Overview of Databases and Transaction Processing

Chapter 1

What is a Database?

- Collection of data central to some enterprise
- Essential to operation of enterprise
  - Contains the only record of enterprise activity
- An asset in its own right
  - Historical data can guide enterprise strategy
  - Of interest to other enterprises
- State of database mirrors state of enterprise
  - Database is persistent

What is a Database Management System?

- 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.

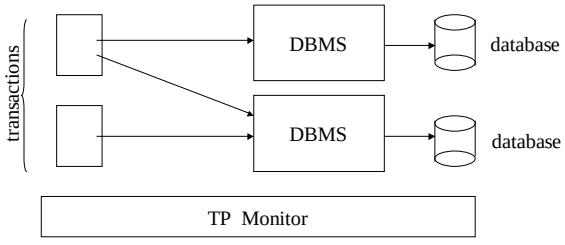
What is a Transaction?

- When an event in the real world changes the state of the enterprise, a transaction is executed to cause the corresponding change in the database state
  - With an on-line database, the event causes the transaction to be executed in real time
- A transaction is an application program with special properties - discussed later - to guarantee it maintains database correctness

What is a Transaction Processing System?

- Transaction execution is controlled by a TP monitor
  - Creates the abstraction of a transaction, analogous to the way an operating system creates the abstraction of a process
  - TP monitor and DBMS together guarantee the special properties of transactions
- A Transaction Processing System consists of TP monitor, databases, and transactions

Transaction Processing System



Systems: Then and Now

- Relational model using SQL - high-level view of data
  - Older systems presented low-level view
- Might contain multimedia data
  - Older systems restricted to alphanumeric data
- On-line: database accessed at time of event
  - Older systems were off-line, batch

7

Systems: Then and Now (con't)

- Concurrent - multiple transactions execute simultaneously
  - Older systems processed transactions sequentially
- Distributed computation - different parts of the application execute on different computers
  - Older systems were centralized

8

Systems: Then and Now (con't)

- Distributed data - different parts of the data are stored in different databases on different computers
  - Older systems were centralized
- Heterogeneous - involves HW and SW modules from different manufacturers
  - Older systems were homogeneous
- Accessed by everyone (e.g., e-commerce)
  - Older systems restricted to trained personnel

9

System Requirements

- **High Availability**: on-line => must be operational while enterprise is functioning
- **High Reliability**: correctly tracks state, does not lose data, controlled concurrency
- **High Throughput**: many users => many transactions/sec
- **Low Response Time**: on-line => users are waiting

10

System Requirements (con't)

- **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

11

Roles in Design, Implementation, and Maintenance of a 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

12

Roles in Design, Implementation and Maintenance of a TPS (con't)

- **Database Administrator** - maintains database once system is operational: space allocation, performance optimization, database security
- **System Administrator** - maintains transaction processing system: monitors interconnection of HW and SW modules, deals with failures and congestion

13

OLTP vs. OLAP

- **On-line Transaction Processing** (OLTP)
  - Day-to-day handling of transactions that result from enterprise operation
  - Maintains correspondence between database state and enterprise state
- **On-line Analytic Processing** (OLAP)
  - Analysis of information in a database for the purpose of making management decisions

14

OLAP

- Analyzes historical data (terabytes) using complex queries
- Due to volume of data and complexity of queries, OLAP often uses a data warehouse
- **Data Warehouse -** (offline) repository of historical data generated from OLTP or other sources
- **Data Mining -** use of warehouse data to *discover* relationships that might influence enterprise strategy

15

Examples - Supermarket

- OLTP
  - Event is 3 cans of soup and 1 box of crackers bought; update database to reflect that event
- OLAP
  - Last winter in all stores in northeast, how many customers bought soup and crackers together?
- Data Mining
  - Are there any interesting combinations of foods that customers frequently bought together?

16