Sales Tracking System for CPG Giant - Business Intelligence System

** Vignesh K. Aiyer*
*** Sujaan Choksi*
**** Abishek Murali*

Abstract

***Abstract*-Business intelligence commonly known as BI may be referred to as a set of technologies, applications and practices for the collection, integration, analysis, and presentation of business information to support better business decision making. BI systems are basically data driven Decision Support Systems (DSS). BI can be used to support a wide range of business decisions ranging from operational decisions include product positioning or pricing to strategic business decisions including priorities, goals, and directions at the broadest level. When combined, external data (data derived from the market in which a company operates) and internal data (data from sources internal to the company such as financial and operations data) can provide a more complete picture, which creates 'intelligence' that cannot be derived by any singular set of data. Hence BI techniques are going to be used to analyze an insurance company's data and see how we are going to make the data mean something more. In this case it will be used to make decisions regarding formulation of new policies, termination of old policies, and analyzing the route of credit and debit of a company.**

***Index Terms* : Business Intelligence, Decision Support System, SQL Server, analytics**

I. INTRODUCTION

A system is being developed to gain insights into sales data like sales volume for products of a Global CPG major and to identify the location which made the lowest sales for improving marketing in that region. The company currently tracks all its data through its ERP solutions which serves the purpose of data gathering but does not allow analysis. This system focuses on analysis of the data generated from the ERP system of the company. Data is pulled from ERP to ODS through linked servers. From ODS, the data is brought to the staging phase where the data is segregated and cleaned. The staging contains all the data recovered from the ODS. The essential data which is needed for analysis and reporting is then extracted to a data mart or the warehouse [1]. The mart data is then used for cube construction which helps in faster and deeper data analysis. Cubes are multi-dimensional data sources which have dimensions and facts (also known as measures) as their basic constituents. The constructed cube can be used for analysis of data. This can be done by visualization and reporting tools which allow easily connecting the data and forming interactive dashboards. The system hence provides a better view of the data pulled.

A. SM@RT Brand management at CPG major

The CPG major manages an international brand portfolio across more than 160 countries, and needs to be able to respond fast to both competitive threats and new opportunities. The solution was SM@RT - an application that provides fast feedback on the competitive situation; globally, regionally, and in each market, by combining information on the consumer sales of Imperial brands versus competitor brands.

B. Delivering global insights into the duty free market

It constructs a single portal that would bring together

* *Electronics & Communication Engineering,* VIT University, Vellore, India. E-mail : vignesh_aiyer@yahoo.com
** *Electronics & Communication Engineering,* Manipal University, Manipal, India. E-mail : sujaanchoksi@gmail.com
*** *Electronics and Communication Engineering,* BMSCE, Bangalore, India. E-mail : abimur@gmail.com

all the relevant duty free sales, domestic markets data and traveler information. The unique challenges of a complex market, with disparate data sources of variable quality and the specific purchasing behavior of the duty free customer would require significant intervention and data alignment to achieve this holistic view.

C. Equipping sales teams with data to seal deals

A single, elegant business intelligence solution provides the information required by multiple pharmaceutical sales and marketing operations across Europe and aligns the organization behind the performance priorities of the European management team of Chiron.

II. REQUIREMENT SPECIFICATIONS

The project is for a Consumer Packed Goods (CPG) company that deals with Tobacco. The Electronic Point of Sale (EPOS) data is collected by Sabre. This data is then stored in a database. The existing system allows the users to directly build excel reports. The problem with the existing system is that the reports are dull and there is too much data to look into and decide; therefore, this project was needed.

The ultimate aim of the project is to create user friendly reports that emphasize on better visualization, thereby allowing the users to make better judgments. For example, consider a table containing sales data of a product A for cities under a particular country (Fig. 1.) and a geographical map showing the cities in different color where a particular color represents a sales range (Fig. 2). Though the first figure will have accurate data, the second one will have more graphical sort of representation which will enable easy interpretation.

Region	Area								
Central	Area 1	46016	46032	46041	46060	46142	46203	46226	46307
	Area 2	64804	67037	67114	67212	67501	67601	67801	67846
East	Area 3	01001	01007	01013	01027	01028	01040	01056	01060
	Area 4	03246	03301	03839	04005	04011	04038	04070	04072
South	Area 5	20151	20170	20190	22003	22015	22025	22102	22124
	Area 6	32004	32114	32127	32137	32168	32174	32216	32259
	Area 7	35401	35473	35630	36330	36608	38017	38109	38134
West	Area 8	59102	59405	59601	59715	59750	59801	59901	80004
	Area 9	85023	85204	85224	85234	85254	85281	85301	85323
	Area 10	89431	89502	89701	94015	94024	94025	94043	94044

Fig. 1. Tabular Data

Along with the visually enhanced data, the reports should be dynamic, should enable faster retrieval of data and the coloring schema should be in line with the company standards. For this purpose, we will create multidimensional cubes and our reports or dashboards will be based on these cubes.

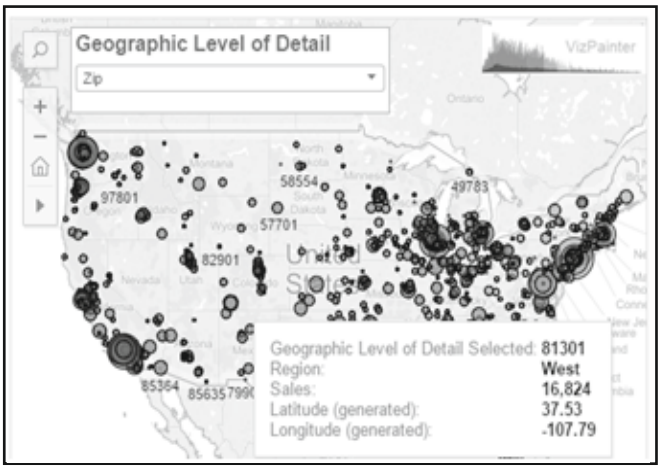

Fig. 2. Geographical Heat Map

III. COMPRESSIVE SENSING

All BI solutions start with analyzing the existing system or the proposed system in order to provide the client with leveraging solutions. The basic flow would be data extraction, loading and cleaning.

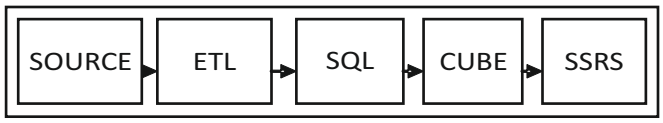

Fig. 3.Dataflow in project

The source of data is in excel files and in databases. The ETL process that will be triggered through a series of packages will load the data into the respective tables in the databases after which views will be created. The cube will take data from the views and provide us with multidimensional view of the data. The dashboards will be created using SSRS and will use the cube data to create charts.

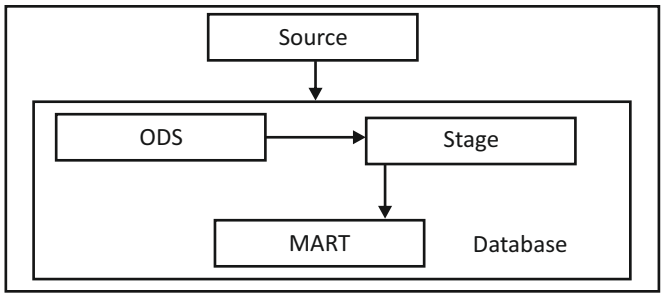

Fig. 4 Database Structure – CPG Major

A. Database

There are around three databases- ODS, Stage, and Mart. The source that is in the form of database is stored in the ODS database and then moved to the stage database. The other source that is stored in the excel files will be moved straight to the stage database. The mart database will be the final database. The data in stage will be moved on to the mart after cleaning. The error records will be moved on the error table in the mart database after which views will be created on all the tables to support model building. The ODS database is a repository. It holds external data from the file server and other external feeds. The database has a schema called repository. The staging database will be used to hold the external data from the files/repository. The database will consist of various staging tables which would serve as intermediary tables between repository/files and mart database [2]. These tables will be used to cleanse, transform, and manipulate different forms of data prior to loading the same into the mart database tables. This database also contains stored procedures which provide a mechanism to transform and load data from the staging tables to the fact and dimension tables. The mart database consists of fact and dimension tables and views that will be used to build the SSAS cube.

B. *SQI Server Integration Services (SSIS) Package*

The system will have two packages.

Load Dimension Data

This package would load all the following data from the staging tables into the mart tables

1. Product: The package should transfer all the data from the product table in ODS into the stage table. The data is then loaded into the mart table. The loading into the mart tables is handled by the stored procedure.
2. The generic flow of the package would be as follows:
   i. Get the respective connection settings from the settings table.
   ii. Log package start
   iii. Check if any other package is running. If yes, then stop the package else continue.
   iv. Load product data
   v. Log package end.

Error handling: In case of any error, the package should error out and stop. The error should be logged in the log table.

Pre/Post execute: The package should also include tasks to calculate the step number and log every task detail in the log detail table.

Load Fact Data

This package will load the fact data.

Invoice data*:* Invoice data is loaded from flat files into the staging database. The file format name is Invoice_YYYYMM. The package will then cleanse the data and load the data from stage into the mart. The flow of the package would be as follows:

1. Get the respective connection settings from the settings table
2. Log package start
3. Check if any other package is running. If yes, then stop the package else continue.
4. Load invoice data - Check if an invoice file exists in the expected location. If yes, proceed with the below steps. If not, log an appropriate message and proceed to step 6.
   i. Obtain the month for which invoices are being loaded from the file name.
   ii. Perform basic validations like eliminating blank spaces/blank lines.
   iii. Load file into stage table.
   iv. Load the data from stage to mart table
5. Set the current day in the time dimension table based on the latest day for which fact data exists.
6. Log package end.

Error Handling: In case of any error, the package should error out and stop. The error should be logged in the fact table.

Pre/Post Execute: The package should also include tasks to calculate the step number and log every task detail in the log table.

Wrapper Package

This package is used to call all of the above two packages in a sequential manner.

1. Get the respective connection settings from the settings table
2. Log package start
3. Call dimension loader
4. Call fact loader
5. Log package end.

Error handling: In case of any error, the package should error out and stop. The error should be logged in the fact table.

Pre/Post execute: The package should also include tasks

to calculate the step number and log every task detail in the log table.

C. Multi-Dimensional Cube

There are two types of schema that can be used while trying to build a multidimensional cube. These are the Snowflake schema and the Star schema. For this project we will use the star schema.

The star schema separates business process data into facts, which hold the measurable, quantitative data about a business, and dimensions which are descriptive attributes related to fact data [3]. Examples of fact data include sales price, sale quantity, and time, distance, speed, and weight measurements. Related dimension attribute examples include product models, product colours, product sizes, geographic locations, and salesperson names.

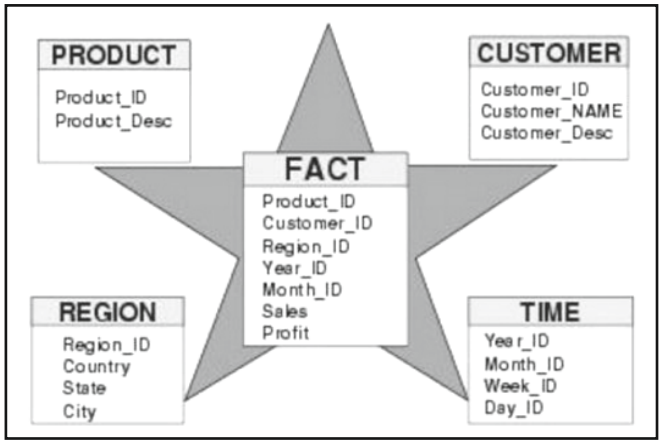

Fig. 5. Example of Star Schema

After creating the star schema the relationships should be established after which the code for calculated columns have to be scripted using MDX [4].

D. SSRS Dashboard

In the dashboard connect to the SSAS cube that we have created. Drag in the appropriate columns and measure on the basis of the dashboard requirement. Choose the appropriate graphical representation [5]. Make the dash board visually attractive.

IV. METHODOLOGY

i. Business Case: This process builds on the work of the startup process, and the project brief is augmented to form a business case. The approach taken to ensure quality of the project is agreed together with the overall approach to controlling the project itself. Project files are also created as is an overall plan for the project. A plan for the next stage of the project is also created.

ii. Project initiation model: Key activities include: planning quality, planning a project, refining the business case and risks, setting up project controls, setting up project files, and assembling a Project Initiation Document. A general E.R diagram has also been defined which shows relation of an entity with others. This domain model will contain the main domain entities, their major attributes, and the relationships between these entities.

iii. User interface model: User interface diagram has been defined under this model which shows flow of screens within the system [6].

V. RESULTS AND ANALYSIS

Figures 6,7, and 8 show the process in which the SSIS processes are initiated to ensure that the extract, transform, and load process handles all exceptions and loads data at given intervals. These packages are then deployed to the SQL Server 2012 interface so as to deploy with environment variables [7].

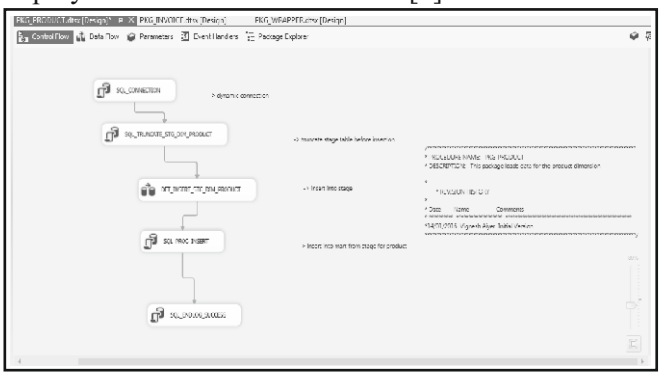

Fig. 6. Product package in SQL Server Integration Services

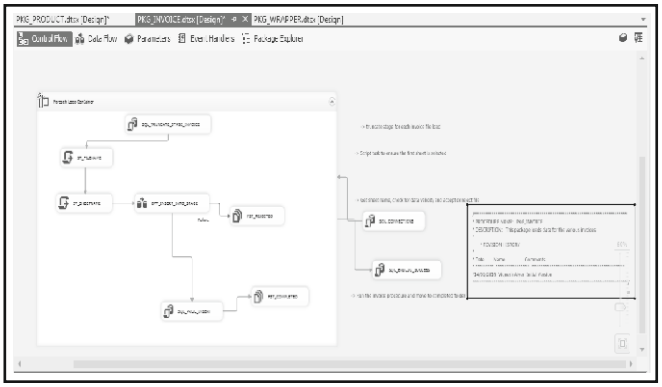

Fig. 7. Invoice package in SQL Server Integration Services

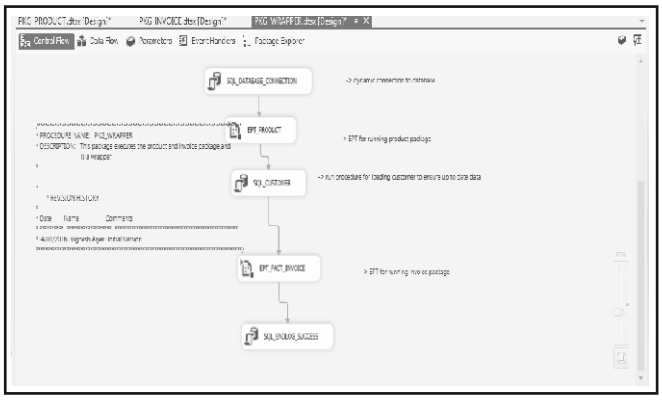

Fig. 8. Wrapper Package in SQL Server Integration Services

In fig. 9, we see the overview in SSAS for the cube that has been created. It shows the relationship between tables and the hierarchy that has been used.

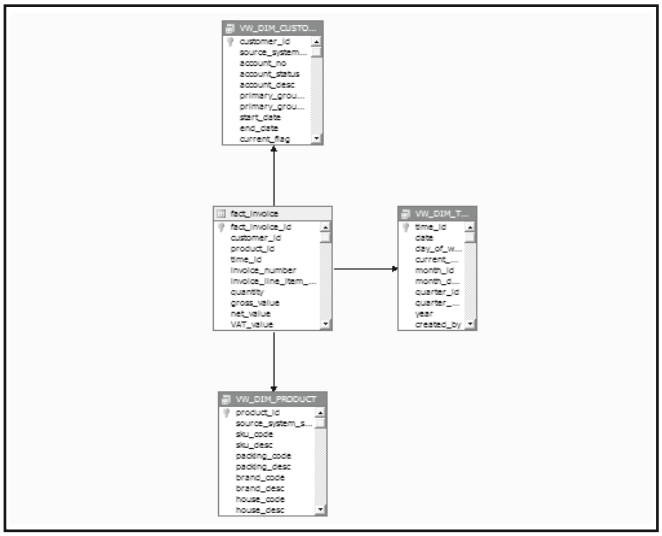

Fig. 9 SSAS Cube dimension overview

Finally, the reports are developed based on the cube in SSRS which shows the net invoice value for the product at different levels and for different customers in fig. 10.

Fig. 10. Dashboard view

VI. CONCLUSION

The development of the application has been successfully completed with the results of all the tests being very positive with minute bugs found and fixed. There have been no issues regarding the application and it is serving its purposes very efficiently. The system provides end to end development of Business Intelligence solution. It facilitates the development of analytical reports rather than transactional reports that help top level managers in decision making. It enables systematic management of the master and the fact data. The system also ensures the optimization of each unit leads to faster overall execution time [8].

The technical advantages of the application are:

I. Flexibility

ii. Highly reliable

iii. Simple and easy to understand by business users

REFERENCES

[1] K.Kevin and K.Daniel, SQL in a nutshell, O'Reilly &Associates, Inc., Sebastopol, CA, 2001.

[2] K.Brian, M. Allan, and G. Darren, Professional SQL Server 2008 Integration Services (Programmer to Programmer), Wiley, John & Sons, 2008.

[3] W.Mark, Z. Robert, and P. Mosha, Fast track to MDX. New Springer Publications, 2005.

[4] Database journal, the knowledge center for database professionals.Available: http://www.databasejournal.com. Accessed: May, 2016.

[5] Excel Macros and programming. Available: http://excel-vba.com. Accessed on May 1, 2016.

[6] MSDN: Microsoft Developer Network, Available: http://www.msdn.microsoft.com. Accessed on : May, 1, 2016.

[7] Microsoft SQL Server tutorials. Available: http://www.sqlcentral.com. Accessed on: May ,1, 2016.

[8] DBA, developer and performance tuning articles. Available: http://www.sql-serverperformance.com. Accessed on: May, 1, 2016.

About the Authors

Vignesh Aiyer was born in Chennai on 09/04/1995. He has a B.Tech degree in Electronics & Communication Engineering from VIT University, Vellore, India awarded in the year 2016.

Sujaan Choksi was born in Mumbai on 18/05/1993. He has a B.Tech. degree in Electronics & Communication Engineering from Manipal University, Manipal, India awarded in the year 2015.

Sujaan has completed an internship at Thorogood Associates, a Business Intelligence Consultancy firm in Bengaluru, India from December 2014-May 2015. He has been working as a Business Intelligence Consultant at Thorogood Associates since June 2015.

Abishek Murali was born in Chennai on 28/03/1993. He has a B.Tech. degree in Electronics and Communication Engineering from BMSCE, Bangalore, India awarded in the year 2015.

He has been working as a Business Intelligence Consultant at Thorogood Associates since June 2015.