales with comparing with predicted sales. Error Rate are mostly around +-10 % between actual vs. predicted values.



Menu	Month	Predicted Qty	Actual Qty	Error Rate(A-P)/A*100
Devilled Chicken-Pan Large	Aug/2014	9002	8503	5.86
	Sep/2014	8451	8151	3.68
	Oct/2014	8611	8215	4.82
	Nov/2014	9360	8567	9.25
	Dec/2014	4974	5125	2.94
	Jan/2015	7582	8133	6.77
	Feb/2015	8099	7895	2.58
	Mar/2015	8406	8894	5.48



Menu	Month	Predicted Qty	Actual Qty	Error Rate(A-P)/A*100
Cheese Lovers - Stuff Crust Large	Aug/2014	149	178	16.29
	Sep/2014	152	181	16.02
	Oct/2014	149	136	9.55
	Nov/2014	189	157	20.38
	Dec/2014	164	195	15.89
	Jan/2015	164	128	28.12
	Feb/2015	159	153	3.92
	Mar/2015	165	161	2.48



Menu	Month	Predicted Qty	Actual Qty	Error Rate(A-P)/A*100
Sausage Delight - Pan Large	Aug/2014	2066	2156	4.17
	Sep/2014	2249	1985	13.29
	Oct/2014	2120	1869	13.42
	Nov/2014	2071	2016	2.72
	Dec/2014	1803	1985	9.16
	Jan/2015	2065	1984	4.08
	Feb/2015	1940	2564	24.33
	Mar/2015	1912	1984	3..62

Table 10.1 Test Results

Chapter 11
Conclusion

11.1 Introduction

In the previous chapter it has discussed, how the objective of the solution are make sure that it has been achieved.

This is the end of the chapter and presented Conclusion and further work of this project. Here explain the overall achievement, problem encountered, limitation, lesson learnt and further work

11.2 Achievements

The solution of this project has been successfully implemented with data warehouse technology and mining data set. This solution improved the easy access of data accessing and look into deep on when making promotional strategies. The association rules are most valuable output of the solution. This is help to reach the customer attraction and retention of the organization business goal in continuous format. And Time series help to identify periodic sales of last years and most successful to get predicted value with minimum less differentiate.

The Microsoft Association Rules algorithm used to find which the items are most frequently ordered with other items. The minimum support and minimum probability variables defined the scope of the mining rules and item set. So as per the changing those parameter the mining rules may get change.

The Microsoft Time Series algorithm used to find the prediction of sales for coming months. And also we can know the quantity of sales for each item for late months. From that we can plan the promotional strategies for increase the sales of unmoving items and increase the sales of more moving items. The seasonal promotion strategies also implemented by go through the time series

The marketing strategy is been implemented through predicted sales with combination of most associate items is ensure the most valuable promotion to customer with increase of sales quantity. The pricing strategy also needs to be revised and to be checked how the predicted sales are more differentiate with actual sales quantity.

11.3 Problems encountered

The ETL process gets more time to complete the data load into data warehouse. The transactional data get loaded for one week in a break. And also less knowledge of the SSIS tool make very hard to configure data flow task without the error. So there is need to study the integration services for some period and re do the process again and again, to run the ETL process successfully with minimum errors.

As a same when doing mining also, in correct format of dataset kill the process time and need to wait long hours.

Making all databases in one machine is not possible according to capacity of the current database and can't get processing power as much as possible.

When moved the databases the SQL instance need to be same, otherwise need to reconfigure each SSIS package.

11.4 Limitations

The data are not much including of customer demography data. The Customer wise clustering is not possible with those data. At least of Age and Income clustering is not achieved. So customer wise promotional strategy is not able to design.

And also the number of the items in order is mostly one/two. And there is no variety of items in an order. Most of the time Coca cola/Garlic bread currently available when make an order. So these two items are mostly visible items. So is not make any sense of identifying an insight of the data.

The complex query structure of the DMX and MDX query is needed to be understood well.

The Storage capacity, Limitation of RAM, Less CPU speed make decrease the processing capacity and kill more time on ETL and mining.

11.5 Lesson Learnt

We need to choose the right algorithm and required data set for doing the mining. Otherwise is kill the process time and end result. There are more algorithms and we are try it them. But convert that end result into strategy is more irrelevant.

On mining, the data set must to be arranged or restructured according to requirement of the algorithm input. Otherwise is create in correct rules or no rues and also wrong output.

11.6 Further works

The sales need to be checked after the promotional strategy implemented and how is the predicted sales vary with those values. And also need to identify customer order frequency within a period and is more help to retain the customer.

And also there is a need to develop customized dashboard from available data warehouse, now is able to make reports from Power BI desktop software.

The customer wise menu prediction can be applied to online order app. The prediction may be happened after users enter his/her telephone number in the screen. And same time after selecting of one item, the menu prediction can apply by using of the association rules.

Reference

[1] Daniel J.Power, "Decision Support Systems" , University of Northern Iowa and DSSResources.com, USA(http://www.irma-international.org/viewtitle/14330/)

[2] S. Chaudhuri et al., "Database Technology for Decision Support Systems", IEEE Computer Society, Dec.2001,pp.48-55.

[3]http://www.hospitalityupgrade.com/_magazine/magazine_Detail-ID-522-Mining Restaurant- Data:- Know-your-customer.asp

[4] D.F.Rateb, 2013, "Database and Data Mining for Coffee Shops in Egypt" IBIMA Business review , Vol.2013,Article ID 94244 (www.ibimapublishing.com-journals- IBIMBR-2013-942449-942446.pdf)

[5] Lon-Mu Liu et al, Data mining on time series"Computational Statistics & Data analysis,37(2001)455-476 (http://stat.rutgers.edu/home/rongchen/publications/01CSDA_timeseries-datamining.pdf)

[6] P.Paulraj, 2001,"Data warehousing Fundamentals" ISBN:0-471-22162-7

[7] S.Chaudhuri, "An Overview of data Warehousing and OLAP Technology" (Appears in ACM igmodRecord,March 1997).

[8] 'Data Warehouse Design Considerations", SQL Server 2000, Microsoft Corporation (http://technet.microsoft.com/en-us/library/aa902672(v=sql.80).aspx)

[9]http://www.zdnet.com/gartner-releases-2013-data-warehouse-magic-quadrant-7000010796/ (database vendors)

[10] M.Velicanu. "Buiding a Data warehouse step by step", InformaticaEconomica. nr.2(42) 2007(http://revistaie.ase.ro/content/42/velicanu.pdf)

[11] http://www.databasejournal.com/sqletc/article.php/3888696/Data-Warehousing-Architecture---Designing-the-Data-Staging-Area.htm

[12] http://www.bidwbooks.com/what-is-datawarehouse/

[13] Brain knight, et al."Microsoft SQL server 2012 Integration Services ",ISBN:978-81-265-3561-3,

[14] Jamie MacLennan, et al. "Data Mining with SQL Server 2008 ",ISBN:978-81-265-1918-7,

[15] Margaret Rouse, Feb.2011, association rules (in data mining) http://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining

[16] http://www-users.cs.umn.edu/~desikan/research/dataminingoverview.html

[17] https://www.nchannel.com/blog/how-to-perform-sales-trend-analysis/

[18] http://www.slideshare.net/tharindurusira/time-series-prediction-algorithms-literature- Review.

[19] http://www.forecastingsolutions.com/arima.html

[20] Narendra Bute (2004), Important Features of .NET, http://www.codeproject.com/KB/vb/features_of_net.aspx

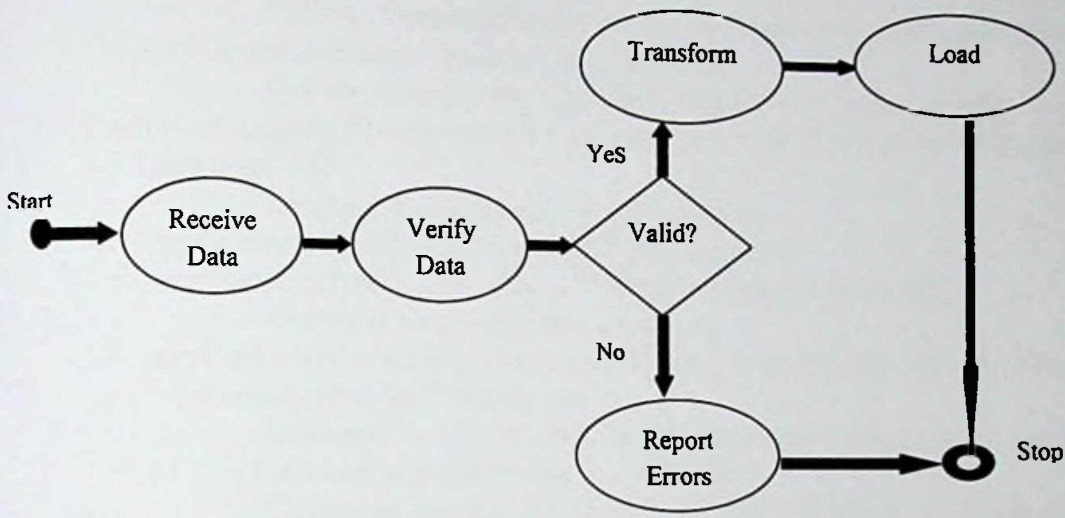
[21] Tiwebb Ltd (2007), what is .NET? , http://www.devtopics.com/what-is-net/

Appendix A – Matrix Design

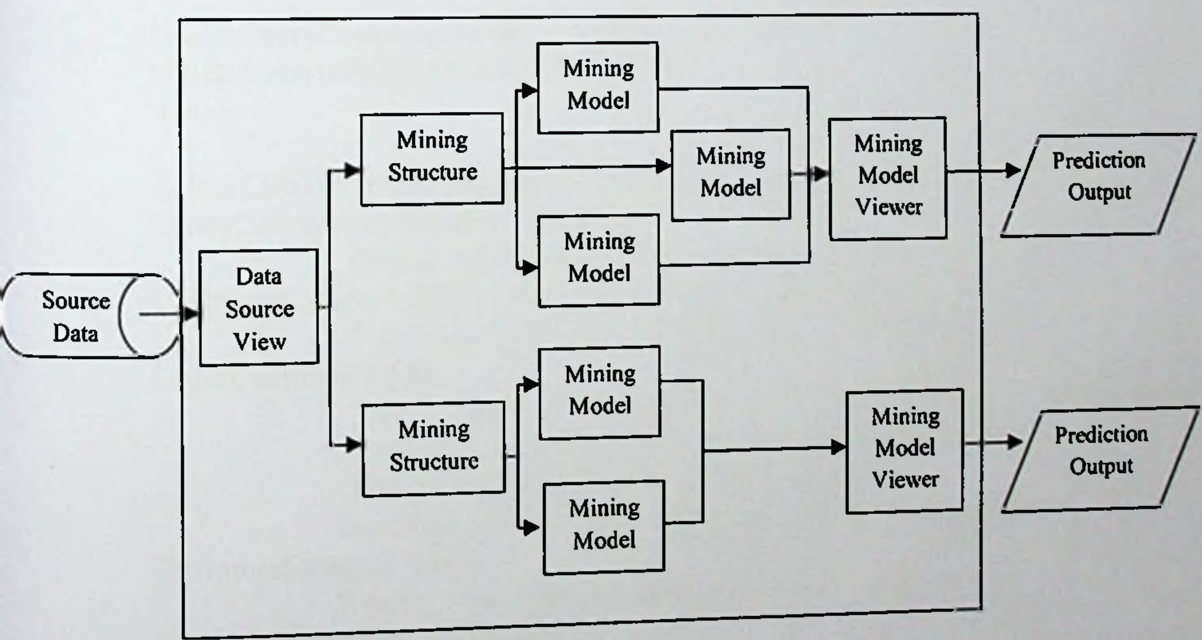
	Outlet	Men u	Cust omer	Disc ount	Date	Tim e	Orde r Type	Rider	Cit y	Cou ntry	Meal deal Cate gory	Meal Deal Item	Cust Cate gory	Cust Lanq age
Orde r	X	X		X	X	X	X						X	X
Deliv erers	X	X	X		X	X	X	X	X					
Onlin e Orde rs					X					X				
Meal Deal Tran	X				X						X	X		

Appendix B –Diagram

B.1 ETL State Diagram



B.2 Data mining components



Appendix C – Stored Proc in ETL

C.1 Extract Customer Data

```
SELECT DISTINCT
                DeliveryCustomerProfile.TelNo, DeliveryCustomerProfile.Title + '' +
DeliveryCustomerProfile.Customer AS Customer,
                DeliveryCustomerProfile.CustCategory AS CategoryCode,
CustomerCategory.Description AS Category, DeliveryCustomerProfile.CusLanguage
AS LanquageCode,
                LanguageRef.Description AS Lanquage,
DeliveryCustomerProfile.CustType AS CustomerType,
DeliveryCustomerProfile.CivilStatus, DeliveryCustomerProfile.DOB,
                DeliveryCustomerProfile.Company,
DeliveryCustomerProfile.Department, DeliveryCustomerProfile.SpouseDOB,
DeliveryCustomerProfile.WeddingAniv,
                DeliveryCustomerProfile.EntryDate, DeliveryCustomerProfile.Active
FROM        DeliveryCustomerProfile INNER JOIN
                CustomerCategory ON DeliveryCustomerProfile.CustCategory =
CustomerCategory.Code INNER JOIN
                LanguageRef ON DeliveryCustomerProfile.CusLanguage =
LanguageRef.Code
                                                Where
len(DeliveryCustomerProfile.Customer)>=4 and
len(DeliveryCustomerProfile.TelNo)<=9
 Union all

 select DIstinct  OrderCustomer.TelNo,OrderCustomer.Title + '' +
OrderCustomer.Customer as Customer,
                 OrderCustomer.CustCategory as CategoryCode,
CustomerCategory.Description AS Category,'' as LanquageCode,
                 '' as Lanquage, '' as CustomerType, '' as CivilStatus,
OrderCustomer.DOB,
                 '' as Company, '' as Department, '' as SpouseDOB, '' as WeddingAniv,
                 OrderCustomer.SysDate as EntryDate,'true' as Active
                 FROM OrderCustomer INNER JOIN
                CustomerCategory ON OrderCustomer.CustCategory =
CustomerCategory.Code
                Where  len(OrderCustomer.Customer)>=4 and
len(OrderCustomer.TelNo)<=9
```

C.2 Extract Sales Data

```
set dateformat dmy select  BillHeader.OrderNo,
CONVERT (DATE,BillHeader.OrderDate) as orderdate,
 CONVERT(VARCHAR(8),BillHeader.OrderTime,108) as OrderTime,
 BillHeader.OrderType,
```

```
Convert(varchar(3),BillTrans.Category) as Category, BillTrans.MainMenuCode,
BillTrans.SubMenuCode, BillTrans.Qty, BillTrans.Rate,
BillHeader.Outlet, BillHeader.Pax,
CASE WHEN BillHeader.DisCode is null THEN 'N/A' WHEN
BillHeader.DisCode=" THEN 'N/A' ELSE BillHeader.DisCode END DisCode,
CASE WHEN  BillHeader.TelNo IS null then'N/A' WHEN  BillHeader.TelNo=" then
'N/A' ELSE BillHeader.TelNo end TelNo  ,

 BillTrans.ItemType,
              BillHeader.TableNo,

DATEDIFF(N,CONVERT(VARCHAR(8),BillHeader.OrderTime,108),CONVERT(
VARCHAR(8),BillHeader.EndTime,108)) as TableTime,
 BillHeader.Sus_Date as CallCenterOrderNo
FROM        BillHeader INNER JOIN
        BillTrans ON BillHeader.OrderNo = BillTrans.OrderNo AND
BillHeader.OrderDate = BillTrans.OrderDate
        AND BillHeader.Outlet = BillTrans.Outlet
        Where BillHeader.OrderDate>=@Startdate and
BillHeader.OrderDate<=@EndDate
```

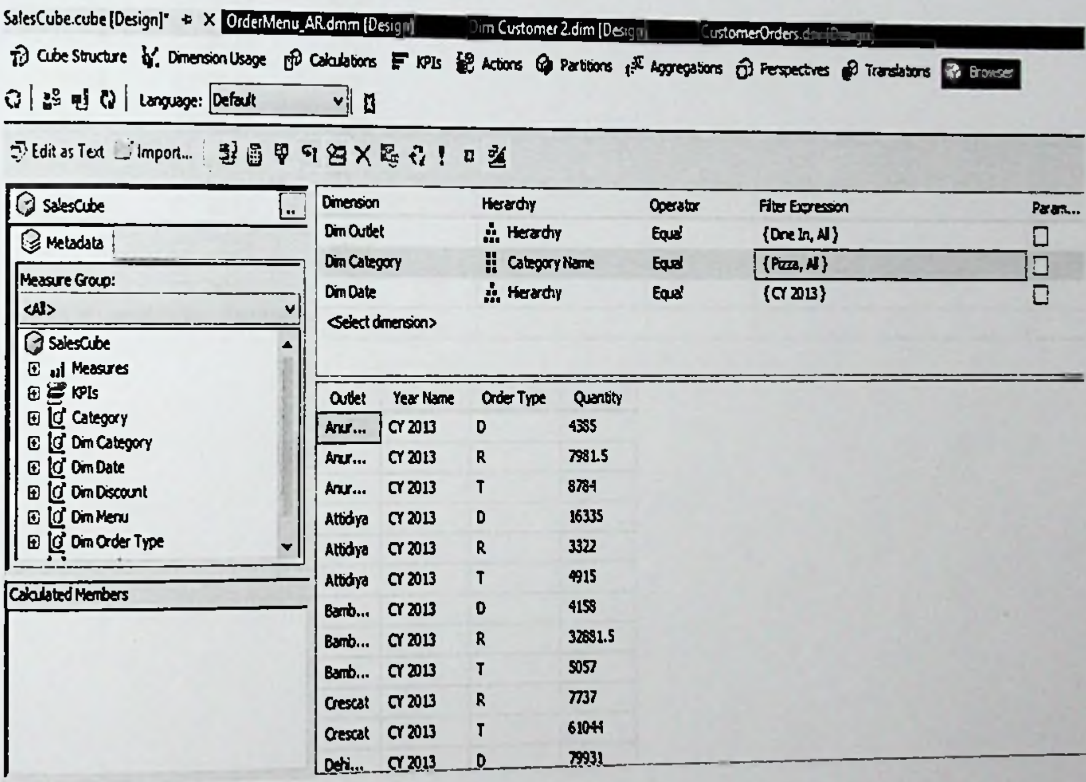
C.3 Extract Delivery data

```
set dateformat dmy  Select HistoryDeliveryOrderHeader.OrderNo,
HistoryDeliveryOrderHeader.MasterTelNo,
              HistoryDeliveryOrderHeader.Outlet,
              HistoryDeliveryOrderHeader.SubTotal,
HistoryDeliveryOrderHeader.PickupOrder, HistoryDeliveryOrderHeader.OpenTime,
HistoryDeliveryOrderHeader.ChangeOrder,
              HistoryDeliveryOrderHeader.Rider, CONVERT(VARCHAR(8),
HistoryDeliveryOrderHeader.DeliveryTime, 108) AS DeliveryTime,
CONVERT(VARCHAR(8),
              HistoryDeliveryOrderHeader.DispatchTime, 108) AS DispatchTime,
CONVERT(VARCHAR(8), HistoryDeliveryOrderHeader.DeliveredTime, 108) AS
DeliveredTime,
              HistoryDeliveryOrderHeader.HotDeal,
HistoryDeliveryOrderHeader.Department, HistoryDeliveryOrderHeader.Company,
CONVERT(DATE,
              HistoryDeliveryOrderHeader.OrderDate) AS OrderDate,
CONVERT(VARCHAR(8), HistoryDeliveryOrderHeader.OrderTime, 108) AS
OrderTime, StreetMaster.City
FROM        HistoryDeliveryOrderHeader INNER JOIN
              StreetMaster ON HistoryDeliveryOrderHeader.Street =
StreetMaster.Code INNER JOIN
              CityMaster ON StreetMaster.City = CityMaster.Code
```

Appendix D – MDX Queries

D.1 Following MDX query will apply for viewing Sales of quantity of year 2013 by order type in all outlets

SELECT NON EMPTY { [Measures].[Quantity] } ON COLUMNS, NON EMPTY { ([Dim Outlet].[Outlet].[Outlet].ALLMEMBERS * [Dim Date].[Year Name].[Year Name].ALLMEMBERS * [Dim Order Type].[Order Type].[Order Type].ALLMEMBERS) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM (SELECT ({ [Dim Date].[Hierarchy].[Year Name].&[CY 2013] }) ON COLUMNS FROM [SalesCube]) WHERE ([Dim Date].[Hierarchy].[Year Name].&[CY 2013]) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS

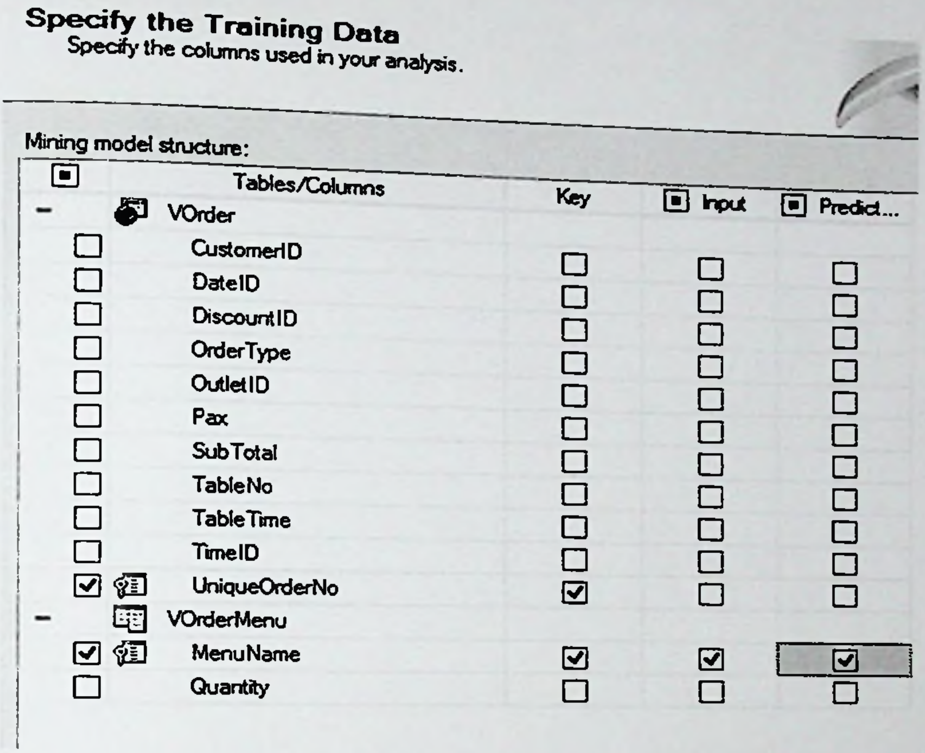


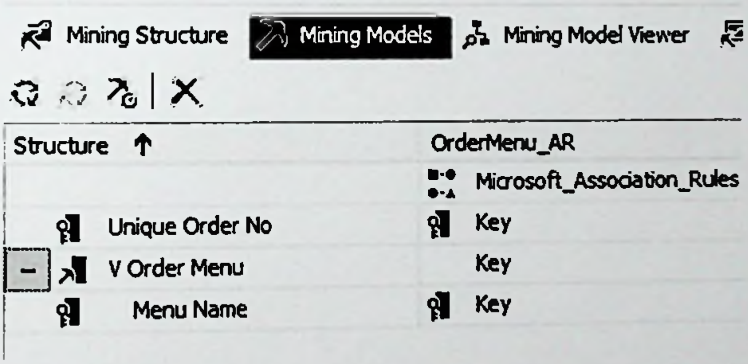
D.2 Following MDX query will apply for viewing delivery count by city on year 2013

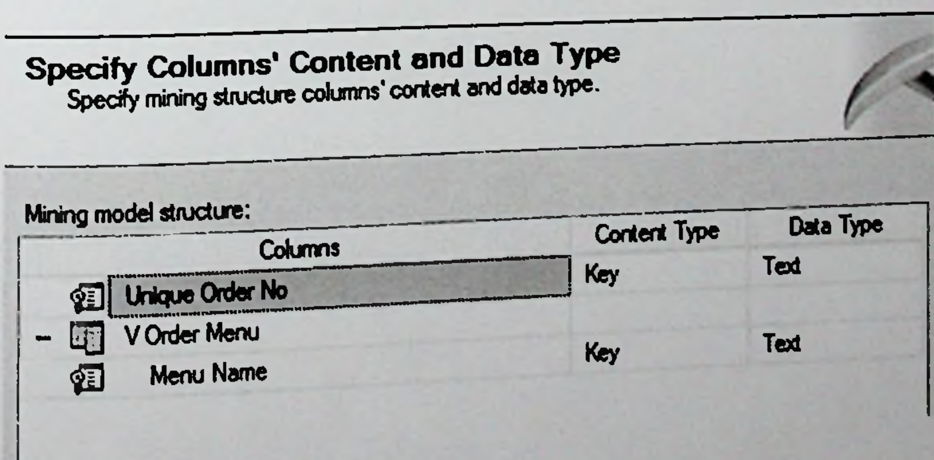
SELECT NON EMPTY { [Measures].[Fact Deliveries Count] } ON COLUMNS, NON EMPTY { ([Dim City].[City Name].[City Name].ALLMEMBERS * [Dim Date 1].[Year Name].[Year Name].ALLMEMBERS) } DIMENSION PROPERTIES MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM (SELECT ({ [Dim Date 1].[Year Name].&[CY 2013] }) ON COLUMNS FROM [DeliveryCube]) CELL PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE, FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS

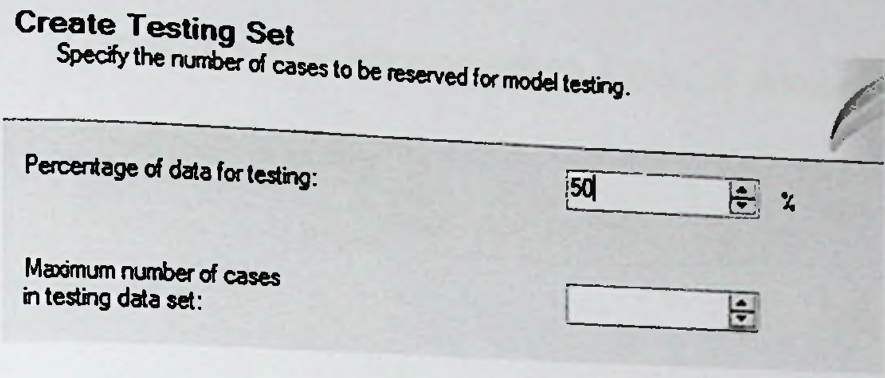
City Name	Year Name	Fact Deliveries Count
ANGODA	CY 2013	97
ANPITIYA	CY 2013	372
ANURAD...	CY 2013	3224
ARANGALA	CY 2013	2112
ARUPPOLA	CY 2013	107
ASGIRIYA	CY 2013	1170
ATHURU...	CY 2013	98
ATTIDIYA	CY 2013	213
BADAHEL...	CY 2013	1051
BAHIRAV...	CY 2013	1521

Appendix E – Implementing Steps of Association Rule

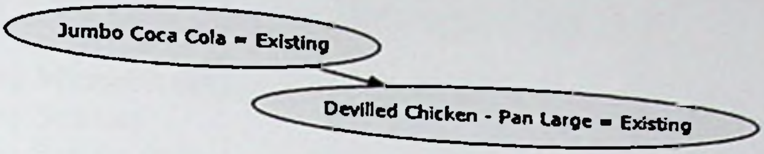


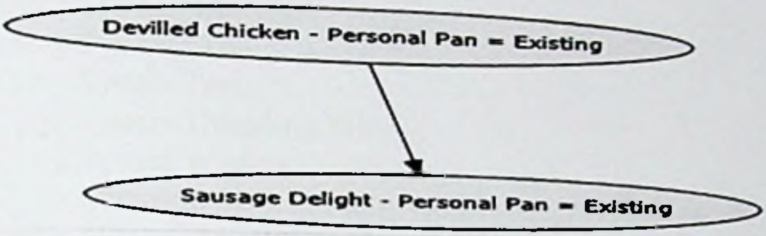


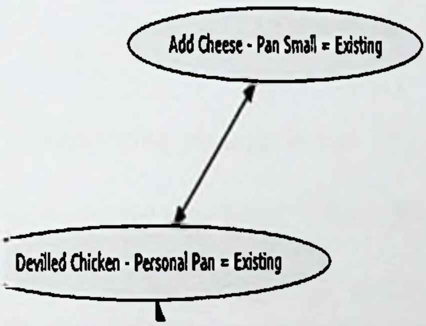


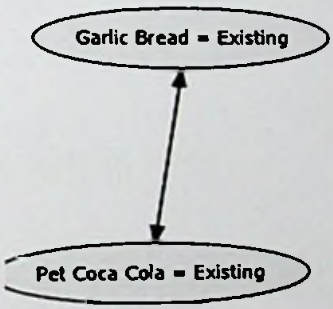

Create Testing Set
Specify the number of cases to be reserved for model testing.
Percentage of data for testing:
50
%
Maximum number of cases
in testing data set:

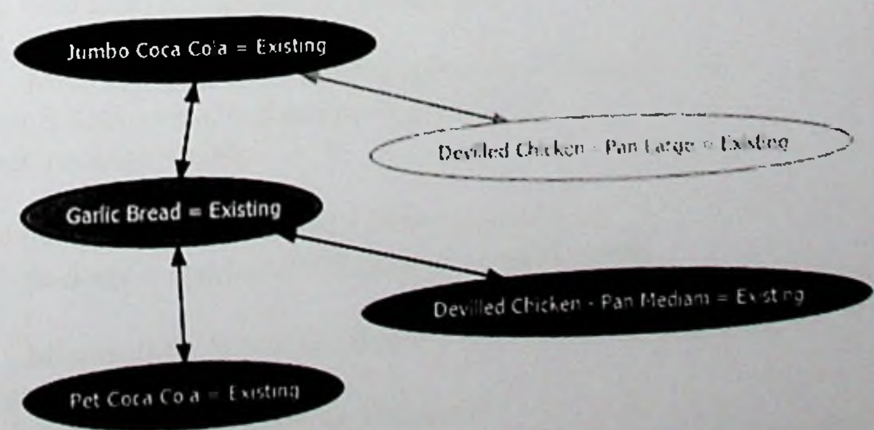
Appendix F – Strongest to Lowest Associate Items











Appendix G – Coding for ETL package execution through Win App

```
using Microsoft.SqlServer.Dts.Runtime;
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using Microsoft.SqlServer.Dts.Runtime;

namespace WindowsFormsApplication1
{
publicpartialclassForm1 : Form
   {
public Form1()
      {
         InitializeComponent();
      }

privatestring packagePath = "";

privatevoid btnOutlet_Click(object sender, EventArgs e)
      {
         packagePath = @"D:\My
Project\SSIS\IntergrationServices\IntegrationService\IntegrationService\Dim_Outlet.
dtsx";
         RunPackage(packagePath);
      }

Public void RunPackage(string packagepath)
      {
         Microsoft.SqlServer.Dts.Runtime.Application app = new
Microsoft.SqlServer.Dts.Runtime.Application();
Package package = null;

//Load the SSIS Package which will be executed
         package = app.LoadPackage(packagepath, null);

         Microsoft.SqlServer.Dts.Runtime.DTSExecResult results =
package.Execute();
//Check the results for Failure and Success
if (results == Microsoft.SqlServer.Dts.Runtime.DTSExecResult.Failure)
         {
string err = "";
```

```
foreach (Microsoft.SqlServer.Dts.Runtime.DtsError local_DtsError in
package.Errors)
            {
string error = local_DtsError.Description.ToString();
                err = err + error;
            }
            lblMessage.Text = err;
            lblMessage.ForeColor = Color.Red;
        }
if (results == Microsoft.SqlServer.Dts.Runtime.DTSExecResult.Success)
        {
string message = "Package Executed Successfully....";
            lblMessage.Text = message;
            lblMessage.ForeColor = Color.Green;
        }

    }

privatevoid btnUploadCategory_Click(object sender, EventArgs e)
    {
        packagePath = @"D:\My
Project\SSIS\IntergrationServices\IntegrationService\IntegrationService\Dim_Categor
y.dtsx";
        RunPackage(packagePath);
    }

privatevoid btnProcessMenu_Click(object sender, EventArgs e)
    {
        packagePath = @"D:\My
Project\SSIS\IntergrationServices\IntegrationService\IntegrationService\Dim_Menu.d
tsx";
        RunPackage(packagePath);
    }

privatevoid btnProcessDiscount_Click(object sender, EventArgs e)
    {
        packagePath = @"D:\My
Project\SSIS\IntergrationServices\IntegrationService\IntegrationService\Dim_Discou
nt.dtsx";
        RunPackage(packagePath);
    }

privatevoid btnProcessMealDealItem_Click(object sender, EventArgs e)
    {
        packagePath = @"D:\My
Project\SSIS\IntergrationServices\IntegrationService\IntegrationService\Dim_MealDe
alItem.dtsx";
        RunPackage(packagePath);
    }
```

Appendix H – Coding of Mining user interface

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;
using System;
using System.Data.SqlClient;
using System.Data.OleDb;
using System.Runtime.Remoting.Channels;
using System.Windows.Forms.VisualStyles;
using ADOMD;
using Microsoft.AnalysisServices.AdomdClient;
using System.Linq;



namespace MenuPromotion
{
   public partial class btnGridclear : Form
   {

      AdomdConnection _adomdConnectioncon;
      string connectionString = "Provider=MSOLAP;Integrated Security=SSPI;Persist
Security Info=False;DataSource=SHAMARAN-LENOVO;Initial
Catalog=Pizzahut_Sales;";

      public btnGridclear()
      {
         InitializeComponent();
         RefreshData();
      }



      public void RefreshData()
      {
         DataRow dr;


         SqlConnection con = new SqlConnection(@"Data Source=SHAMARAN-
LENOVO;Initial Catalog=Pizza-DatawareHouse;Integrated Security=True");
         con.Open();
```

```
            SqlCommand cmd = new SqlCommand("select MenuId,MenuName from
Dim_menu where isactive=1 order by MenuName", con);
            SqlDataAdapter sda = new SqlDataAdapter(cmd);
            DataTable dt = new DataTable();
            sda.Fill(dt);

            dr = dt.NewRow();
            dr.ItemArray = new object[] { 0, "--Select Menu--" };
            dt.Rows.InsertAt(dr, 0);

            cmbMenu.ValueMember = "MenuId";

            cmbMenu.DisplayMember = "MenuName";
            cmbMenu.DataSource = dt;

            con.Close();
        }

        private void btnMining_Click(object sender, System.EventArgs e)
        {
            if (minItemSet.Value == 0)
            {
                MessageBox.Show("Please enter the valid value", "Message",
MessageBoxButtons.OK);

            }
            else if (cmbMenu.Text == "--Select Menu--")
            {
                MessageBox.Show("Please select the menu item", "Message",
MessageBoxButtons.OK);
            }
            else
            {
                ExecuteAndFetch(GenerateMenuPrediction(), "PredictMenu");
            }



        }

        public string GenerateMenuPrediction()
        {
            string menuName = cmbMenu.Text;
            int minimumItemSet = Convert.ToInt16(minItemSet.Value);

            string DMX1 = "SELECT flattened PredictAssociation([OrderMenu_AR].[V
Order Menu],INCLUDE_STATISTICS, " + minimumItemSet + ")  as Result ";
            string DMX2 = " From[OrderMenu_AR] NATURAL PREDICTION JOIN
(SELECT (SELECT '" + menuName + "' AS [Menu Name]) AS [V Order Menu]) AS
t";
```

```
        string DMX = "";

        DMX = DMX1  + DMX2;
        return DMX;

    }

    public string GenerateCustometrMenuPrediction()
    {
        string telno = txtTelNo.Text;
        string DMX = "";


        int minimumItemSet = Convert.ToInt16(minItemSet.Value);

        string DMX1 = "SELECT flattened
PredictAssociation([CustomerOrderMenu_AR].[V Customer Order],
INCLUDE_STATISTICS, " + minimumItemSet + ") ";
        string DMX2 = " From [CustomerOrderMenu_AR] PREDICTION JOIN
SHAPE {OPENQUERY([Pizza_DatawareHouse],'SELECT[MasterTelNo], ";
        string DMX3 = " [CustomerID] FROM [dbo].[Dim_Customer] ORDER BY
[CustomerID]')} APPEND ({OPENQUERY([Pizza_DatawareHouse], ";
        string DMX4 = "'SELECT[MenuName],[CustomerID] FROM
[dbo].[VCustomerOrder] ORDER BY [CustomerID]')} RELATE[CustomerID] ";
        string DMX5 = " TO [CustomerID]) AS [VCustomerOrder] AS t ON
[CustomerOrderMenu_AR].[Master Tel No] = t.[MasterTelNo] AND ";
        string DMX6 = " [CustomerOrderMenu_AR].[V Customer Order].[Menu
Name] = t.[VCustomerOrder].[MenuName] WHERE ";
        string DMX7 = " t.[MasterTelNo] ='" + telno + "'";

        DMX = DMX1 + DMX2 + DMX3 + DMX4 + DMX5 + DMX6 + DMX7;

        return DMX;

    }

    public bool ExecuteAndFetch(string strCommand, string gridType)
    {

        using (var connection = new OleDbConnection(connectionString))
        {

            connection.Open();

            using (var command = new OleDbCommand(strCommand, connection))
            {

                command.CommandTimeout = 10000;
                OleDbDataAdapter sqlDataAdap = new OleDbDataAdapter(command);
```

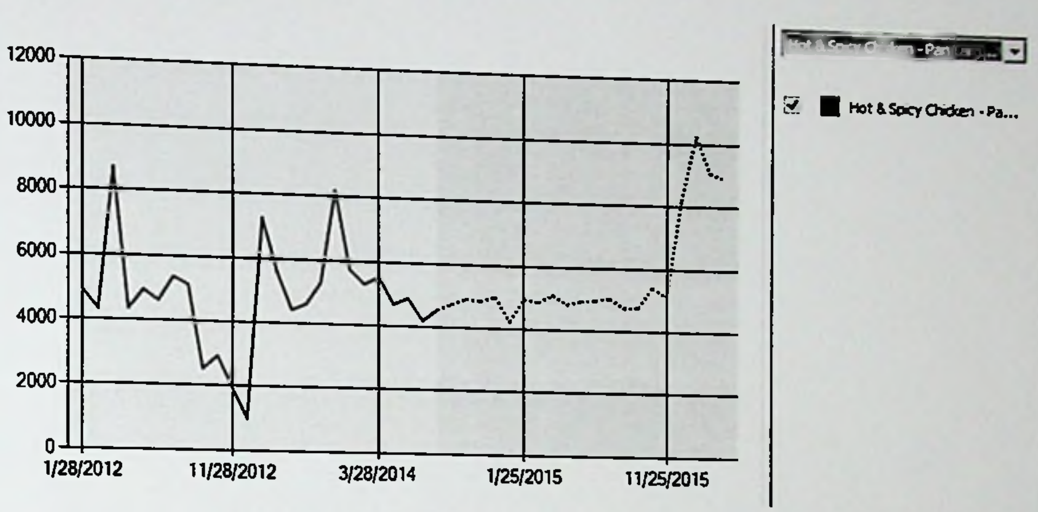
```
            DataTable dtRecord = new DataTable();
            sqlDataAdap.Fill(dtRecord);
            if (gridType=="PredictMenu")
            {
                dgvMenu.DataSource = dtRecord;
            }
            else if (gridType=="PredictCustomerMenu")
            {
                dgvPredictByTelNo.DataSource = dtRecord;
            }
            else if (gridType == "PredictSales")
            {
                dgvPredictSalesQty.DataSource = dtRecord;
            }
            else if (gridType == "PredictSalesARTXP")
            {
                dgvPredictSalesQtyARTXP.DataSource = dtRecord;
            }

            else if (gridType == "PredictSalesARIMA")
            {
                dgvPredictSalesQtyARIMA.DataSource = dtRecord;
            }

          }

        }
        return true;
      }

      private void btnCustomerPersonalisedMenu_Click(object sender,
System.EventArgs e)
      {
        string telno = "";
        telno = txtTelNo.Text.ToString().Trim();
        if (minItemSet.Value == 0)
        {
          MessageBox.Show("Please enter the valid value", "Message",
MessageBoxButtons.OK);

        }
        else if (telno.Length== 0)
        {
          MessageBox.Show("Please enter a valid phone number", "Message",
MessageBoxButtons.OK);
        }
        else
        {
```

```
            ExecuteAndFetch(GenerateCustometrMenuPrediction(),
"PredictCustomerMenu");
        }

    }

    private void button1_Click(object sender, System.EventArgs e)
    {
        dgvMenu.DataSource = null;
        dgvPredictByTelNo.DataSource = null;
        dgvPredictSalesQty.DataSource = null;
        dgvPredictSalesQtyARTXP.DataSource = null;
        dgvPredictSalesQtyARIMA.DataSource = null;
    }

    private void btnPredictableSalesQty_Click(object sender, System.EventArgs e)
    {
        if (minimumPrediction.Value == 0)
        {
            MessageBox.Show("Please enter the valid value", "Message",
MessageBoxButtons.OK);

        }
        else if (cmbMenu.Text == "--Select Menu--")
        {
            MessageBox.Show("Please select the menu item", "Message",
MessageBoxButtons.OK);
        }
        else
        {
            ExecuteAndFetch(GenerateSalesQtyPrediction(), "PredictSales");
        }

    }


    public string GenerateSalesQtyPrediction()
    {

        int minimumItemPrediction = Convert.ToInt16(minimumPrediction.Value);
        string menuName = cmbMenu.Text;

        string DMX = "";

        string DMX1 = " SELECT Flattened
(PredictTimeSeries([MenuSales1_TS].[Qty], " + minimumItemPrediction +
                    ")) as [Predict Qty] ";
        string DMX2 = " From [MenuSales1_TS] WHERE [MenuSales1_TS].[Menu
Name] ='" + menuName + "'";
```

```
        DMX = DMX1 + DMX2;

        return DMX;

    }

    public string GenerateSalesQtyPredictionARTXP()
    {

        int minimumItemPrediction = Convert.ToInt16(minimumPrediction.Value);
        string menuName = cmbMenu.Text;

        string DMX = "";

        string DMX1 = " SELECT Flattened
(PredictTimeSeries([MenuSales1_ARTXP].[Qty], " + minimumItemPrediction +
                "")) as [Predict Qty] ";
        string DMX2 = " From [MenuSales1_ARTXP] WHERE
[MenuSales1_ARTXP].[Menu Name] ='" + menuName + "'";

        DMX = DMX1 + DMX2;

        return DMX;

    }

    public string GenerateSalesQtyPredictionARIMA()
    {

        int minimumItemPrediction = Convert.ToInt16(minimumPrediction.Value);
        string menuName = cmbMenu.Text;

        string DMX = "";

        string DMX1 = " SELECT Flattened
(PredictTimeSeries([MenuSales1_ARIMA].[Qty], " + minimumItemPrediction +
                "")) as [Predict Qty] ";
        string DMX2 = " From [MenuSales1_ARIMA] WHERE
[MenuSales1_ARIMA].[Menu Name] ='" + menuName + "'";

        DMX = DMX1 + DMX2;

        return DMX;

    }

    private void btnARTXP_Click(object sender, System.EventArgs e)
    {
        if (minimumPrediction.Value == 0)
        {
```

```
            MessageBox.Show("Please enter the valid value", "Message",
MessageBoxButtons.OK);

        }
        else if (cmbMenu.Text == "--Select Menu--")
        {
            MessageBox.Show("Please select the menu item", "Message",
MessageBoxButtons.OK);
        }
        else
        {
            ExecuteAndFetch(GenerateSalesQtyPredictionARTXP(),
"PredictSalesARTXP");
        }

    }

    private void btnARIMA_Click(object sender, System.EventArgs e)
    {
        if (minimumPrediction.Value == 0)
        {
            MessageBox.Show("Please enter the valid value", "Message",
MessageBoxButtons.OK);

        }
        else if (cmbMenu.Text == "--Select Menu--")
        {
            MessageBox.Show("Please select the menu item", "Message",
MessageBoxButtons.OK);
        }
        else
        {
            ExecuteAndFetch(GenerateSalesQtyPredictionARIMA(),
"PredictSalesARIMA");
        }
    }
  }
}
```

Appendix I – Test Results



Menu	Month	Predicted Qty	Actual Qty	Error Rate(A-P)/A*100
Hot & Spicy Chicken - Pan Large	Aug/2014	4730	5230	9.56
	Sep/2014	4908	4235	15.89
	Oct/2014	4852	4526	7.20
	Nov/2014	5003	5156	2.96
	Dec/2014	4228	4695	9.94
	Jan/2015	4950	4599	7.63
	Feb/2015	4860	5132	5.30
	Mar/2015	5097	4952	2..92