Data Warehouse and Its Trends Driving the Revolution

***Prof. Renuka Sabapathi**
**** Prof.Pradnya Pawar**

Abstract

This paper exhibits information related to data warehouse and its trends. Data warehouse is one of the hot topics in computer technology application. It's 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.The term "data warehouse" itself might still endure, the demands of today's always-on global economy have shattered that old-fashioned metaphor of a limited-access and static warehouse and have inspired leading companies to think instead of relentlessly enriching their data, accelerating it, manipulating it, and ensuring it gets into the hands of as many relevant people as possible. The significance of these changes is that just as manufacturing and logistical prowess were once indispensable qualities of leading businesses, today's battles for competitive advantage will be won by those companies with superior capabilities in business analytics, the democratization of data encompassing not just employees but also partners and customers, and the ability to turn Big Data into Big Opportunities.In the spirit of capturing and describing this data-warehouse revolution, this paper highlights the most noteworthy trends such as *On-demand analytics environments meet the growing demand for rapid prototyping and information discovery ,In-database analytics simplify analysis, Database platforms support mixed workloads.*

Keywords: Data warehouse, Designing models, Trends of data warehouse.

Introduction

A set of significant new concept and tools has evolved in to new technology that addresses the problem of providing all key people with access to information needed for the enterprise to survive and prosper in an increasingly competitive world.

Data warehousing is a field that has grown from the integration of a number of different technologies and experience over the past two decades. Perhaps the most important concept that has come out of the data warehouse movement is the recognition that there are two fundamentally different types of information systems in all organization- operational systems and informational systems. Operational systems are the systems that help the everyday operations of enterprise and were almost always the first part of the enterprise to be computerized. Information systems deal with analyzing data and making a decision about how the enterprise will operate now, and in the future.Data warehouse is subject-oriented, non-volatile, integrated, time variant collection of data, which helps in developing strategic decisions. The popular way of describing data warehouse is through multi-dimensional model. Multidimensional model is based on some central theme which is represented by fact table .A fact table is constructed from multiple dimension tables. The association between these fact table and dimension tables are generally represented through three data warehouse schemas namely star schema, snowflake schema and fact Constellation.

Data Warehouse Architecture

Data warehouse architecture is a way of representing the overall structure of data,

Assistant Professor, IT and CS Department, B.N.Bandodkar College of Science, Thane, Email_id:renuka.prema@gmail.com,pradnyachaudhari5@gmail.com

communication, processing and presentation that exist for end user computing within the enterprise.

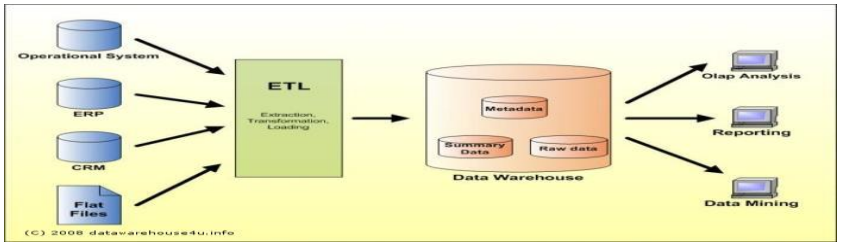


Figure 2.1

.

The **conceptual design** is to have each business unit have its own subject area tailored to their reporting and analytical requirements. The simple diagram above shows the basic design for how source data is extracted, transformed and loaded into the data warehouse on a daily basis through an ETL process. Each subject area is updated with the recently loaded data and made available to the users. Reports that have been scheduled to be distributed and delivered on a predefined schedule are released, and last, users can perform ad-hoc reporting and custom analysis by using the front end tool.

One of the **great functions that a Data Warehouse** provides is to develop reports that are used monthly, weekly, quarterly, and year, No longer it is necessary to continually re-create or re-run these types of standard reports. As a result, individual components can scale and be made highly available with ease.

Key characteristics of a Data warehouse

The key characteristics of a data warehouse are as follows:

- Data is structured for simplicity of access and high-speed query performance.
- End users are time-sensitive and desire speed-of-thought response times.
- Large amounts of historical data are used.
- Queries often retrieve large amounts of data, perhaps many thousands of rows.
- Both predefined and ad hoc queries are common.
- The data load involves multiple sources and transformations.

Database Design

Dimensional model and schemas are most often used in data warehousing systems.

Dimensional Model: In this model, all data is contained in two types of tables called Fact **table** and Dimension Table.

- **Fact Tables:** Fact tables have measurement data. They have many rows but typically not many columns. A fact table usually contains facts with the same level of aggregation.
- **Dimension Tables:** Dimension tables provide category data to give context to the fact data. A key aspect of dimension tables is the hierarchy information they provide. Dimension data typically has rows for the lowest level of detail plus rows for aggregated dimension values.

Schemas: A schema is a collection of database objects, including tables, views, indexes, and synonyms. The most commonly used schema types are Star Schema and Snowflake Schema.

- **Star Schemas**: A star schema has the form of a star where the dimension tables are grouped around the fact table in the center. The goal for star schemas is structural simplicity and high performance data retrieval. Because most queries in the modern era are generated by reporting tools and applications, it's vital to make the query generation convenient and reliable for the tools and application.

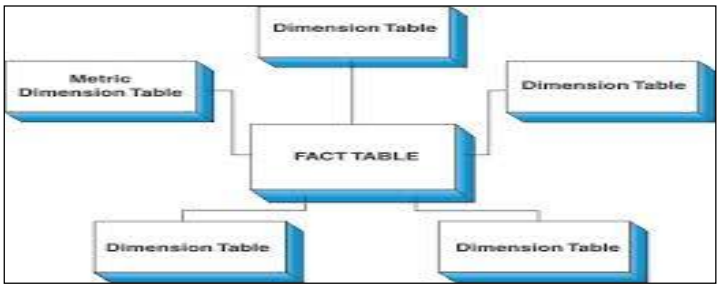


Figure 4.1

- **Snowflake Schema:** The snowflake schema is an extension of the star schema, where each point of the star explodes into more points. In a snowflake schema, that dimensional table is normalized into multiple lookup tables, each representing a level in the dimensional hierarchy. The main advantage of the snowflake schema is the improvement in query performance due to minimized disk storage requirements and joining smaller lookup tables.

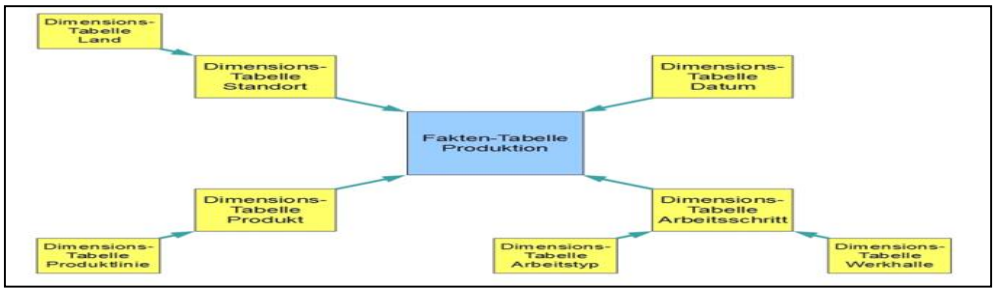
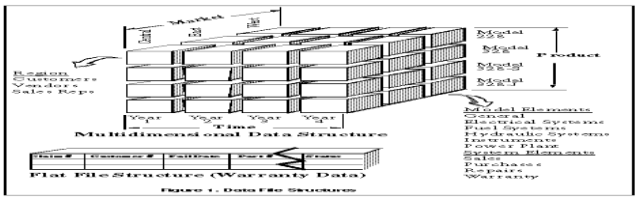


Figure 4.2

- **Multidimensional Data Model:**

The multidimensional data model is an integral part of On-Line Analytical Processing, or OLAP. Multidimensional mode is the default server mode of Analysis Services. It includes a query and calculation engine for OLAP data, with MOLAP, ROLAP, and HOLAP storage modes to balance performance with scalable data requirements. Because OLAP is on-line, it must provide answers quickly; analysts pose iterative queries during interactive sessions. The multidimensional data model is designed to solve complex queries in real time. The central attraction of the dimensional model of a business is its simplicity.... that simplicity is the fundamental key that allows users to understand databases, and allows software to navigate databases efficiently. It is composed of logical cubes, measures, dimensions, hierarchies, levels, and attributes. The simplicity of the model is inherent because it defines objects that represent real-world business entities.



Need of data warehouse:

- **A Data Warehouse Delivers Enhanced Business Intelligence** :

  By providing data from various sources, managers and executives will no longer need to make business decisions based on limited data or their gut. In addition, "data warehouses and related BI can be applied directly to business processes including marketing segmentation, inventory management, financial management, and sales."

- **A Data Warehouse Saves Time :**

  Since business users can quickly access critical data from a number of sources—all in one place—they can rapidly make informed decisions on key initiatives. They won't waste precious time retrieving data from multiple sources. Not only can that but the business execs query the data themselves with little or no support from IT—saving more time and more money. That means the business users won't have to wait until it gets around to generating the reports, and those hardworking folks in it can do what they do best—keep the business running.

- **A Data Warehouse Enhances Data Quality and Consistency :**

  A data warehouse implementation includes the conversion of data from numerous source systems into a common format. Since each data from the various departments is standardized, each department will produce results that are in line with all the other departments. So you can have more confidence in the accuracy of your data and accurate data is the basis for strong business decisions.

- **A Data Warehouse Provides Historical Intelligence :**

  A data warehouse stores large amounts of historical data so you can analyze different time periods and trends in order to make future predictions. Such data typically cannot be stored in a transactional database or used to generate reports from a transactional system.

Latest Trends in Data Warehousing

Most enterprises are overwhelmed with data. Information is now being generated much more quickly than it can be consumed. Meanwhile, leveraging information to make complete business decision is becoming ever more critical to enterprises. All these factors increase the demand for comprehensive information management solutions.

What are the most compelling developments in data warehousing that motivate IT leaders to undertake new initiatives? This brief highlights the most noteworthy trends in data warehouse technology.

- **The "datafication" of the enterprise spawns more-capable data warehouses:**
  Data warehouses were historically populated with structured business data from enterprise applications. This data can be collected not only from computers, but also from billions of mobile phones, tens of billions of social media posts and many other sources. With the help of new big data technologies, data warehouses are expanding in variety and scope. This improves the quality and speed of business decision-making as people learn how to acquire, organize, and analyze this massive influx of information

- **On-demand analytics environments meet rising demand for rapid prototyping and information discovery** :

Business intelligence and analytics are resource-intensive activities that lend themselves to on-demand computing due to their iterative nature and fluctuating workloads. These versatile "sandbox "environments can flex with a shifting volume and velocity of data, making them ideal for analyzing energy usage, monitoring shop floor operations, gauging consumer sentiment, and undertaking many other large-scale analytics challenges. Cutting-edge technologies, such as multitenant databases, enable organizations to set up one cloud environment with dozens or even hundreds of "pluggable "databases for people to use.

- **In-database analytics simplify analysis:**

  In-database analytics is a scheme for processing data within the database, avoiding the data movement that slows response time. It includes a variety of techniques for finding patterns and relationships in large volumes of data. Popular in-database analytics capabilities include data mining algorithms implemented in the database, native SQL functions for basic statistical activities and integration with statistical programming languages like R. Because these techniques are applied directly within the database, analysts can eliminate data movement to and from servers, which accelerates cycle times and reduces total cost of ownership. Organizations can more easily tackle big data projects by minimizing data movement and ensuring better security, scalability, and performance

- **Hadoop optimizes data warehousing environments by accelerating data transformation:**

  The process of converting data from one format to another format is data transformation. Organizations that must contend with big data are increasingly using Hadoop—an open source software that enables organizations to process huge amounts of data across inexpensive servers and storage devices—to augment their traditional data warehouse assets. Hadoop present a cost-effective platform to manage large volumes of unstructured data, storing and transforming any kind of data before pushing it into a data warehouse for analysis. It is also flexible enough to accommodate machine data, documents, and social media and perform the high-volume operations by offloading resource intensive activities to a dedicated, low-cost transformation platform.

- **Engineered systems become the standard for large-scale information management activities:**

  Fast-moving organizations embrace engineered systems for their data warehouse initiatives because they provide a simpler, cheaper, more flexible way to get the job done. These organizations work with one vendor to order complete, preconfigured systems. Engineered systems ensure optimum performance, with software and hardware working in harmony to create a cohesive database platform. They are especially popular for data warehouse solutions because they perform better and are fully integrated and simpler to maintain.

- **In-memory technologies supercharge data warehouse performance:**

  All warehouse data used to be stored on magnetic disks. Now that data is being moved into RAM to achieve performance improvements that are orders of magnitude faster than previous methods. Data warehouse administrators can configure these environments to optimize data among RAM, Flash, and disk-based access methods based on heuristic access patterns. This lightning-fast database processing can entirely eliminate the need to create analytics indexes.

- **Database platforms support mixed workloads:**

  Engineered systems that support mixed workloads bring data warehouses and operational data together to simplify reporting and analysis. Operational system which helps to process data to

support critical operation needs. In order to do this, operational databases have been historically created to provide an efficient processing structure for a relatively small number of well defined business transactions. The database design is done to support operational systems like management or informational systems where critical data is getting access. This in turn means that the data access technology available to obtain operational data and clearly indicates that data ware housing is free for the information which is locked in operational system and can easily combine it with information from other often external source of data .Increasingly large organizations are acquiring additional data from outside database. This information includes demographic, econometric, competitive and purchasing trends. It is very essential to identify the right data source and determine an efficient process to collect the fact.

7. Summary and Conclusion

Data warehouse is an important database information technology that many companies implement as an information system component. In this paper, we make a study of different approaches used for data warehouse Design on following criteria: Framework/Architecture, key characteristics, need of data warehouse, star and snowflake schemas which are more efficient for data warehouse design as they are easy to learn. This paper also focused on latest trends on data warehouse which discuss best practices that will help the organization deal with data deluge. These trends describe a new generation of data warehouses that are bigger, better, and faster than ever before, transforming data into information and information into actionable insights, enabling businesses to forge ahead with unprecedented speed and agility.

References

- *. Data Warehousing – Design Development and Best Practices –SOUMENDRA MAHANTY*
- *Data Warehousing: Using the Wal-Mart Model – In Data Management system –MORGAN KAUFMANN*
- *Mastering Data warehouse Design – Relational and Dimensional techniques- CLAUDIA IMHOFF, NICHOLAS GALEMMO*