Review on Data Warehouse: As Prerequisite Aspect of Business Decision Making Activity

Prof. Kruti B Koshiya
Assistant Professor
Department of Computer Engineering
Vadodara Institute of Engineering, Kotambi, India

Abstract— This paper describes the technology of data warehouse in decision making. Data warehousing and on-line analytical processing (OLAP) are prerequisite aspects of decision support, which has increasingly become a focus of the database industry. The construction of data warehouses involves data cleaning, data integration, data transformation and as important pre-processing step for data mining. The data warehouse supports on-line analytical processing (OLAP) which has the functional and performance requirements. Data warehousing have evolved as one of primary technologies that facilitate data storage, organization and, denoting retrieval. Different requirements on database technology compared to traditional on-line transaction processing applications.

Key words: Data Warehousing, OLAP, Data Mining, Decision-Making and Decision Support

I. INTRODUCTION

Data warehousing provides architectures and tools for business executives to consistently organize, understand, and use their knowledge to form a strategic decisions. Data warehouse systems are the valuable tools in today's fast-evolving world. Data warehouse systems allow the combination of a range of application systems. They support information science by providing a solid platform of consolidated knowledge for analysis. "A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile assortment of knowledge in support management's decision making process". Data warehouse is an assortment of decision support technologies, geared toward enabling the knowledge worker (executive, manager, and analyst) to form higher and quicker decisions. It is a physical implementation of a decision support knowledge model and stores the knowledge on that an enterprise has to build strategic decisions. The data will be keep in many alternative sorts of databases. One database architecture that has recently emerged is the "data warehouse", a repository of multiple heterogeneous knowledge sources, organized underneath a unified schema at a single website so as to facilitate management decision-making. Data warehouse technology includes knowledge cleansing, data integration and on-line Analytical processing.

II. DATA WAREHOUSING

A. Definition of Data Warehousing

According to W.H.Inmon, a number one designer within the construction of knowledge warehouse systems, a data warehouse is a subject-oriented, integrated, time-variant and non-volatile assortment of knowledge in support of management's deciding making process. So, data warehouse are often same to be a semantically consistent knowledge store that is a physical implementation of a decision support data model associated stores the knowledge on that an enterprise must create strategic selections. So, its design is claimed to be made by integration data from multiple heterogeneous sources to support and /or adhoc queries, analytical news and decision-making. Data warehouses give on-line analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities that facilitates effective data processing. The useful and performance needs of OLAP area unit quite completely different from those on-line transaction processing applications traditionally supported by the operational databases. Data can currently be keep in many various forms of databases. One form of database design that has recently emerged is data warehouse that is a repository of multiple heterogeneous information sources, organized underneath a unified schema at one web site so as to facilitate management decision-making. A data warehouse is outlined as a "subject-oriented, integrated, time variant, non-volatile assortment of data that serves as a physical implementation of a decision support information model associated stores the data on that an enterprise must create strategic selections. Data warehousing is an assortment of decision support technologies, aimed toward enabling the data employee (executive, manager, and analyst) to create higher and quicker selections. Data warehousing technologies are successfully deployed in several industries: producing (for order cargo and client support), retail (for user profiling and inventory management), financial services (for claims analysis, risk analysis, MasterCard analysis, and fraud detection), transportation (for fleet management), telecommunications (for decision analysis and fraud detection), utilities (for power usage analysis), and care (for outcomes analysis). This paper presents a roadmap of information warehousing technologies, specializing in the special necessities that information warehouses place on database management systems (DBMSs).

B. Data Warehousing Fundamentals

A data warehouse (or smaller -scale data mart) is a specially ready repository of information designed to support decision making. The information comes from operational systems and external sources. To make the information warehouse, knowledge are extracted from source systems, cleaned (e.g., to detect and correct errors), remodeled (e.g., put into subject teams or summarized), and loaded into a knowledge store (i.e., placed into a knowledge warehouse). The data in a very data warehouse have the subsequent characteristics:

- Subject oriented — The data are logically organized around major subjects of the organization, e.g., around customers, sales, or items produced.

All rights reserved by www.ijsrd.com

- Integrated — All of the data about the subject are combined and can be analyzed together.
- Time variant — Historical data are maintained in detail form.
- Nonvolatile — the data are read only, not updated or changed by users.

C. Architecture and End-to-End Process

Figure 1 shows a typical data warehousing architecture.

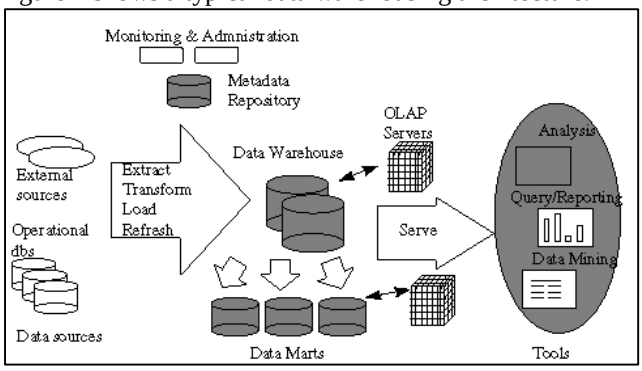

Fig. 1: Data Warehousing Architecture

It includes tools for extracting knowledge from multiple operational databases and external sources; for cleanup, transforming and integrating this knowledge; for loading data into the data warehouse; and for sporadically refreshing the warehouse to replicate updates at the sources and to purge knowledge from the warehouse, maybe onto slower depository storage. Additionally to the most warehouse, there could also be many division data marts. Data within the warehouse and data marts is hold on and managed by one or a lot of warehouse servers that present multidimensional views of information to a spread of front tools: query tools, report writers, analysis tools, and data processing tools. Finally, there is a repository for storing and managing metadata, and tools for observance and administering the warehousing system.

III. Decision Making Using a Data Warehouse

A Decision Support System (DSS) is any tool used to improve the method of deciding in complicated systems. A DSS will vary from a system that answer easy queries and permit an ensuant decision to be created, to a system that use AI and provides elaborated querying across a spectrum of connected datasets. Amongst the most important application area unites of DSS are those difficult systems that directly "answer" queries, in particular high-level "what-if "scenario modeling. Over the last decade there was a transition to decision supporting using knowledge warehouses .The data warehouse surroundings is a lot of controlled and thus a lot of reliable for decision support than previous strategies. The data warehouse environment supports the whole decision support necessities by providing high-quality data, created accessible by correct and effective cleanup routines and using consistent and valid information transformation rules and document pre-summarization of information values. It contains one single supply of correct, reliable data which will be used for analysis. Prediction is a large topic and runs from predicting the failure of additives or instrumentation, to distinctive fraud and even the prediction of leader financial gain. By means that of analyzing on the far side activities or times, you'll build a prediction or so an occasion. The combination of all the data warehouse viewpoints is depicted in Figure 3.

A. Usage Methodology

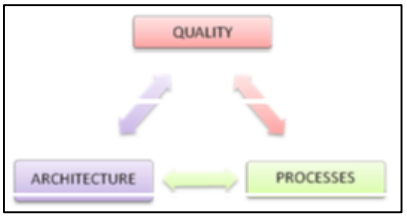

Fig. 3: The different viewpoints for the metadata repository of a data warehouse.

In a basic Meta model for information warehouse architecture and quality has been conferred as in Fig. 3. The framework describes an information warehouse in 3 perspectives: a conceptual, a logical and a physical perspective. Every perspective is divided into the 3 traditional information warehouse levels: source, data warehouse and client level. All the data warehouse viewpoints is depicted in Figure 3.

IV. Steps for Designing Data Warehouse

Designing a data warehouse is a complex process, which consists of following activities:

- Define the architecture, do capacity planning and select the storage servers, database and OLAP servers, and tools.
- Integrate the servers, storage and client tools.
- Design the warehouse schema and views.
- Define the physical warehouse organization, data placement, and partitioning and access methods.
- Connect the sources using gateways, ODBC drivers or other wrappers.
- Design and implement scripts for data extraction, cleaning, transformation, load and refresh.
- Populate the repository with the schema and view definitions, scripts, and other metadata.
- Design and implement end-user applications.
- Roll out the warehouse and applications

V. Data Warehouse Models

There square measure three knowledge warehouse models, per design purpose of view:

- Enterprise warehouse
- Data mart
- Virtual warehouse

A. Enterprise Warehouse

- Collects all of the data concerning subjects spanning the complete organization.
- Provides corporate wide information integration, typically from one or a lot of operational systems or external info suppliers, and is cross-functional in scope.
- Typically contains careful information yet as summarized in formation, and might home in size from many gigabytes to terabytes or on the far side.
- May be enforced on ancient mainframes, UNIX system super servers, or paralleled design platforms.

All rights reserved by www.ijsrd.com

B. Data Mart

- Contains a set of corporate-wide knowledge that's important to a particular cluster of users, however, scope is confined to specific chosen subjects.
- Square measure sometimes enforced on inexpensive division servers that UNIX or windows/NT –based.
- Square measure classified as freelance or dependent, counting on the supply of information operational systems or external data suppliers, or from knowledge generated domestically at intervals a specific department. But, dependent knowledge marts square measure sourced directly from enterprise knowledge warehouse.
- The knowledge contained in data outlet tend to be summarized.

C. Virtual Warehouse

- Is a group of views over operational databases?
- Only a number of the doable outline views is also materialized for economical question process.
- Is straightforward to create however needs excess capability on operational database servers.

D. Why OLAP in Data Warehouse

Simply told, an {information} warehouse stores military science information that answers "who?" and "what?" questions about past events. Whereas OLAP systems have the power to answer "who?" and "what?" queries, it's their ability to answer "what if?" and "why?" that sets them except information warehouses.

- OLAP permits deciding concerning future actions. In distinction to knowledge warehouse, this can be sometimes based on relative technology. OLAP uses a dimensional read of mixture knowledge to supply quick access to strategic info for additional analysis.
- OLAP and knowledge warehouses area unit complementary. A knowledge warehouse manages and stores data. OLAP transforms knowledge warehouse "data" into "strategic information". It ranges from basic navigation and browsing (often referred to as 'slice and dice') to calculations, to a lot of serious analysis like time series and complicated modeling.

VI. Conclusion

Data warehouse is aforesaid to be a semantically consistent knowledge store that is a physical implementation of a decision support knowledge model associated stores the knowledge on that an enterprise must build strategic decisions. So, its design is alleged to be made by integration knowledge from multiple heterogeneous sources to support adhoc queries, analytical reportage and decision-making. Data warehouses give on-line analytical process (OLAP) tools for the interactive analysis of multidimensional knowledge of assorted granularities that facilitates effective data mining.

Data warehousing and on-line analytical processing (OLAP) are a unit essential components of decision support, which has progressively become attention of the database industry. More management and timely access to strategic data facilitates effective decision making. This provides leverage to library managers by providing the power to model real world projections and a more economical use of resources. OLAP permits the organization as a full to reply a lot of quickly to promote demands. Market responsiveness, in turn, typically yields improved revenue and profitableness. And there's no got to emphasize that present libraries got to offer market-oriented services.

References

[1] Han, J & Kamber, M 2001, Data Mining Concepts and Techniques, Morgan Kaufmann.
[2] Inmon, WH 2002, Building the Data Warehouse, 3rd Edition, Wiley.
[3] www.wikipedia.org

All rights reserved by *www.ijsrd.com*