International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A Data Warehouse Design for Institutions Information System

A G P Kujur[1], Ajay Oraon[2]

[1]Department of Computer Sc and Engg, BIT Sindri

[2]Department of Chemical Engg, BIT Sindri

Abstract: ***Data Warehousing is a technique of bringing collectively all of a company's data from different computer systems, together with those connecting to customers, employees, vendors, product, inventory, and financial etc . The data warehouse connects different database together in order to offer a more inclusive data set for making decision. Presently, large enterprises work on database systems to manage their data and information. These databases are useful for conducting daily business transactions. However, the tight competition in the marketplace has led to the concept of data mining in which data are analyzed to derive effective business strategies and discover better ways in carrying out business. The paper considers how different ways of determining such warehouse have been developed and how confident organizations have used them to increase control over data and decision making. This reveals that organization that can develop a strong system, data warehousing is value the cost. A physical repository where relational data are specially organized to provide enterprise, cleansed data in a standardized format.***

Keywords: Data Warehouse, DBMS, Data Mining, Information System

1. Introduction

Now a days, almost every enterprise uses a database to store its vital data and information . For instance, dynamic websites, accounting information systems, payroll systems, stock management systems all rely on internal databases as a container to store and manage their data. The competition in the marketplace has led business managers and directors to seek a new way to increase their profit and market power, and that by improving their decision making processes. In this sense, the idea of data warehouse and data mining was born . In fact, data warehousing is the process of collecting data from operational functional databases, transforming, and then archiving them into special data repository called data warehouse with the goal of producing accurate and timely management information ; whereas, data mining is the process of discovering trends and patterns from data warehouse, useful to carry out data analysis .

A typical Institute or university often comprises a lot of subsystems crucial for its internal processes and operations. Examples of such subsystems include the student registration system, the Library system, the accounting system, the course management system, the staff Management system, and many others. All these systems are connected to many distributed databases that are employed for every day transactions and processes. However, Institute rarely employ systems for handling data analysis, forecasting, prediction, and decision making. This paper proposes a data warehouse design for a typical university information system whose role is to help in and support decision making. The proposed design transforms the existing operational databases into an information database or data warehouse by cleaning and scrubbing the existing operational data. Besides, several columns and structures are dropped as they are useless for data mining. Finally, all captured and cleaned data are loaded and indexed in the warehouse making them ready for conducting data mining tasks.

2. Methodology of warehouse design

1) **Operational Database design:** An operational database is a regular database meant to run the business on a current basis and support everyday transactions and processes .
2) **Informational Database design:** An informational database is a special type of database that is designed to support decision making based on historical point-in-time and prediction data for complex queries and data mining applications. A data warehouse is an example of informational database.
3) **Data Warehouse design:** It is a subject-oriented, integrated, time-variant, non-updatable collection of data used in support of management decision-making processes. It is subject-oriented as it studies a specific subject such as sales and customers behavior. It is integrated as it defines consistent naming conventions, formats, and encoding structures from multiple data sources. It is time-variant as it studies trends and changes over time. It is non- updatable as it is read-only, i.e. cannot be updated by regular users.
4) **Data Mining:** In its broader sense, it is a knowledge discovery process that uses a blend of statistical, machine learning, and artificial intelligence techniques to detect trends and patterns from large data-sets, often represented as data warehouse. The purpose of data mining is to discover news facts about data helpful for decision makers .

3. Data warehouse design for BIT Sindri

1) Operational Database used in BIT Sindri

Following operational database used at institute are in different database like Sql server,Ms Access, Excel etc.

Student registration system for handling students' registration processes in Sql server accounting system for managing students' fees payments, a course management system for managing courses and assigning them particular sections and instructors, an alumni system for archiving students' records after graduation in MS Access

An assets system for distributing items such as machines, equipment, and computers over different departments In oracle. Some information save in image format like jpg, bmp etc.

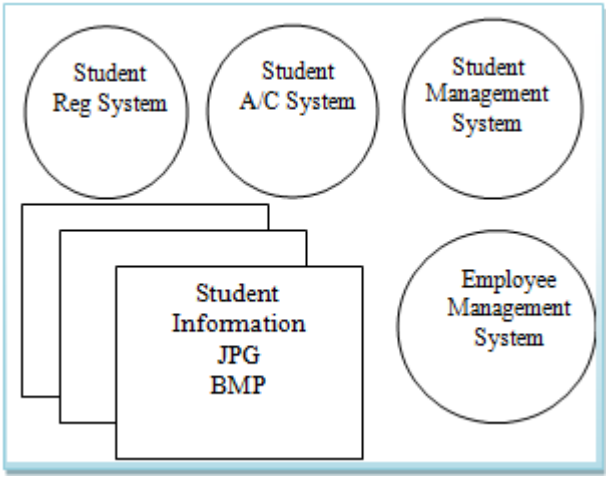

Figure 1: Operational database.

2) Data Warehouse Design – Informational Database design for BIT Sindri

This section discusses the design of the proposed informational database along with its architecture, building process, SQL queries, and conceptual schema. In order to build the informational database, four vital steps are required to be completed, and they are respectively: Capture and extract, scrub and data cleansing, transform, and load and index.

.

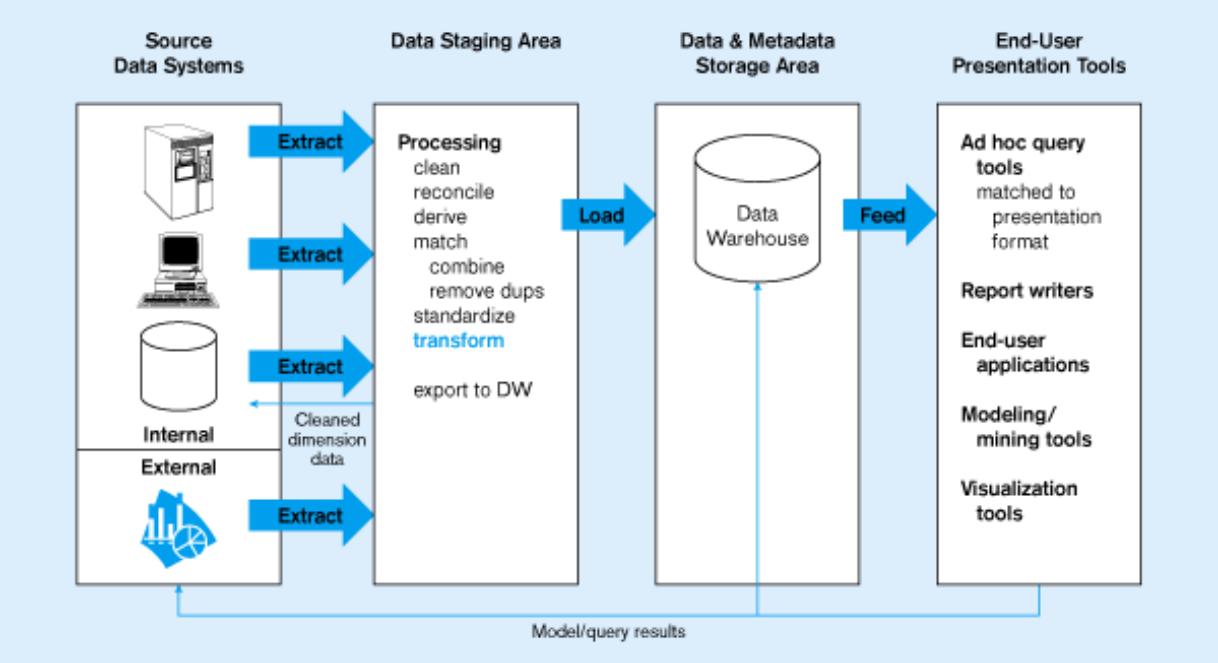

Figure 2: Different stages required to transform the operational database into an informational database or data warehouse

3) Some basic step of warehouse design

- **Design Snowflake Schema** The proposed data warehouse is a Snowflake type design
- **Data Cleansing and Transforming**:This section discusses all major transformations and SQL queries required in order to build and load the data warehouse.
- **Dropping Tables**
- **Merging Tables**
- **Add New Columns**
- **Removing Columns**
- **Implementation: Loading and Indexing**

4. Conclusions

This paper introduced a model for building a data warehouse for a typical **BIT Sindri** information system. The warehouse is an informational database whose data are extracted from an already existing operational database.

The purpose of the proposed design is to help decision makers and **BIT Sindri** principles in performing data mining and data analysis over the data stored in the warehouse which eventually helps them in discovering critical patterns and trends.

References

[1] Pant, S., Hsu, C., "Strategic Information Systems Planning: A Review", Information Resources Management Association International Conference, May 21–24, 1995, Atlanta.
[2] Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", Hershey, New York, 2007.
[3] Preeti S., Srikantha R., Suryakant P., "Optimization of Data Warehousing System: Simplification in Reporting and Analysis", International Journal of Computer Applications, vol. 9 no. 6, pp. 33–37, 2011.
[4] Kantardzic, Mehmed, "Data Mining: Concepts, Models, Methods, and Algorithms", John Wiley & Sons, 2003.
[5] William Inmon, "Building the Operational Data Store", 2nd ed., John Wiley & Sons, 1999.
[6] Matteo Golfarelli & Stefano Rizzi, "Data Warehouse Design: Modern Principles and Methodologies", McGraw-Hill Osborne Media, 2009.
[7] Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introduction to Data Mining", Addison Wesley, 2005.
[8] Microsoft Access Home Page, http://office.microsoft.com/en-us/access/default.aspx
[9] Jeffrey A. Hoffer, Mary Prescott, Heikki Topi, "Modern Database Management", 9th ed, Prentice Hall, 2008.
[10] Robert Laberge, "The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence Insights", McGraw-Hill Osborne Media, 2011.
[11] Anahory, S., D. Murray, "Data Warehousing in the Real World: A Practical Guide for Building Decision Support Systems", Addison Wesley Professional. 1997.
[12] Michael Goebel, Le Gruenwald, "A Survey of Data Mining and Knowledge Discovery Software Tools", Proceedings of the ACM SIGKDD, vol. 1, no. 1, pp. 20-33, 1999.