A Study on Data Warehouse – An Emerging Technology

Prof S L Patil

Kle Societys J G College of Commerce Hubballi

Abstract—**A digital storage system that links and unifies vast volumes of data from numerous sources is called a data warehouse (DW). In order for businesses to transform their data into insight and make informed decisions, it serves to support regulatory needs as well as business intelligence (BI), reporting, and analytics. Data warehouses serve as an organization's single source of truth by centrally storing both historical and current data.**

Data flows into a data warehouse from operational systems (like ERP and CRM), databases, and external sources such as partner systems, Internet of Things (IoT) devices, weather apps, and social media – usually on a regular cadence. The emergence of cloud computing has caused a shift in the landscape. In recent years, data storage locations have moved away from traditional on-premises infrastructure to multiple locations, including in private and public clouds.

Both organized and unstructured data, such as images, videos, and sensor data, can be handled by modern data warehouses. Some leverage integrated analytics and in-memory database technology (which holds the data set in computer memory rather than in disk storage) to provide real-time access to trusted data and drive confident decision-making. Without data warehousing, it's very difficult to combine data from heterogeneous sources, ensure it's in the right format for analytics, and get both a current and long-range view of data over time.

Index Terms—**Meaning, Characteristics, Data Warehouse Architecture, Types, etc.**

I. INTRODUCTION

Organizations are using cloud-based technology for easy data collecting, reporting, and analysis in the quickly evolving business world of today. This is where data warehousing, a crucial part of business intelligence that helps companies perform better, comes in. It is important to understand what data warehouse is and why it is evolving in the global marketplace.

An overview of data warehouses will be given in this article, along with an exploration of important ideas such as data warehouse architecture, data warehouse characteristics, data management, data warehouse advantages, and data warehouse applications in data science.

What Is a Data Warehouse

In order to make better decisions, data warehouses act as a central location for information analysis and storage. Transactional systems, relational databases, and other sources are among the many sources of data that are regularly added to an organization's data warehouse. A data warehouse is a type of centralized storage system that makes it possible to store, analyze, and interpret data to help with decision-making. Data warehouses frequently receive data from relational databases, transactional systems, and other sources.

A data warehouse is a type of data management system that facilitates and supports business intelligence (BI) activities, specifically analysis. Data warehouses are primarily designed to facilitate searches and analyses and usually contain large amounts of historical data. A collection of organizational data and information retrieved from external and operational data sources is known as a data warehouse. Data is periodically extracted from client interface apps, external partner systems, and a variety of internal applications, including marketing, finance, and sales. Decision-makers are then able to access and examine this data. What is a data warehouse, then? To begin with, it is an extensive collection of historical and present data intended to improve an organization's performance.

Essential Features of a Data Warehouse

➢ Focused on the subject: Since a data warehouse offers information based on topics rather than a company's overall operations, it is subject-oriented. These topics could include inventory, sales, promotions, etc. For instance, you must create a data warehouse focused on sales if you wish to examine the sales data for your business.

Important data like "who was your best customer last year?" or "who is likely to be your best customer in the coming year?" could be found in such a warehouse.

➤ Combined: Data from multiple sources is combined into a standardized format to create a data warehouse. In terms of title, structure, and coding, the data must be kept in the warehouse in a uniform and widely recognized way. This makes data analysis more efficient.

➤ Not Changing: After being entered into a data warehouse, data must not be altered. Every bit of data is read-only. Entering current data does not remove previous data. This aids in your analysis of what transpired and when.

➤ Variable in Time: Time is either explicitly or implicitly noted in the data that is kept in a data warehouse. The Primary Key, which needs to contain a time element such as the day, week, or month, is an illustration of temporal variance in a data warehouse.

II. ARCHITECTURE OF DATA WAREHOUSES

➤ The lowest tier: A relational database system is typically represented by the data warehouse server or bottom tier. Data is cleaned, transformed, and sent into this layer using back-end techniques.

➤ The Middle Level: The middle tier represents an OLAP server that can be implemented in two ways. The ROLAP or Relational OLAP model is an extended relational database management system that maps multidimensional data process to standard relational process. The MOLAP or multidimensional OLAP directly acts on multidimensional data and operations.

➤ Highest Level: This is the front-end client interface that gets data out from the data warehouse. It holds various tools like query tools, analysis tools, reporting tools, and data mining tools.

III. THE OPERATION OF A DATA WAREHOUSE

Information and data gathered from multiple sources are combined into a single, extensive database through data warehousing. A data warehouse might, for instance, compile client data from a company's website, comment cards, mailing lists, and point-of-sale systems. Additionally, it may include pay information, private employee data, etc. These data warehouse components are used by businesses to examine their clientele. One of the capabilities of a data warehouse is data mining, which is searching through enormous amounts of data for significant trends and coming up with creative ways to boost sales and profits.

IV. DATA WAREHOUSE TYPES

➤ Warehouse for Enterprise Data (WED): This kind of warehouse supports decision-support services across the organization by acting as a key or central database. Access to cross-organizational data, a uniform approach to data representation, and the ability to conduct sophisticated queries are the benefits of this kind of warehouse.

➤ Operational Data Store (ODS): This kind of warehouse acts as a central database or key that makes decision-support services available across the entire company. This kind of warehouse has the advantages of allowing complicated queries to be executed, offering a consistent approach to data representation, and facilitating access to cross-organizational information.

➤ The Data Mart: A subset of a data warehouse designed to support a certain division, area, or business unit is called a data mart. Data is stored in a central repository, often known as a data mart, in each business department. Periodically, the ODS stores the data from the data mart. The EDW stores and uses the data after receiving it from the ODS.

Advantages of Data Warehouses
- Enhanced consistency of the data
- Improved business choices
- End consumers' easier access to company data
- Improved data documentation
- Better data documentation
- gathering relevant information from multiple sources and storing it in one location

Key Components of Data Warehousing
A central database, ETL (extract, transform, load) tools, metadata, and access tools are the four primary parts of a typical data warehouse. These parts are all

designed to be fast so that you may analyze data instantly and receive results.
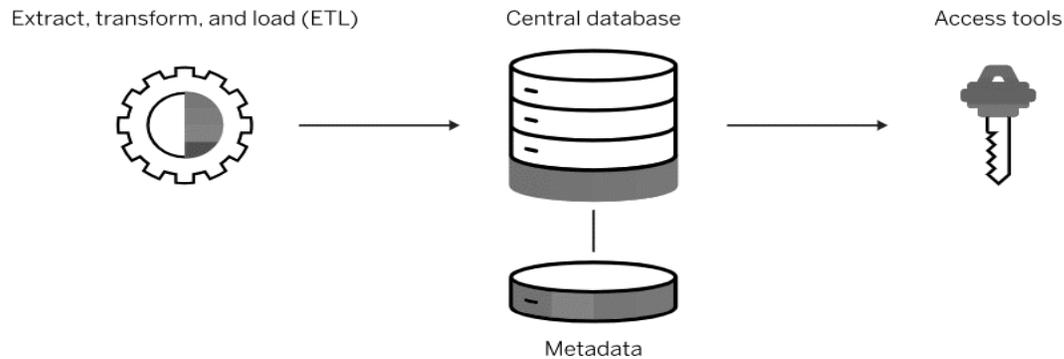
- The Central Database: Your data warehouse is built on top of a database. These have typically been conventional relational databases that operate on-site or in the cloud. But because of Big Data, the need for true, real-time performance, and a drastic reduction in the cost of RAM, in-memory databases are rapidly gaining in popularity.
- Data integration: Data integration: Data is pulled from source systems and modified to align the information for rapid analytical consumption using a variety of data integration approaches such as ETL (extract, transform, load) and ELT as well as real-time data replication, bulk-load processing, data transformation, and data quality and enrichment services.
- Metadata: Data about your data is called metadata. It details the data sets in your data warehouse, including their source, usage, values, and other characteristics. Technical metadata explains how to retrieve data, including where it is stored and how it is organized, while business metadata gives your data context.
- Data warehouse access tools: Users can interact with the data in your data warehouse by using access tools. Data mining tools, application development tools, query and reporting tools, and OLAP tools are a few types of access tools.

Best practices in data warehousing

You can save time and money by following tried-and-true methods when building a new data warehouse or adding new applications to an existing warehouse.

Other practices are a component of your overall IT program, while others are exclusive to your company use. You will learn more best practices as you collaborate with your technology and services partners, but the following list is an excellent place to start.

Best practices for businesses
- Specify what information you need: You can locate the data sources to back up your initial needs after you have a clear idea of them. Customers, suppliers, and trade associations will frequently have data recommendations for you.
- Document the location, structure, and quality of your current data. Then, you can identify data gaps and business rules for transforming the data to meet your warehouse requirements.
- Assemble a group. Executive sponsors, supervisors, and employees who will use and supply the information are all included in this. Determine the KPIs and standard reporting, for instance, that they require in order to do their duties.
- Set your data warehouse apps in order of priority. Select one or two pilot initiatives that have strong business value and realistic needs.
- Select a reliable technology partner for data warehouses. They must possess the experience and implementation services required for your projects. Verify that they can accommodate your deployment requirements.

- Create an effective project plan. Together with your team, create a practical plan and timeline that facilitates status updates and communication.

Best practices for IT
- Keep an eye on security and performance. Your data warehouse has essential information, but in order for it to benefit the company, it needs to be easily accessible. Keep a close eye on system utilization to guarantee optimal performance.
- Keep an eye on security and performance. Your data warehouse has essential information, but in order for it to benefit the company, it needs to be easily accessible. Keep a close eye on system utilization to guarantee optimal performance.
- Offer a flexible architecture. You will find a variety of data mart and warehouse requirements as your corporate and business unit usage grows. They will be much better served by a flexible platform than by a constrained, limited offering.
- Automate maintenance and other procedures. Machine learning can automate technical management tasks in data warehouses to maintain speed and lower operational costs, in addition to enhancing business intelligence.
- Make strategic use of the cloud. The deployment requirements of departments and business units vary. When necessary, use on-premises solutions; otherwise, take use of cloud data warehouses for scalability, cost savings, and access via phone and tablet.

V. METHODOLOGY

The data collection purely on secondary basis through online, magazines and other publications.