Data Warehouse and Data Mining

Lecture No. 07

Terminologies

Naeem Ahmed
Email: naeemmahoto@gmail.com

Department of Software Engineering
Mehran University of Engineering and Technology Jamshoro
Database

- Database is an organized collection of data
- Databases are created to operate large quantities of information by inputting, storing, retrieving, and managing that information.
- Collection of data central to some enterprise
- Essential to operation of enterprise
  - Contains the only record of enterprise activity
Database Management System

• DBMS is a suite of computer software providing the interface between users and a database or databases.

• A Database Management System (DBMS) is a program that manages a database:
  – Supports a high-level access language (e.g. SQL)
  – Application describes database accesses using that language
  – DBMS interprets statements of language to perform requested database access
Transaction

• A transaction is any kind of action involved in conducting business, or an interaction between people

• A transaction is executed to make change into database state of organization, whenever an event in real world changes the business state
Transaction Processing System

• Transaction processing (TP) is a style of computing that divides work into individual, indivisible operations, called transactions.

• A Transaction Processing System (TPS) or Transaction Server is a software system, or software/hardware combination, that supports transaction processing.

• A TPS consists of Transaction Processing Monitor (TPM), databases, and transactions.
Transaction Processing System

Diagram showing the flow of data from Clients to Computers through TPM to Application server with TP Monitor, leading to services and finally to Database servers.
• **High Availability**: on-line implies that it must be operational while enterprise is functioning

• **High Reliability**: correctly tracks state, does not lose data, controlled concurrency

• **High Throughput**: many users implies that many transactions per second

• **Low Response Time**: on-line implies that users are waiting
TPS Requirements

• **Long Lifetime**: complex systems are not easily replaced
  – Must be designed so they can be easily extended as the needs of the enterprise change

• **Security**: sensitive information must be carefully protected since system is accessible to many users
  – Authentication, authorization, encryption
Roles in TPS

- **System Analyst** - specifies system using input from customer; provides complete description of functionality from customer’s and user’s point of view

- **Database Designer** - specifies structure of data that will be stored in database

- **Application Programmer** - implements application programs (transactions) that access data and support enterprise rules
Roles in TPS

- **Database Administrator** - maintains database once system is operational: space allocation, performance optimization, database security

- **System Administrator** - maintains transaction processing system: monitors interconnection of Hardware and Software modules, deals with failures and congestion
Terminologies

• **On-Line:** A process controlled by a computer
• **Analytical Processing** needs Analytical Data
• **Analytical Data:** Data that involve analysis
  – Analytical Data consist of Business Data
• **Business Data:** Time, Customers, Sales, Stores, Products, etc
On-Line Transaction Processing

- **OLTP** a class of information systems that facilitate and manage *transaction-oriented applications*, typically for data entry and retrieval transaction processing.
- **OLTP** is day-to-day handling of transactions that result from enterprise operation.
- It maintains correspondence between database state and enterprise state.
- **OLTP** has also been used to refer to *processing* in which the system responds immediately to user requests. An automated teller machine (ATM) for a bank is an example of a commercial transaction processing application.
# OLTP versus Data Warehouse

<table>
<thead>
<tr>
<th></th>
<th>OLTP</th>
<th>Data Warehouse</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>users</strong></td>
<td>clerk, IT professional</td>
<td>knowledge worker</td>
</tr>
<tr>
<td><strong>function</strong></td>
<td>day to day operations</td>
<td>decision support</td>
</tr>
<tr>
<td><strong>DB design</strong></td>
<td>application-oriented</td>
<td>subject-oriented</td>
</tr>
<tr>
<td><strong>data</strong></td>
<td>current, up-to-date, detailed, flat relational, isolated</td>
<td>historical, summarized, multidimensional, integrated, consolidated</td>
</tr>
<tr>
<td><strong>usage</strong></td>
<td>repetitive</td>
<td>ad-hoc</td>
</tr>
<tr>
<td><strong>access</strong></td>
<td>read/write</td>
<td>lots of scans</td>
</tr>
<tr>
<td><strong>unit of work</strong></td>
<td>short, simple transaction</td>
<td>complex query</td>
</tr>
<tr>
<td><strong># records accessed</strong></td>
<td>tens</td>
<td>millions</td>
</tr>
<tr>
<td><strong># users</strong></td>
<td>thousands</td>
<td>hundreds</td>
</tr>
<tr>
<td><strong>DB size</strong></td>
<td>100MB-GB</td>
<td>100GB-TB</td>
</tr>
<tr>
<td><strong>metric</strong></td>
<td>transaction throughput</td>
<td>query throughput, response</td>
</tr>
</tbody>
</table>
Enterprise Resource Planning

- **ERP** is business management software (a suite of integrated applications)
- An organization can use to **store** and **manage** data from every stage of business, including:
  - Product planning, cost and development
  - Manufacturing
  - Marketing and sales
  - Inventory management
  - Shipping and payment
- **ERP** provides an **integrated real-time** view of core business processes, using **common databases** maintained by a database management system (DBMS)
Enterprise Resource Planning