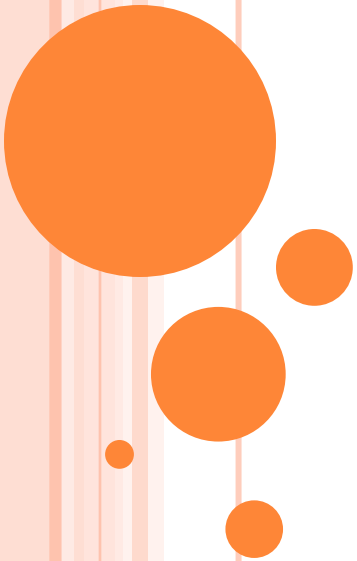


OLTP VS OLAP

OLTP (On-line Transaction Processing)

- is characterized by a large number of short on-line transactions (INSERT, UPDATE, DELETE).
- The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second.
- In OLTP database there is detailed and current data, and schema used to store transactional databases is the entity model (usually 3NF).

OLAP (On-line Analytical Processing)

- is characterized by relatively low volume of transactions.
- Queries are often very complex and involve aggregations.
- For OLAP systems a response time is an effectiveness measure.
- OLAP applications are widely used by Data Mining techniques.
- In OLAP database there is aggregated, historical data, stored in multi-dimensional schemas (usually star schema).

OLTP vs. OLAP

- OLTP: On-Line Transaction Processing
  - Many short transactions (queries + updates)
  - Examples:
    - Update account balance
    - Enroll in course
    - Add book to shopping cart
  - Queries touch small amounts of data (one record or a few records)
  - Updates are frequent
  - Concurrency is biggest performance concern
- OLAP: On-Line Analytical Processing
  - Long transactions, complex queries
  - Examples:
    - Report total sales for each department in each month
    - Identify top-selling books
    - Count classes with fewer than 10 students
  - Queries touch large amounts of data
  - Updates are infrequent
  - Individual queries can require lots of resources

Why OLAP & OLTP don't mix (1)

Different performance requirements

- Transaction processing (OLTP):
  - Fast response time important (< 1 second)
  - Data must be up-to-date, consistent at all times
- Data analysis (OLAP):
  - Queries can consume lots of resources
  - Can saturate CPUs and disk bandwidth
  - Operating on static "snapshot" of data usually OK
- OLAP can "crowd out" OLTP transactions
  - Transactions are slow → unhappy users
- Example:
  - Analysis query asks for sum of all sales
  - Acquires lock on sales table for consistency
  - New sales transaction is blocked

Why OLAP & OLTP don't mix (2)

Different data modeling requirements

Transaction processing (OLTP):

- Normalized schema for consistency
- Complex data models, many tables
- Limited number of standardized queries and updates

Data analysis (OLAP):

- Simplicity of data model is important
  - Allow semi-technical users to formulate ad hoc queries
- De-normalized schemas are common
  - Fewer joins → improved query performance
  - Fewer tables → schema is easier to understand

Why OLAP & OLTP don't mix (3)

Analysis requires data from many sources

- An OLTP system targets one specific process
  - For example: ordering from an online store
- OLAP integrates data from different processes
  - Combine sales, inventory, and purchasing data
  - Analyze experiments conducted by different labs
- OLAP often makes use of historical data
  - Identify long-term patterns
  - Notice changes in behavior over time
- Terminology, schemas vary across data sources
  - Integrating data from disparate sources is a major challenge

DATA WAREHOUSING AND OPERATIONAL DBMS

What is Data Warehousing?

Definition:

- A data warehouse is a copy of transaction data specifically structured for querying and reporting.
- An expanded definition for data warehousing includes business intelligence tools, tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata.

NOTHING TO DO WITH WHETHER SOMETHING IS A DATA WAREHOUSE.

- This definition of the data warehouse focuses on data storage.
- A data warehouse can be normalized or denormalized.
- It can be a relational database, multidimensional database, flat file, hierarchical database, object database, etc.
- Data warehouse data often gets changed.
- And data warehouses often focus on a specific activity or entity.

Why Do We Need Data Warehouses?

- Consolidation of information resources
- Improved query performance
- Separate research and decision support functions from the operational systems
- Foundation for data mining, data visualization, advanced reporting and OLAP tools

- The data stored in the warehouse is uploaded from the operational systems.
- The data may pass through an operational data store for additional operations before it is used in the DW for reporting.
- Operational DBMS is used to deal with the everyday running of one aspect of an enterprise.
- OLTP (on-line transaction processor) or Operational DBMS are usually designed independently of each other and it is difficult for them to share information.

How Do Data Warehouses Differ From Operational DBMS?

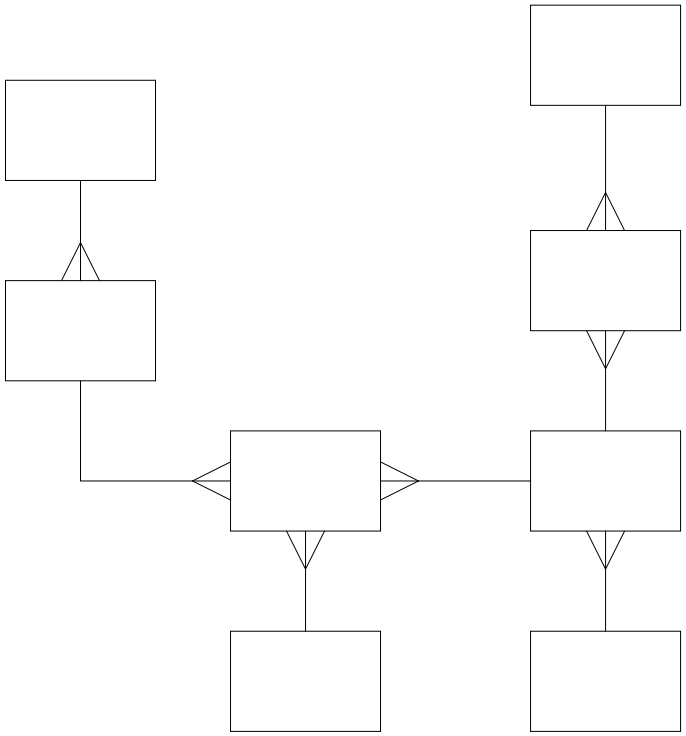
- Goals
- Structure
- Size
- Performance optimization
- Technologies used

How Do Data Warehouses Differ From Operational Systems?

Data warehouse	Operational DBMS
Subject oriented	Transaction oriented
Large (hundreds of GB up to several TB)	Small (MB up to several GB)
Historic data	Current data
De-normalized table structure (few tables, many columns per table)	Normalized table structure (many tables, few columns per table)
Batch updates	Continuous updates
Usually very complex queries	Simple to complex queries

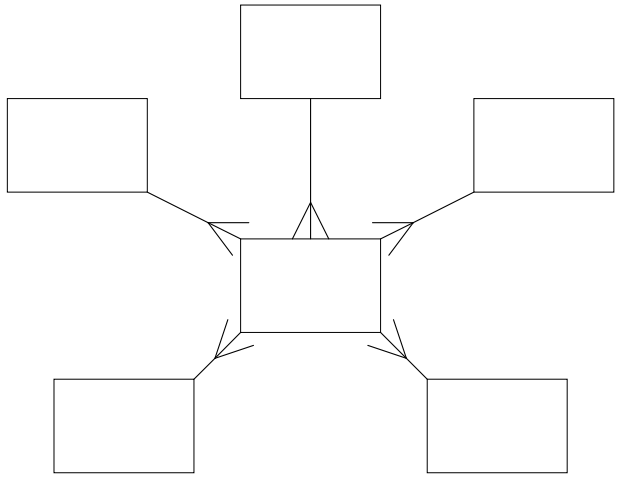
Design Differences

Operational DBMS



ER Diagram

Data Warehouse



Star Schema

FUNCTIONS

- A data warehouse maintains its functions in three layers: staging, integration, and access.
- Staging is used to store raw data for use by developers (analysis and support).
- The integration layer is used to integrate data and to have a level of abstraction from users.
- The access layer is for getting data out for users.

What Is a Data Warehouse Used for?

- Knowledge discovery
  - Making consolidated reports
  - Finding relationships and correlations
  - Data mining
  - Examples
    - Banks identifying credit risks
    - Insurance companies searching for fraud
    - Medical research

THE END