Data Warehouses and Data Marts

Today, the most successful companies are those that can respond quickly and flexibly to market changes and opportunities. A key to this response is the effective and efficient use of data and information by analysts and managers. The problem is providing users with access to corporate data so that they can analyze the data to make better decisions.

Let's look at an example. If the manager of a local bookstore wanted to know the profit margin on used books at her store, she could obtain that information from her database, using SQL or QBE. However, if she needed to know the trend in the profit margins on used books over the past 10 years, she would have to construct a very complicated SQL or QBE query

This example illustrates several reasons why organizations are building data warehouses and/or data marts. First, the bookstore's databases contain the necessary information to answer the manager's query, but this information is not organized in a way that makes it easy for her to find what she needs. Second, the organization's databases are designed to process millions of transactions every day.

Therefore, complicated queries might take a long time to answer, and they also might degrade the performance of the databases. Third, transactional databases are designed to be updated. This update process requires extra processing. Data warehouses and data marts are read-only, and the extra processing is eliminated because data already contained in the data warehouse are not updated. Fourth, transactional databases are designed to access a single record at a time. Data warehouses are designed to access large groups of related records. As a result of these problems, companies are using a variety of tools with data warehouses and data marts to make it easier and faster for users to access, analyze, and query data.

Describing Data Warehouses and Data Marts

In general, data warehouses and data marts support business intelligence (BI) applications. business intelligence is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions.

A data warehouse is a repository of historical data that are organized by subject to support decision makers in the organization. Because data warehouses are so expensive, they are used primarily by large companies. A data mart is a low-cost, scaled-down version of a data warehouse that is designed for the end-user needs in a strategic business unit (SBU) or an individual department.

Data marts can be implemented more quickly than data warehouses, often in less than 90 days. Further, they support local rather than central control by conferring power on the user group. Typically, groups that need a single or a few BI applications

require only a data mart, rather than a data warehouse.

The basic characteristics of data warehouses and data marts include the following:

• Organized by business dimension or subject. Data are organized by subject—for example, by customer, vendor, product, price level, and region. This arrangement differs from transactional systems, where data are organized by business process, such as order entry, inventory control, and accounts receivable.

• Use online analytical processing. Typically, organizational databases are oriented toward handling transactions. That is, databases use online transaction processing (OLTP), where business transactions are processed online as soon as they occur. The objectives are speed and efficiency, which are critical to a successful Internet-based business operation. Data warehouses and data marts, which are designed to support decision makers but not OLTP, use online analytical processing.

Online analytical processing (OLAP) involves the analysis of accumulated data by end users.

• Integrated. Data are collected from multiple systems and then integrated around subjects. For example, customer data may be extracted from internal (and external) systems and then integrated around a customer identifier, thereby creating a comprehensive view of the customer

• Time variant. Data warehouses and data marts maintain historical data (i.e., data that include time as a variable). Unlike transactional systems, which maintain only recent data (such as for the last day, week, or month), a warehouse or mart may store years of data. Organizations utilize historical data to detect deviations, trends, and long-term relationships.

• Nonvolatile. Data warehouses and data marts are nonvolatile—that is, users cannot change or update the data. Therefore the warehouse or mart reflects history, which, as we just saw, is critical for identifying and analyzing trends. Warehouses and marts are updated, but through IT-controlled load processes rather than by users.

• Multidimensional. Typically the data warehouse or mart uses a multidimensional data structure. Recall that relational databases store data in two-dimensional tables. In contrast, data warehouses and marts store data in more than two dimensions. For this reason, the data are said to be stored in a multidimensional structure. A common representation for this multidimensional structure is the data cube

A Generic Data Warehouse Environment

The environment for data warehouses and marts includes the following:

- Source systems that provide data to the warehouse or mart
- Data-integration technology and processes that prepare the data for use
- Different architectures for storing data in an organization's data warehouse or data marts
- Different tools and applications for the variety of users.
- Metadata, data-quality, and governance processes that ensure that the warehouse or mart meets its purposes

Source Systems.

There is typically some "organizational pain" (i.e., business need) that motivates a fi rm to develop its BI capabilities. Working backward, this pain leads to information requirements, BI applications, and source system data requirements. The data requirements can range from a single source system, as in the case of a data mart, to hundreds of source systems, as in the case of an enterprise wide data warehouse. Modern organizations can select from a variety of source systems: operational/transactional systems, enterprise resource planning (ERP) systems, Web site data, third-party data (e.g., customer demographic data), and more. The trend is to include more types of data (e.g., sensing data from RFID tags). These source systems often use different software packages (e.g., IBM, Oracle) and store data in different formats (e.g., relational, hierarchical)

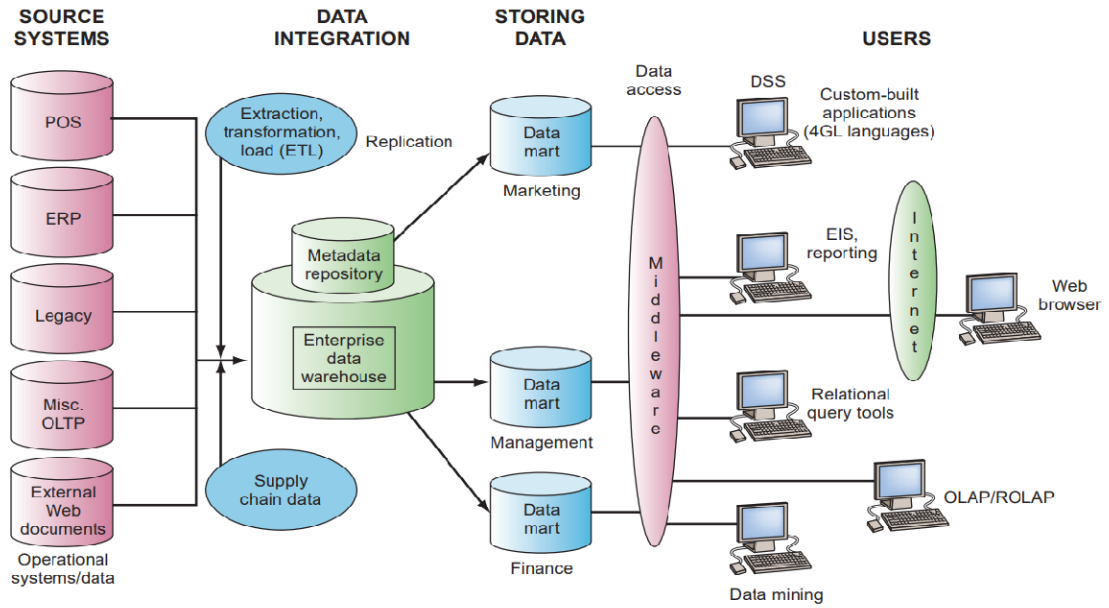

FIGURE 5.9 Data warehouse

- **Data Integration.**
- In addition to storing data in their source systems, organizations need to extract the data, transform them, and then load them into a data mart or warehouse. This process is often called ETL, but the term data integration is increasingly being used to refl ect the growing number of ways that source system data can be handled. For example, in some cases, data are extracted, loaded into a mart or warehouse, and then transformed

- **Data extraction** can be performed either by handwritten code (e.g., SQL queries) or by commercial data-integration software. Most companies employ commercial software. This software makes it relatively easy to specify the tables and attributes in the source systems that are to be used, map and schedule the movement of the data to the target (e.g., a data mart or warehouse), make the required transformations, and ultimately load the data.

- Storing the Data. A variety of architectures can be used to store decision-support data. The most common architecture is one central enterprise data warehouse, without data marts. Most organizations use this approach, because the data stored in the warehouse are accessed by all users and represent the single version of the truth. Another architecture is independent data marts. This architecture stores data for a single application or a few applications, such as marketing and fi nance. Limited thought is given to how the data might be used for other

applications or by other functional areas in the organization. This is a very application-centric approach to storing data.

- **Metadata**. It is important to maintain data about the data, known as metadata, in the data warehouse. Both the IT personnel who operate and manage the data warehouse and the users who access the data need metadata. IT personnel need information about data sources; database, table, and column names; refresh schedules; and data-usage measures. Users' needs include data defi nitions, report/query tools, report distribution information, and contact information for the help desk.

- **Data Quality**. The quality of the data in the warehouse must meet users' needs. If it does not, the data will not be trusted and ultimately will not be used. Most organizations fi nd that the quality of the data in source systems is poor and must be improved before the data can be used in the data warehouse. Some of the data can be improved with data-cleansing software, but the better, long-term solution is to improve the quality at the source system level. This approach requires the business owners of the data to assume responsibility for making any necessary changes to implement this solution

Application areas of data warehouse

- Following are the most common areas where data warehouses are
- extensively used these days:
- **Airline**: In the airline system, it Is used for operation purpose like crew assignment, analysis of route profitability, frequent flyer program promotions, etc.
- **Banking**: It is widely used in the banking sector to manage effectively the resources available on desk. Few banks also use it for the market research, performance analysis of the product and operations.
- Healthcare: Healthcare sector use dat warehouse to make strategies and predict outcomes,treatment reports,share reports
- Public sector:in public sector data warehouse is used for intelligent gathering,it help government agencies to maintain and analyse tax records,health policy

records for every individual

- Advantages of Data Warehouse
- Data warehouse allows business users to quickly access critical data from sources all in one place
- Data warehouses helps to integrate many sourses of data to reduce stress on the production system
- Data warehouse store large amount of historical data . This helps users to analyse different time periods and trends to make future prediction