

Data Warehousing



Concept of Data Warehousing

Data warehousing is the process of constructing and using a data warehouse. A data warehouse is constructed by integrating data from multiple heterogeneous sources that support analytical reporting, structured and/or ad hoc queries, and decision making. Data warehousing involves data cleaning, data integration, and data consolidations.

Data warehousing may be defined as a collection of corporate information and data derived from operational systems and external data sources. A data warehouse is designed with the purpose of inducing business decisions by allowing data consolidations, analysis, and reporting at different aggregate levels. Data is populated into the data warehouse by extraction, transformation, and loading.

Data warehousing incorporates data stores and conceptual, logical, and physical models to support business goals and end-user information needs. Creating a data warehouse requires mapping data between sources and targets, then capturing the details of the transformation in a metadata repository. The data warehouse provides a single, comprehensive source of current and historical information.

Functions of Data Warehouse Tools and Utilities

- **Data Extraction:** Involves gathering data from multiple heterogeneous sources.
- **Data Cleaning:** Involves finding and correcting the errors in data.
- **Data Transformation:** Involves converting the data from legacy format to warehouse format.
- **Data Loading:** Involves sorting, summarizing, consolidating, checking integrity, and building indices and partitions.
- **Refreshing:** Involves updating from data sources to warehouse.

Data Warehouse Models

Operational Data Warehouse

The view over an operational data warehouse is known as a virtual warehouse. It is easy to build a virtual warehouse. Building a virtual warehouse requires excess capacity on operational database servers. Hence, it is widely preferred for routine activities like storing records of the employees.

Data Mart

Data mart contains a subset of organization-wide data. It is specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.

Enterprise Data Warehouse

Enterprise Data Warehouse is a centralized warehouse. It provides decision support service across the enterprise. It offers a unified approach for organizing and representing data. It also provides the ability to classify data according to the subject and give access according to those divisions.

Components of Data Warehouse

Load manager

Load manager is also called the front component. It performs with all the operations associated with the extraction and load of data into the warehouse. These operations include transformations to prepare the data for entering into the Data warehouse.

Warehouse Manager

Warehouse manager performs operations associated with the management of the data in the warehouse. It performs operations like analysis of data to ensure consistency, creation of indexes and views, generation of denormalization and aggregations, transformation and merging of source data and archiving and backing-up data.

Query Manager

Query manager is also known as backend component. It performs all the operation operations related to the management of user queries. The operations of these Data warehouse components are direct queries to the appropriate tables for scheduling the execution of queries.

End-user access tools

This is categorized into five different groups like: 1) Data Reporting 2) Query Tools 3) Application development tools 4) EIS tools, 5) OLAP tools and data mining tools.

Who needs Data warehouse?

Data warehouse is needed for all types of users like:

- Decision makers who rely on mass amount of data
- Users who use customized, complex processes to obtain information from multiple data sources.
- It is also used by the people who want simple technology to access the data
- It is also essential for those people who want a systematic approach for making decisions.
- If the user wants fast performance on a huge amount of data which is a necessity for reports, grids or charts, then Data warehouse proves useful.
- Data warehouse is a first step if you want to discover 'hidden patterns' of data-flows and groupings.

Types of Architecture

There are mainly three types of Data warehouse Architectures:

Single-tier architecture

The objective of a single layer is to minimize the amount of data stored. This goal is to remove data redundancy. This architecture is not frequently used in practice.

Two-tier architecture

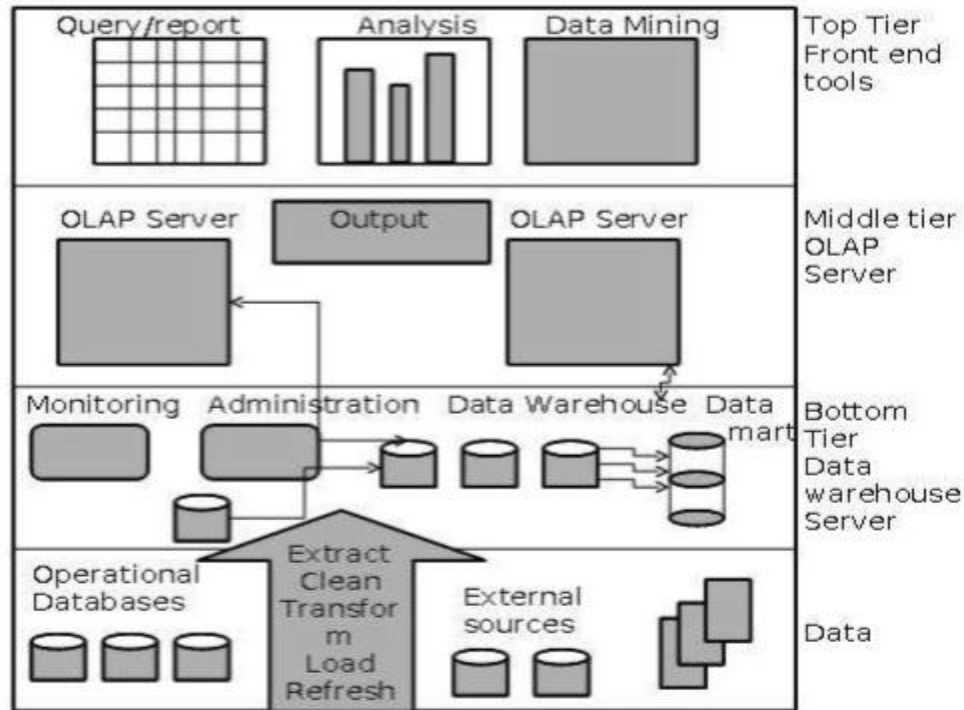
Two-layer architecture separates physically available sources and data warehouse. This architecture is not expandable and also not supporting a large number of end-users. It also has connectivity problems because of network limitations.

Three-tier architecture

This is the most widely used architecture. It consists of three layers:

- **Bottom Tier:** The bottom tier of the architecture is the data warehouse database server. It is the relational database system. We use the back end tools and utilities to feed data into the bottom tier. These back end tools and utilities perform the Extract, Clean, Load, and refresh functions.
- **Middle Tier:** The middle tier in data warehouse is an OLAP server which is implemented using either ROLAP or MOLAP model. For a user, this application tier presents an abstracted view of the database. This layer also acts as a mediator between the end-user and that database.

- **Top-Tier:** This tier is the front-end client layer. Top tier is the tools and API that you connect and get data out from the data warehouse. This layer holds the query tools and reporting tools, analysis tools and data mining tools.



Advantages of Data Warehousing

- Data warehouse allows business users to quickly access critical data from some sources all in one place.
- Data warehouse provides consistent information on various cross-functional activities. It is also supporting ad-hoc reporting and query.
- Data Warehouse helps to integrate many sources of data to reduce stress on the production stem.
- Data warehouse helps to reduce total turnaround time for analysis and reporting.
- Restructuring and Integration make it easier for the user to use for reporting and analysis.
- Data warehouse allows users to access critical data from the number of sources in a single place. Therefore, it saves user's time of retrieving data from multiple sources.
- Data warehouse stores a large amount of historical data. This helps users to analyze different time periods and trends to make future predictions.

Disadvantages of Data Warehousing

- Not an ideal option for unstructured data.
- Creation and Implementation of Data Warehouse is surely time consuming affair.
- Data Warehouse can be outdated relatively quickly
- Difficult to make changes in data types and ranges, data source schema, indexes, and queries.
- The data warehouse may seem easy, but actually, it is too complex for the average users.
- Despite best efforts at project management, data warehousing project scope will always increase.
- Sometime warehouse users will develop different business rules.
- Organizations need to spend lots of their resources for training and Implementation purpose.

Applications of Data Warehousing

Banking Industry

In the banking industry, concentration is given to risk management and policy reversal as well analyzing consumer data, market trends, government regulations and reports, and more importantly financial decision making.

Most banks also use warehouses to manage the resources available on deck in an effective manner. Certain banking sectors utilize them for market research, performance analysis of each product, interchange and exchange rates, and to develop marketing programs.

Analysis of card holder's transactions, spending patterns and merchant classification, all of which provide the bank with an opportunity to introduce special offers and lucrative deals based on cardholder activity. Apart from all these, there is also scope for co-branding.

Finance Industry

Similar to the applications seen in banking, mainly revolve around evaluation and trends of customer expenses which aids in maximizing the profits earned by their clients.

Consumer Goods Industry

They are used for prediction of consumer trends, inventory management, market and advertising research. In-depth analysis of sales and production is also carried out. Apart from these, information is exchanged business partners and clientele.

Government and Education

The federal government utilizes the warehouses for research in compliance, whereas the state government uses it for services related to human resources like recruitment, and accounting like payroll management.

The government uses data warehouses to maintain and analyze tax records, health policy records and their respective providers, and also their entire criminal law database is connected to the state's data warehouse. Criminal activity is predicted from the patterns and trends, results of the analysis of historical data associated with past criminals.

Universities use warehouses for extracting of information used for the proposal of research grants, understanding their student demographics, and human resource management. The entire financial department of most universities depends on data warehouses, inclusive of the Financial Aid department.

Healthcare

One of the most important sectors which utilize data warehouses is the Healthcare sector. All of their financial, clinical, and employee records are fed to warehouses as it helps them to strategize and predict outcomes, track and analyze their service feedback, generate patient reports, share data with tie-in insurance companies, medical aid services, etc.

Hospitality Industry

A major proportion of this industry is dominated by hotel and restaurant services, car rental services, and holiday home services. They utilize warehouse services to design and evaluate their advertising and promotion campaigns where they target customers based on their feedback and travel patterns.

Insurance

As the saying goes in the insurance services sector, "Insurance can never be bought, it can be only be sold", the warehouses are primarily used to analyze data patterns and customer trends, apart from maintaining records of already existing participants. The design of tailor-made customer offers and promotions is also possible through warehouses.

Manufacturing and Distribution Industry

This industry is one of the most important sources of income for any state. A manufacturing organization has to take several make-or-buy decisions which can influence the future of the sector, which is why they utilize high-end OLAP tools as a part of data warehouses to predict

market changes, analyze current business trends, detect warning conditions, view marketing developments, and ultimately take better decisions.

They also use them for product shipment records, records of product portfolios, identify profitable product lines, analyze previous data and customer feedback to evaluate the weaker product lines and eliminate them.

For the distributions, the supply chain management of products operates through data warehouses.

The Retailers

Retailers serve as middlemen between producers and consumers. It is important for them to maintain records of both the parties to ensure their existence in the market.

They use warehouses to track items, their advertising promotions, and the consumers buying trends. They also analyze sales to determine fast selling and slow selling product lines and determine their shelf space through a process of elimination.

Services Sector

Data warehouses find themselves to be of use in the service sector for maintenance of financial records, revenue patterns, customer profiling, resource management, and human resources.

Telephone Industry

The telephone industry operates over both offline and online data burdening them with a lot of historical data which has to be consolidated and integrated.

Apart from those operations, analysis of fixed assets, analysis of customer's calling patterns for sales representatives to push advertising campaigns, and tracking of customer queries, all require the facilities of a data warehouse.

Transportation Industry

In the transportation industry, data warehouses record customer data enabling traders to experiment with target marketing where the marketing campaigns are designed by keeping customer requirements in mind.

The internal environment of the industry uses them to analyze customer feedback, performance, manage crews on board as well as analyze customer financial reports for pricing strategies.