

2

DATA WAREHOUSE

2.1 What is data Warehouse?

(DEFINATIONS)

- Data warehouse is a central location where consolidated data from multiple locations are stored .
- A data warehouse is the data(meta/fact/dimension/aggregation) and the process managers(query/tool)that makes information available enabling people to make informal decision
- Data warehousing is a technology that aggregates structured data from one or more sources so that it can be compared and analyzed for greater business intelligence.
- A Data Warehouse is a repository of information collected from multiple sources, stored under a unified scheme, and which usually resides a a single site. Data warehouse is constructed via a process of data cleaning, data transformation, data integration, data loading and periodic data refreshing.

2.2 Introduction

Data Warehouse (DW) stores corporate information and data from operational systems and a wide range of other data resources. Data Warehouses are designed to support the decision-making process through data collection, consolidation, analytics, and research. They can be used in analyzing a specific subject area, such as “sales,” and are an important part of modern Business Intelligence.

The architecture for Data Warehouses was developed in the 1980s to assist in transforming data from operational systems to decision-making support systems. Normally, a Data Warehouse is part of a business’s mainframe server or in the Cloud.

In a Data Warehouse, data from many different sources is brought to a single location and then translated into a format the Data Warehouse can process and store. For example, a business stores data about its customer’s information, products, employees and their salaries, sales, and invoices. The boss may ask about the latest cost-reduction measures, and getting answers will require an analysis of all of the previously mentioned data. Unlike basic operational data storage, Data Warehouses contains aggregate historical data (highly useful data taken from a variety of sources).

Punch cards were the first solution for storing computer generated data. By the 1950s, punch cards were an important part of the American government and businesses. The warning “Do not fold, spindle, or mutilate” originally came from punch cards. Punch cards continued to be used regularly until the mid-1980s. They are still used to record the results of voting ballots and standardized tests. “Magnetic storage” slowly replaced punch cards starting in the 1960s. Disk storage came as the next evolutionary step for data storage. Disk storage (hard drives and floppies) started becoming popular in 1964 and allowed data to be accessed directly, which was a significant improvement over the clumsier magnetic tapes. IBM was primarily responsible for the early evolution of disk storage. They invented the floppy disk drive as well as the hard disk drive. They are also credited with several of the improvements now supporting their products. IBM began developing and manufacturing disk storage devices in 1956. In 2003, they sold their “hard disk” business to Hitachi.

The Need for Data Warehouses

During the 1990s major cultural and technological changes were taking place. The internet was surging in popularity. Competition had increased due to new free trade agreements, computerization, globalization, and networking. This new reality required greater business intelligence, resulting in the need for true data warehousing. During this time, the use of application systems exploded.

By the year 2000, many businesses discovered that, with the expansion of databases and application systems, their systems had been badly integrated and that their data was inconsistent. They discovered they were receiving and storing lots of fragmented data. Somehow, the data needed to be integrated to provide the critical “Business Information” needed for decision-making in a competitive, constantly-changing global economy.

Data Warehouses were developed by businesses to consolidate the data they were taking from a variety of databases, and to help support their strategic decision-making efforts

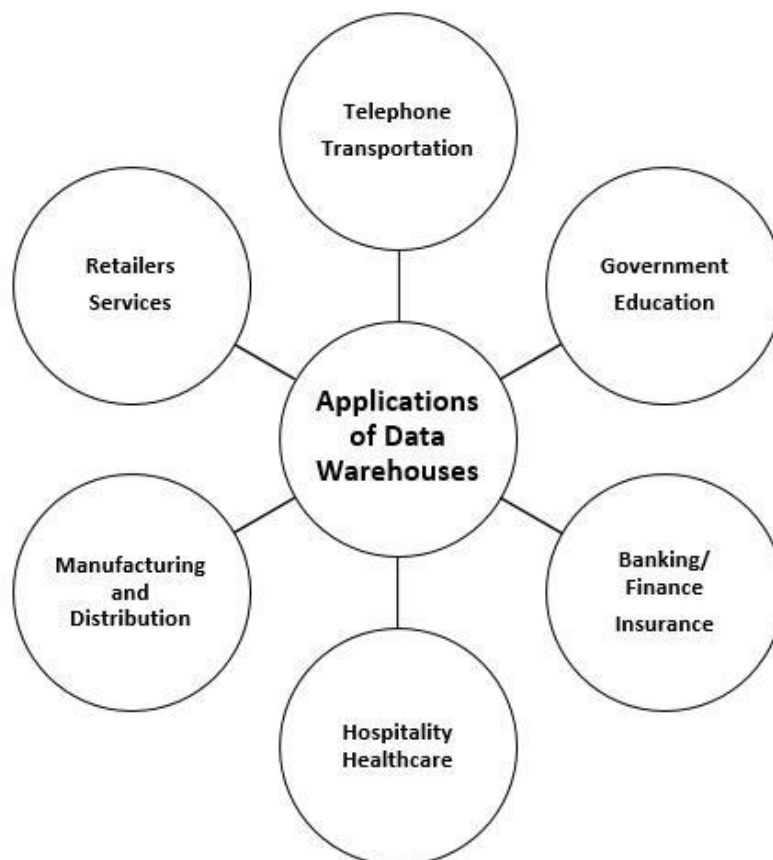
Benefits of Data Warehouse

1. High Return of investment:
 - a. Implementation of data warehousing by an organisation requires a huge investment typically from Rs 10-15 lakhs, however a study by the internal data corporation (IDC) in 1996 reported that avg 3 year returns on investment (ROI) in data warehousing reached 401%.
2. More Cost-effective decision making:
 - a. Data Warehousing helps to reduce the overall cost of product by reducing the number of channels.

3. Competitive advantage:
 - a. The competitive advantage is gained by allowing decision-makers access to data that can reveal previously unavailable, unknown and untapped information. For e.g. Customers, trends, and demand.
4. Increased productivity of corporate decision makers-
 - a. Data warehousing improves the productivity of corporate decision-makers by creating and integrated database of consistent, subject-oriented, and historical data.
5. Better enterprise intelligence:
 - a. It helps to provide better enterprise intelligence.

Application of Data Warehouse

Data Warehouses owing to their potential have deep-rooted applications in every industry which use historical data for prediction, statistical analysis, and decision making. Listed below are the applications of Data warehouses across innumerable industry backgrounds.



1. Banking Industry

In the banking industry, concentration is given to

- Risk management
- Analyzing consumer data, market trends,
- Government regulations and reports,
- 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.

2. 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.

3. 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.

4. Government

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.
- to maintain and analyze tax records,
- analyse health policy records and their respective providers,
- Analyse entire criminal law database . Criminal activity is predicted from the patterns and trends, results of the analysis of historical data associated with past criminals.

5. Education

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.

6. Healthcare

One of the most important sector which utilizes 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 analyse their service feedback,
- generate patient reports,
- share data with tie-in insurance companies,
- Medical aid services, etc.

7. 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.

8. 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.

9. 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
- 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.

10. 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,
- advertising promotions,
- 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.

11. 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.

12. Telephone Industry & Transportation 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, it also

- analysis of customer's calling patterns for sales representatives to push advertising campaigns,
- Tracking of customer queries, all require the facilities of a data warehouse.

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.

Datawarehouse Model

From the architecture point of view, there are three

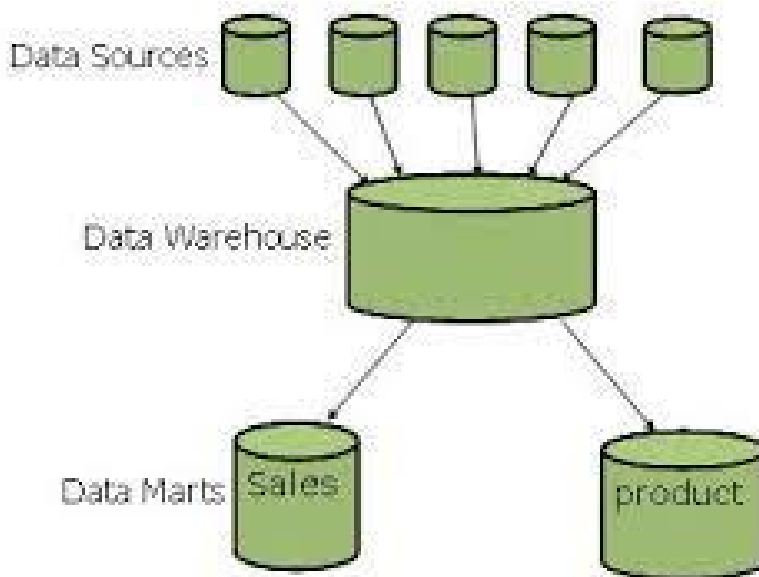
Problems of Data Warehousing

1. Underestimation of resource of data loading.

- a. Sometime we underestimate the time required to extract, clean and load the data into the warehouse.
2. Hidden problem with source systems
 - a. Sometime hidden problems associated with the source system feeding the data warehouse may be identified after years of being undetected.eg. when entering details of the new properties certain field may allow nulls which may result in staff entering incomplete properties data even when available are applicable.
3. Required data not captured
 - a. In some cases the required data is not captured by the source system which may be very important for the data warehouse purpose. e.g. Data of registration for the property may not be used in the source system but it may be very important analysis purpose.
4. Data Homogenization:
 - a. The concept of data warehouse deals with the similarity of data formats between different data source, Thus results in to lose of some important value of data.
5. High demand for resources:
 - a. The data warehouse requires large amount of data
6. High maintenance Cost
 - a. Data warehouse are high maintenance system. Any organisation of the business process and the source system may affect the data warehouse and its results high maintenance cost.
7. Long duration
 - a. The building of warehouse can take up to three years which is why some organisation reluctant in investing in to data warehouse.
8. Data ownership
 - a. Data warehousing may change the attitude of end-user to the ownership of data. Sensitive data that owned by one department has to loaded in data warehouse for decision making purpose. But sometime it result in to reluctance of that department because it may hesitate to share it with other.

Data Mart

- A data mart is a repository of data that is designed to serve a particular community of knowledge workers.
- A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as Sales or Finance or Marketing. Data marts are often built and controlled by a single department within an organization.



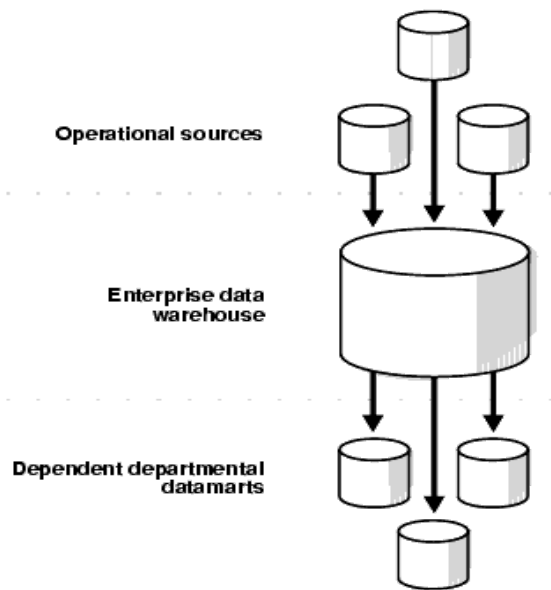
Types of Data Marts

Dependent, Independent, and Hybrid Data Marts

Three basic types of data marts are dependent, independent, and hybrid. The categorization is based primarily on the data source that feeds the data mart. Dependent data marts draw data from a central data warehouse that has already been created. Independent data marts, in contrast, are standalone systems built by drawing data directly from operational or external sources of data or both. Hybrid data marts can draw data from operational systems or data warehouses.

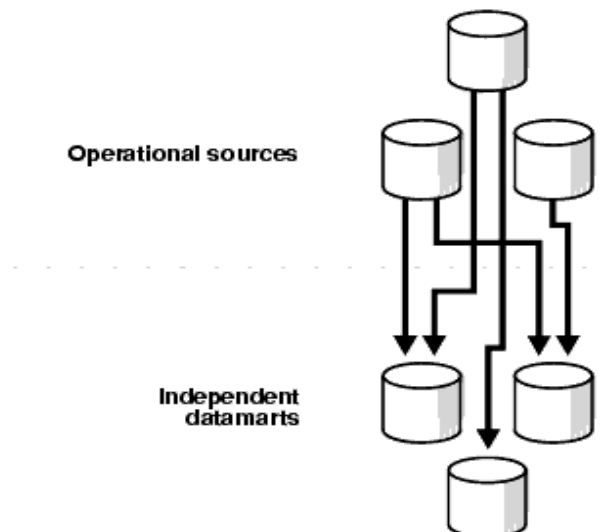
Dependent Data Marts

A dependent data mart allows you to unite your organization's data in one data warehouse. This gives you the usual advantages of centralization. Figure below illustrates a dependent data mart.



Independent Data Marts

An independent data mart is created without the use of a central data warehouse. This could be desirable for smaller groups within an organization. It is not, however, the focus of this Guide. See the Data Mart Suites documentation for further details regarding this architecture. Figure below illustrates an independent data mart.

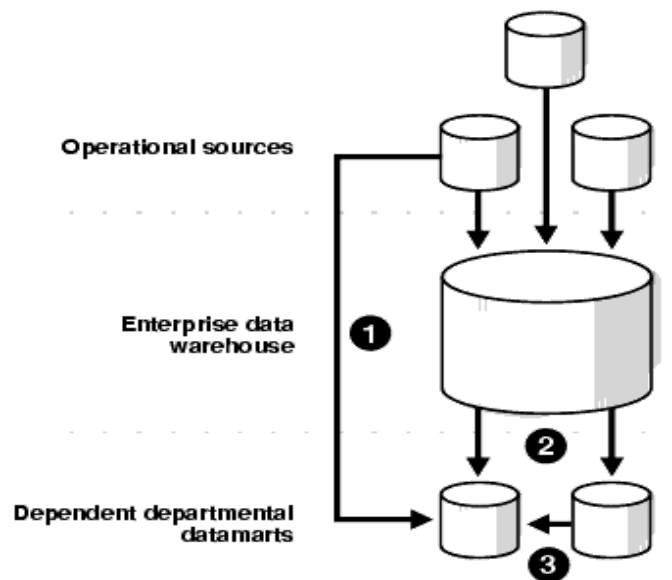


Hybrid Data Marts

A hybrid data mart allows you to combine input from sources other than a data warehouse. This could be useful for many situations, especially when you need ad

hoc integration, such as after a new group or product is added to the organization. Figure below illustrates a hybrid data mart.

Figure Hybrid Data Mart



Difference Between Data Warehouse and Data Mart

	Data Warehouse	Data Mart
Definition	Data Warehouse is a big central repository of historical data	Data Mart can be considered as a subset of data warehouse
Focus	Multiple subject areas	Specific subject area
Control	Central organization unit	Generally, single department

Scope	Corporate	Line of Business
Data Sources	Multiple	Few selected
Size	100 GB-TB+	< 100 GB
Designing	Comparatively difficult	Easy
Implementation	Months to years	Months
Decision	Strategic	Tactical