

Data warehousing in telecom Industry

Dr. Sanjay Srivastava, Kaushal Srivastava, Avinash Pandey, Akhil Sharma

Abstract: Data Warehouse is termed as the storage for the large heterogeneous data collected from different sources and departments for the one organization. It is collected and maintained at a separate location for the analysis purpose by the expert and particular concern team to understand and predict the support and confidence of product and services of any organization. By this we can predict and place the particular item at profitable place .It with the help of data mining tools and techniques, Make us to understand the knowledge of customer and its requirements. Telecom industries have many departmental databases which are integrated to form large databases or Data warehouses. The aim of this paper is to describe about the use of data warehouse in managing and storing for telecom industry. The telecom industry contains very large amount of data, so it is difficult and time consuming to retrieve data related to particular field from such a large databases. So for this, one of the aim of this thesis is to show how data warehouse efficiently manage and store large data marts and how query performance improves due to this.

Keywords: Data warehouses, Design prediction

I. INTRODUCTION

A data warehouse is a relational database that is designed for query and analysis rather than for transaction processing. It usually contains historical data derived from transaction data, but it can include data from other sources. It separates analysis workload from transaction workload and enables an organization to consolidate data from several sources.

In addition to a relational database, a data warehouse environment includes an extraction, transportation, transformation, and loading (ETL) solution, an online analytical processing (OLAP) engine, client analysis tools, and other applications that manage the process of gathering data and delivering it to business users.

The concept of data warehousing has evolved out of the need for easy access to a structured store of quality data that can be used for decision making. It is globally accepted that information is a very powerful asset that can provide significant benefits to any organization and a competitive advantage in the business world. Organizations have vast amounts of data but have found it increasingly difficult to access it and make use of it. This is because it is in many different formats, exists on many different platforms, and resides in many different file and database structures developed by different vendors. Thus organizations have had to write and maintain perhaps hundreds of programs that are used to extract, prepare, and consolidate data for use by many different applications for analysis and reporting. Also, decision makers often want to dig deeper into the data once initial findings are made. This would typically require modification of the extract programs or development of new ones. This process

is costly, inefficient, and very time consuming. Data warehousing offers a better approach.

Data warehousing implements the process to access heterogeneous data sources clean, filter, and transform the data and store the data in a structure that is easy to access, understand, and use. The data is then used for query,

Dr. Sanjay Srivastava, Professor ,Computer Science, MGMCoET, Noida, India, E-mail: san.sri2k@gmail.com
Kaushal Srivastava, Scholar ,Computer Science, MGMCoET, Noida, India, E-mail: lightkaushal@gmail.com
Avinash Pandey, Scholar ,Computer Science, MGMCoET, Noida, India, E-mail:avipandey1508@gmail.com
Akhil Sharma, Scholar ,Computer Science, MGMCoET, Noida, India, E-mail: Akhil.sharma0612@gmail.com

reporting, and data analysis. As such, the access, use, technology and performance requirements are completely different from those in a

Transaction-oriented operational environment. The volume of data in data warehousing can be very high, particularly when considering the requirements for historical data analysis. Data analysis programs are often required to scan vast amounts of that data, which could result in a negative impact on operational applications, which are more performance sensitive. Therefore, there is a requirement to separate the two environments to minimize conflicts and

Degradation of performance in the operational environment.

II. LITERATURE SURVEY

The origin of the concept of data warehousing can be traced back to the early 80's. The concept of data warehousing has evolved out of the need for easy access to a structured store of quality data that can be used for decision making. It is globally accepted that information is a very powerful asset that can provide significant benefits to any organization and a competitive advantage in the business world. Organizations have vast amounts of data but have found it increasingly difficult to access it and make use of it. This is because it is in many different formats, exists on many different platforms, and resides in many different file and database structures developed by different vendors. Thus organizations have had to write and maintain perhaps hundreds of programs that are used to extract, prepare, and consolidate data for use by many different applications for analysis and reporting. Also, decision makers often want to dig deeper into the data once initial findings are made. This would typically require modification of the extract programs or development of new ones. This process is costly, inefficient, and very time consuming. Data warehousing offers a better approach.

III. CHARACTERISTICS OF DATA WAREHOUSE

Following are the characteristics given by data warehousing:

➤ Subject Oriented

Data warehouses are designed to help you analyze data. For example, to learn more about our company's sales data, we can build a warehouse that concentrates on sales. Using this warehouse, we can answer questions like "Who was our best customer for this item last year?" This ability to define a data warehouse by subject matter, sales in this case, makes the data warehouse subject oriented.

➤ Integrated

Integration is closely related to subject orientation. Data warehouses must put data from disparate sources into a consistent format. They must resolve such problems as naming conflicts and inconsistencies among units of measure.

➤ Non volatile

Non volatile means that, once entered into the warehouse, data should not change. This is logical because the purpose of a warehouse is to enable you to analyse what has occurred.

➤ Time Variant

In order to discover trends in business, analysts need large amounts of data. This is very much in contrast to online transaction processing (OLTP) systems, where performance requirements demand that historical data be moved to an archive. A data warehouse's focus on change over time is what is meant by the term time variant.

IV. COMPARISON BETWEEN OLAP AND DATA WAREHOUSE

The comparison between OLAP and OLTP data warehouse is mentioned below:

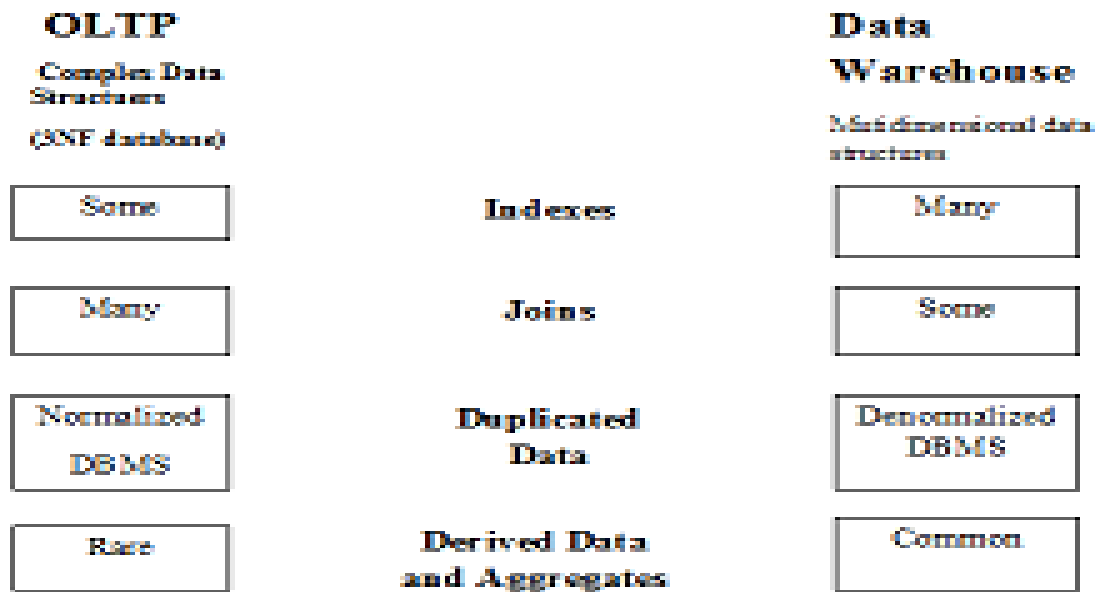


Fig4.1 Comparison figure

S. No	Basis of comparison	OLTP	Data warehouse
1	Historical data	OLTP systems usually store data from only a few weeks or months. The OLTP system stores only historical data as needed to successfully meet the requirements of the current transaction	Data warehouses usually store many months or years of data. This is to support historical analysis
2	Schema design	It uses fully normalized schemas to optimize update/insert/delete performance and guarantee data consistency	It uses denormalized or partially denormalized schemas to optimize query performance.
3	Typical operations	A typical OLTP operation accesses	Data warehouse query scans

		only a handful of records.	thousands or millions of rows.
4	Data modifications	The end users routinely issue individual data modification statements to the database. The OLTP database is always up to date, and reflects the current state of each business transaction	A data warehouse is updated on a regular basis by the ETL process using bulk data modification techniques. The end users of a data warehouse do not directly update the data warehouse.
5	Workload	OLTP systems support only predefined operations.	Data warehouses are designed to accommodate <i>ad hoc</i> queries.

Table 5.1 Table of comparison

Conclusion

The data warehouse is build to provide an easy to access source of high quality data from the easy to access heterogeneous source situated at different location. The data warehouse provide an easy way to model or represent large databases .In telecom industries large amount of data bases are represented using star schema and snow flake schema model. The data warehouse provide best way to access data concerning to particular field.

The data ware house is best way organise and maintain data in large data ware houses. The telecom industry have various analytical transaction per day and these queries can be easily processed if the data is represented using star schema and snowflake schema. In star schema more than one dimension table are there and each dimension table is connected to fact table using foreign keys. But in the case of snow flake schema dimension tables are further extend to other dimension table and the dimension table are connected to fact tables but extended dimension table are not connected to fact table.

REFERENCES

- [1] J.-L. Amat, "Using reporting and data mining techniques to improve knowledge of subscribers; applications to customer profiling and fraud management", *J. Telecommun. Inform. Technol.*, no. 3, pp. 11–16, 2002.
- [2] K. Morik and M. Scholz, "The Minin gMart approach to knowledge discovery in databases" in *Handbook of Intelligent IT*, Ning Zhong and Jiming Liu, Eds. IOS Press, 2003
- [3]. Rosset, S., Neumann, E., Eick, U., & Vatnik (2003). Customer lifetime value models for decision support. *Data Mining and Knowledge Discovery*, 7(3), 321- 339.
- [4] Data mining in Telecommunication by Gray M. Weiss, Fordham University
- [5] Cortes, C., Pregibon, D. Giga-mining. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*; 174-178, 1998 August 27-31; New York, NY:AAAI Press, 1998
- [6] Ezawa, K., Norton, S. Knowledge discovery in telecommunication services data using Bayesian network models. *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*; 1995 August 20-21. Montreal Canada. AAAI Press: Menlo Park, CA, 1995
- [7] The data warehouse toolkit: the complete guide to dimensional modeling by Ralph Kimball, Margy Ross. — 2nd ed. (ISBN 0-471-20024-7).
- [8] Berry, M. and Linoff, G. *Data Mining Techniques for Marketing, Sales and Customer Support*. Wiley (1997).
- [9] Chaudhri, S. and Dayal, U. An overview of Data Warehousing and OLAP Technology. *ACM SIGMOD Record* 26 (1), (1997), 65-74.
- [10] Hilderman, R.J., Carter, C.L., Hamilton, H.J., and Cercone, N. Mining Association Rules from Market Basket Data Using Share Measures and Characterized Itemsets. *International Journal of Artificial Intelligence Tools*. 7 (2), (1998), 189-220.
- [11] Inmon, W. H. *Building the Data Warehouse*. Wiley (1996).
- [12] Kimball, R., Reeves, L., Ross, M., and Thornthwhite, W. *The Data Warehouse Lifecycle Toolkit*. Wiley (1998).
- [13] Leavitt, N. Data Mining for the Corporate Masses. *IEEE Computer*. 35 (5), (2002), 22-24.
- [14] S. Sarawagi, S. Thomas, R. Agrawal. Integrating Mining with Relational Database Systems: Alternatives and Implications. *Proceedings of ACM SIGMOD Conference* (1998), 343-354
- [15] S. Tsur, J. Ullman, S. Abiteboul, C. Clifton, R. Motwani, S. Nestorov, A. Rosenthal. Query Flocks: A Generalization of Association-Rule Mining. *Proceedings of ACM SIGMOD Conference*, (1998), 1-12.
- [16] Wang, K., He Y., and Han J. Mining Frequent Itemsets Using Support Constraints. *Proceedings of International Conference on Very Large Databases VLDB*, (2000), 43-52
- [17] Watson, H. J., Annino, D. A., and Wixom, B. H. Current Practices in Data Warehousing. *Information Systems Management*. 18 (1), (2001).