

SALES BASED DATA EXTRACTION FOR BUSINESS INTELLIGENCE

Sukanta Singh

Department of Computer Science & Engineering, Global Institute of
Management & Technology, Krishnagar, Nadia, West Bengal
sukantasingh2008@gmail.com

ABSTRACT

Data warehouse use has increased significantly in recent years and now plays a fundamental role in many organizations' decision-support processes. An effective business intelligence infrastructure that leverages the power of a data warehouse can deliver value by helping companies enhance their customer experience. In this paper is to generate reports with various drilldowns and slicer conditions with suitable parameters which provide a complete business solution which is helpful for monitor the company inflow and outflow. The goal of the work is for potential users of the data warehouse in their decision making process in the Business process system to get a complete visual effort of those reports by creating the chart and grid interface from warehouse. The example in this paper relate directly to the Adventure Work Data Warehouse Project implementation which helps to know the internet sales amount according to different date.

KEYWORDS

Data Warehouse Architecture, ETL, Business Intelligence (BI), Dimension Table, Fact table, Star Schema, Snowflake Schema, Business Intelligence Reporting

1. INTRODUCTION

A data warehouse is the main repository of an organization's historical data, its corporate memory. It contains the raw material for management's decision support system. The critical factor leading to the use of a data warehouse is that a data analyst can perform complex queries and analysis, such as data mining, on the information without slowing down the operational systems. [3] Data warehouse contain a wide variety of data that present a coherent picture of business conditions at a single point in time. It is a repository of integrated information, available for queries and analysis. [6]

Organizations now consolidate information from several databases into a data warehouse. Organization decision making requires a comprehensive view of all aspects of an enterprise, and many organizations have therefore created consolidated data warehouses that contain data drawn from several databases maintained by different business units together with historical and summary information. The trend toward data warehousing is complemented by an increased emphasis on powerful analysis tools. [2]

By working with real deployment scenarios, we gain a complete understanding of how to work with the tools. My goal is to address the full gamut of concerns that a large company would face during their own real-world deployment. [1]This paper focuses on the SQL Server Integration

Services (SSIS) extraction, transformation, and loading (ETL) design for Project Adventure Work Data Warehouse.

In this system, metadata plays an important role and provides the foundation for all actions in all stages. It can be considered as glue sticking together all individual parts of these systems. In this paper, I propose my data warehouse architecture with new metadata layer and describe the design and implementation of star schema in conceptual data model.

2. THE DATA WAREHOUSE ARCHITECTURE

This architecture consists of core components and processes which constitute a complex analytical information system. The bottom layer of the warehouse connects to the data stores of operational information systems (e.g. accounting, sales, human resources) in order to extract relevant data which are pooled into a staging area. In addition to this, external data (e.g. from market research institutes or financial services) can be acquired to complete the data needed for management decisions.[7] The staging area serves as a temporary data store and allows the consolidation of data from heterogeneous sources. A typical transformation task is to remove homonyms (e.g. using an identical label for different types of attributes) and synonyms (e.g. using two different labels for an attribute). As a result, the integrated data is loaded into a central data warehouse layer which is typically normalized in order to avoid redundancies. In figure 1 it is shown how different data from multiple sources can be fetched and store in data warehouse as form of metadata. Finally, generate reports from this data warehouse.

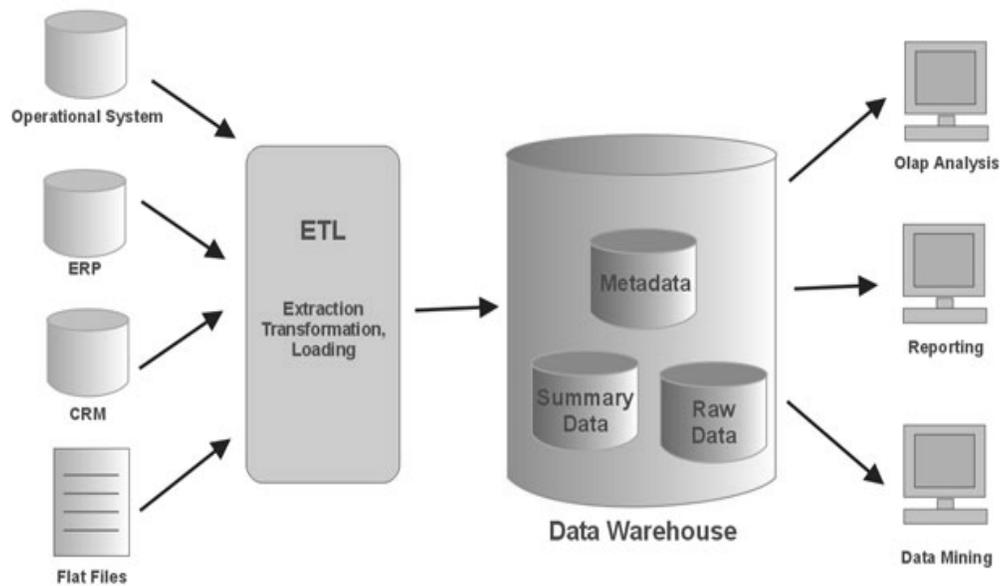


Figure 1: Data Warehouse Architecture

Typically, the informational needs of the marketing department differ from those of the accounting department. As a consequence of this, specific departmental views on the data have to be created.[4] These views, which are called data marts, can be further customized in order to comply with the informational needs of single users (e.g., a specific salesperson in a defined region). To get information from these data marts, end users are provided with a set of tools which allow analytical processing. Most commonly are report generation tools which support simple aggregations (e.g. calculation of statistical measures like mean values, etc.). In order to provide interactive analysis with user defined views, OLAP tools are frequently used. These early tasks of the data warehouse process are executed by use of ETL-tools (extraction, transformation,

loading). These tools provide connectivity to a broad set of different data storage formats (e.g. different database systems like Oracle, DB2 or SQL Server or different text file formats).

3. MULTIDIMENSIONAL DATA MODEL AND SCHEMAS

Multidimensional models highly take advantage of inherent relationships existing in data to populate data in multidimensional matrices referred to as data cubes. The Multidimensional storage model involves two types of tables: dimension tables and fact tables.

In Data Warehouse schema is three types- Star Schema, Fact Constellation Schema and Snowflake schema. Two common multidimensional schemas are star schema and snowflake schema.

3.1. Dimension Table

Dimensional modeling produces dimension tables in which each table contains fact attributes that are independent of those in other dimensions. For example, a customer dimension table contains data about customers, a product dimension table contains information about products, and a store dimension table contains information about stores.

3.2. Fact Table

A key characteristic of a fact table is that it contains numerical data (facts) that can be summarized to provide information about the history of the operation of the organization. Each fact table also includes a multipart index that contains as foreign keys the primary keys of related dimension tables, which contain the attributes of the fact records.

3.3. Star Schema

The star schema consists of a fact table with a single table for each dimension. In star schema a single fact table and for each dimension one dimension table and does not capture hierarchies directly. The primary keys of each of the dimension tables are linked together to form the composite primary key of the fact table. In figure 2 show a star schema design, there is only one de-normalized table for a given dimension.

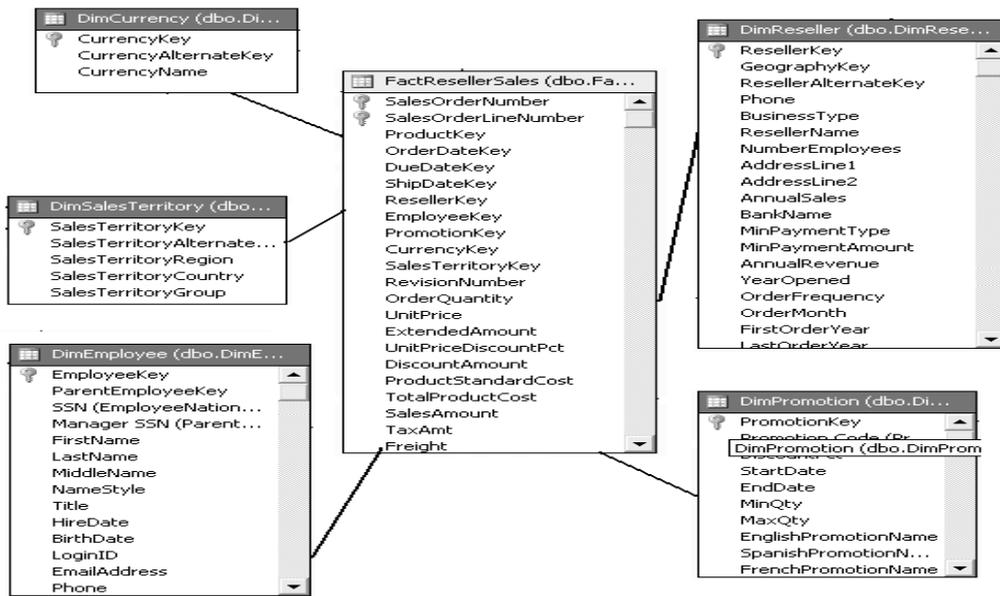


Figure 2: Star Schema for proposed system

3.4. Snowflake Schema

The snowflake schema is a variation on the star schema in which the dimensional tables from a star schema are organized into a hierarchy by normalizing them. But, in figure 3 show snowflake schema represent dimensional hierarchy directly by normalizing tables and easy to maintain and saves storage.

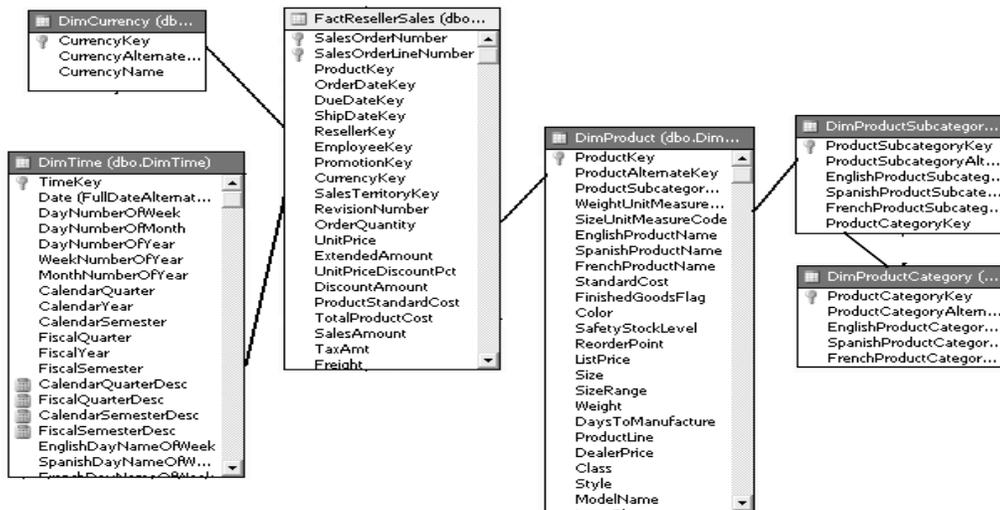


Figure 3: Snowflake Schema for proposed system

4. DATA EXTRACTION FOR PROPOSED SYSTEM

An ETL application is the most comprehensive line between two points. These two points are the enterprise and all its source systems on one end and a data warehouse on the other. The ETL process is much more than code written to move data. The ETL architect also serves as the central

point for understanding the various technical standards that need to be developed if they don't already exist. In figure 4, a query for retrieving data from ExcelSheet to MS SQL Server for Customer Table shown that how data can be extracted from different data source and finally load as data mart in warehouse.

```

INSERT INTO
AdventureWorksDW(CustomerKey, GeographyKey, CustomerAlternateKey,
Title, FirstName, MiddleName, LastName, NameStyle, BirthDate, MaritalStatus, S
uffix,
Gender, EmailAddress, YearlyIncome, TotalChildren, NumberChildrenAtHome,
EnglishEducation, SpanishEducation, FrenchEducation, EnglishOccupation,
SpanishOccupation, FrenchOccupation, HouseOwnerFlag, NumberCarsOwned,
AddressLine1, AddressLine2, Phone, DateFirstPurchase, CommuteDistance)
SELECT CustomerKey, GeographyKey, CustomerAlternateKey, Title, FirstName,
MiddleName, LastName, NameStyle, BirthDate, MaritalStatus, Suffix, Gender,
EmailAddress, YearlyIncome, TotalChildren, NumberChildrenAtHome,
EnglishEducation, SpanishEducation, FrenchEducation, EnglishOccupation,
SpanishOccupation, FrenchOccupation, HouseOwnerFlag, NumberCarsOwned, Adre
ssLine1,
AddressLine2, Phone, DateFirstPurchase, CommuteDistance
FROM OPENROWSET
('Microsoft.Jet.OLEDB.4.0',
'Excel 8.0;Database=C:\Test.xls;',
'SELECT CustomerKey, GeographyKey, CustomerAlternateKey,
Title, FirstName, MiddleName, LastName, NameStyle, BirthDate,
MaritalStatus, Suffix, Gender, EmailAddress, YearlyIncome, TotalChildren,
NumberChildrenAtHome, EnglishEducation, SpanishEducation, FrenchEducation,
EnglishOccupation, SpanishOccupation, FrenchOccupation, HouseOwnerFlag, Num
berCarsOwned,
AddressLine1, AddressLine2, Phone, DateFirstPurchase, CommuteDistance FROM
[Sheet2]')

--EXEC sp_configure 'show advanced options', 1;
--GO
--RECONFIGURE;
--GO
--EXEC sp_configure 'Ad Hoc Distributed Queries', 1;
--GO
--RECONFIGURE;
--GO

```

Figure 4: Retrieving data from ExcelSheet to MS SQL Server for Customer Table

5. IMPLEMENTATION FOR PROPOSED SYSTEM

In proposed system user can be able for visualization about total Internet sales amount, Internet gross profit etc. according to date, geography. BI Reporting is what the data warehouse customers see. All the hardware, software, data architectures, data models, Source System Analysis, Target System Analysis, and ETL applications culminate in data displayed on a computer monitor or printed on a piece of paper. All that work, the effort and investment, will be counted a blazing failure if the BI Reporting fails to deliver or all of that work, the effort and investment, will be counted a tremendous success if the BI Reporting delivers the data and value expected of a data warehouse, which will lead to subsequent iterations of data warehouse development. Here, some screenshots of web based business intelligence report of Adventure Work database included. Figure 5 generate a report about chart view of internet sales amount by year wise which is fetch from Adventure Work Data warehouse cube structure.



Figure 5: Chart View of Internet Sales Amount by Date

Figure 6 show grid view of Internet Sales amount by customer wise according to date. Here different Internet sales amount are listed according to different countries by employee.

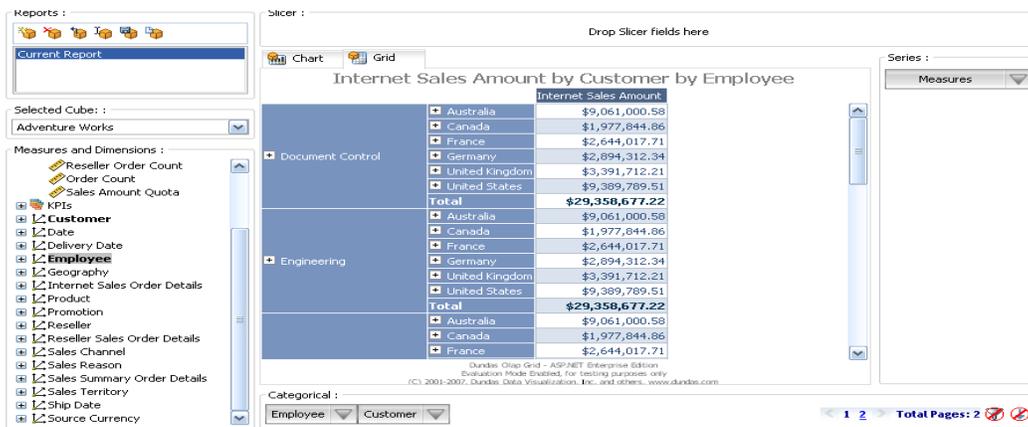


Figure 6: Grid View of Internet Sales Amount by Customer by Employee

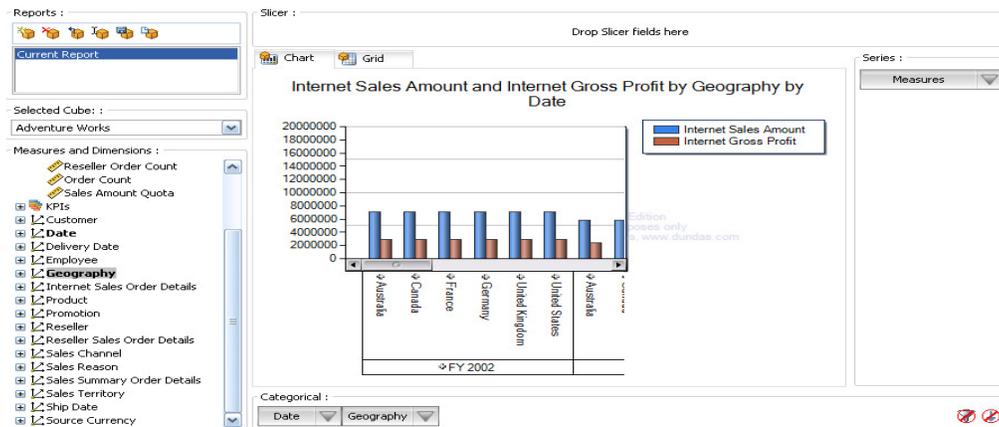


Figure 7: Chart Views of Internet Sales Amount and Internet Gross Profit by Geography by Date

Chart View and Grid View of Internet Sales Amount and Internet Gross Profit by Geography (i.e., different country wise) depicted in figure 7 and figure 8. Here you can see the details of Average unit price, gross profit and sales amount of Internet according to different time slot of a year.

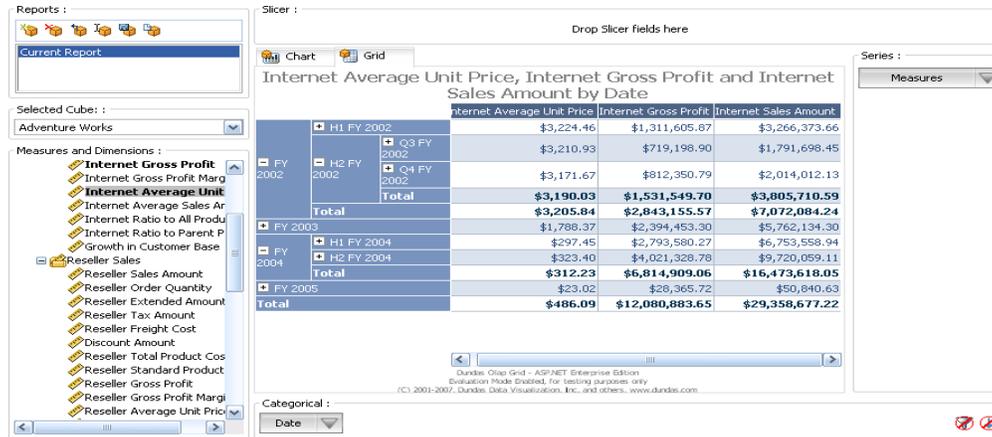


Figure 8: Grid Views of Internet Average Unit Price, Internet Gross Profit and Internet Sales Amount by Date

6. CONCLUSION

Create a complete web Business Intelligence tool based on an organization using Data Warehouse and maintain a strong ETL process which will be dynamic and must more efficient. Hope this web tool is very useful for an organization's financial department. Data warehouses are still an expensive solution and typically found in large firms. The development of a central warehouse is a huge undertaking and capital intensive with large, potentially unmanageable risks.

Data warehousing is the leading and most reliable technology used today by companies for planning, forecasting, and management. After the evolution of the concept of data warehousing during the early 90's it was thought that this technology will grow at a very rapid pace but unfortunately it's not the reality. A major reason for data warehouse project failures is poor maintenance. Without proper maintenance desired results are nearly impossible to attain from a data warehouse. Unlike operational systems data warehouses need a lot more maintenance and a support team of qualified professionals is needed to take care of the issues that arise after its deployment including data extraction, data loading, network management, training and communication, query management and some other related tasks.

REFERENCES

- [1] S. Sai Satyanarayana Reddy, A.Lavanya, Dr.V.Khanna, Dr.L.S.S.Reddy "RESEARCH ISSUES ON DATA WAREHOUSE MAINTENANCE" International Conference on Advanced Computer Control 2008 IEEE
 - [2] Huang, L. & Ulrich, T. & Hemmje, M. & Neuhold, E.J., "Adaptively Constructing the Query Interface for Meta-Search Engines." GMD-IPSI Dvllivostr.15, Darmstadt, 64293 Gemany, 2001.
 - [3] Nilakanta. Sree, Scheibe. Kevin, "Dimensional issues in agricultural data warehouse designs", Computers and electronics in agriculture, Vol.60, No.2, 2008, pp.263-278
 - [4] Jan Chmiel, Tadeusz Morzy, Robert Wrembel, "Multiversion Join Index For Multiversion Data Warehouse", Information and Software Technology □ Vol.51 □ No.1, 2009, pp.98-108.
- A. Simitsis, "Modeling and Managing ETL Processes," Proceedings of the VLDB PhD Workshop, Germany, 2003.

- [5] Than Than Wai, Sint Sint Aung, “*Metadata Based Student Data Extraction from Universities Data Warehouse*” National Inst of Tech Teachers Training and Res 2009 IEEE.
- [6] Hanson, Joseph H. An Alternative Data Warehouse Structure for Performing Updates. December 1996, UMI Press.
- [7] Labio, W.J. and H. Garcia-Molina, Efficient Snapshot Differential Algorithms for Data Warehousing, Technical Report. 1996, Stanford Univ: Palo Alto.
- [8] McElreath, J., Data Warehouses: An Architectural Perspective. Perspectives, November 1995. Computer Sciences Corporation, El Segundo, CA: p. 13.
- [9] Meta Software Corp., Using DesigdIDEF to Simulate Workflow Models with ServiceMOdel, 1995, Meta Software Corp.: Cambridge MA. p. 1-25.
- [10] Zhuge, Y.H.G.-M., Jachim Hammer, Jennifer Widom, *View Maintenance in a Warehousing Environment*. Technical Report. 1994. Stanford: Palo Alto.
- [11] CHEN Xi-ping; ZHANG Liang; LI Chun-xiang; CAO Jie, “Data warehouse and its application to decision supporting system of public traffic”, *Journal of Gansu University of Technology*, 2001(04).

AUTHORS

Sukanta Singh received B.Tech in Computer Science & Engineering from Govt. College of Engineering & Ceramic Technology, WBUT, India in 2008 and M.Tech in Computer Science & Engineering from Kalyani Govt. Engineering College, WBUT, and India in 2010. He has more than 3.5 years of teaching experience and currently working as an Assistant Professor in Global Institute of Management & Technology, West Bengal, India. His fields of interest are soft computing, theory of computation and data warehouse.

