**Operational system:**

run daily operation of a business( transaction processing systems). Collect various types of data. Converted into meaningful information to produce reports

Process information about a single entity such as single order, single invoice or a single customer

**Online transaction processing(OLPT)**

Transaction-oriented tasks. Provide information to managers.

Operational data capture one transaction at a time, they don’t provide trend over time.

**Decision Support System**

Extract data from various operational systems and store as useful information.

How business runs and make decisions to improve the business.

**Primitive data vs derived data**

Primitive data: day-to-day operation of a business

Derived data: summarized data that meets the need of the managers.

Operational data is primitive, DSS is derived.

**Data warehouse:**

Central repository of information, consisting of current and historical data, provides business intelligence to decision makers.

Is an information environment

Organized around major subjects or business dimensions: customer, item, supplier, activity

Store a summary of the transactions rather than each sales transaction

Architecture: modeled by a multidimensional database structure.

**Online Analytical Processing(OLAP)**

Facilitates summarization and aggregation of data, providing the ability to view information from different angles.

For live, ad hoc data access

Technical basis for the calculations and analysis required by business intelligence applications.

**Data Mart**

Data gathered from operational data and other sources that is designed to serve a particular community of knowledge workers.

Meeting specific demands of a particular group of knowledge users.

**Data Mart and Data Warehouse**

Data mart: analysis of user needs. For particular designed purpose, immediate need.

Data warehouse: analysis of what data already exists and collect for later use. Central aggregation of data. Strategic.

Data Warehouse Environment

Subjected-oriented, integrated, nonvolatile and time variant.

Subject orientation: data are organized by business subjects

Integrated: data comes from many operational systems and outside sources. Data structure are designed by different architects and developed in different platforms.

Nonvolatile: once data is transferred into a data warehouse, they are not typically changed. Data from the operational systems are moved into a data warehouse at specific intervals.

Time-Variant: Data in an operational system stores data for current business operation, but data warehouse in for historical purpose.

Data granularity

Level of details or summarization held in the unit of data.

In an operational system, data is usually kept at the lowest level of detail

Data in a data warehouse are kept in summarized form in various levels. Usually starts with the highest level, then drills down to lower levels.

Source data:

Area that data warehouse gets operational data

Production data: comes from various operational systems

Internal data: kept by individuals in the department level

Archived data: old data is periodically transferred for archival such that the system runs efficiently.

External data: market share data of competitors, statistics relating to the industry

Data staging area

Three major functions need to be performed to get the data ready: data extraction, data transformation and data loading

Data extraction: appropriate techniques used to extract these data and placed in a separate physical environment.

Data Transformation: combining many forms of data. First, extracted data are cleaned, which eliminate duplicates of data. And in common format.

Data loading: A large volume of data is loaded when the design of data warehouse is completed

Data in data warehouse are organized in a dimensional schema, which process large volume of data

Relational schema used for fast transactions.

Information delivery/data access tools: used to query data warehouse and provide necessary information

Metadata: information that provides information about other data.

Guides an analyst in the transformation and loading processes.

OLPT vs OLAP

OLTP:

Customer-oriented and used for transaction and query processing

Manages current data, too detailed.

ER model and an application-oriented database design

Current data

Short, atomic transactions.

OLAP:

Market-oriented, used for analysis by knowledge workers.

Historical data, provides facilities for summarization and aggregation.

Star or snowflake model and a subject-oriented database design.

Historical data that originating from many sources.

Mostly read-only operations.

Architecture types:

Centralized corporate data warehouse: no data marts, all info from centralized data warehouse

Independent data marts:

Data from source systems to staging area, then to the various independent data marts to end-users.

Federated:

Data elements in the various data marts and even data warehouses that compose the federation. All info is from centralized data warehouse.

Hub-and-spoke

Information from both centralized data warehouse and the dependent data marts.

Data-Mart Bus

The collection of all data marts form the data warehouse.

Dimension:

A data structure that categorizes facts and measures for answering business questions. Commonly used dimensions are customers, products, locations and time.

Fact:

A measurement captured from an event in the marketplace

Fact table has its primary key made up of a subset of the foreign keys, called a composite or concatenated key.

Star schema is suitable for query processing.

Data cube is suitable for OLAP processing.

Classic cube is the sale of a product by location by time, 3D cube

Hierarchies in Dimensions

Measures are stored in the fact table in such details that users can roll-up in carious level of summarization called aggregation.

Hierarchy defines a sequence of mappings from a set of low-level concepts to higher-level, more general level concepts.

Slice and Dice:

Slice: Selection on one dimension of the given cube, resulting in a sub cube(eg, data for various cities and items for Q1)

Dice: Performing a selection on two or more dimensions(eg, first include sales for Q1, then on location to include sales of some cities)

Snowflake schema easy to maintain

Star schema easy to use