Chapter 12: Data and Database Administration

Modern Database Management
6th Edition
Jeffrey A. Hoffer, Mary B. Prescott, Fred R. McFadden
Robert C. Nickerson
ISYS 464 – Spring 2003
Topic 24

Definitions

- **Data Administration**: A high-level function that is responsible for the overall management of data resources in an organization, including maintaining corporate-wide definitions and standards.
- **Database Administration**: A technical function that is responsible for physical database design and for dealing with technical issues such as security enforcement, database performance, and backup and recovery.
  - Often combined into one position

Data Administration Functions
- Data policies, procedures, standards
- Planning organization’s information architecture
- Data conflict (ownership) resolution
- Internal marketing of DA concepts
- Managing the data repository (metadata)

Database Administration Functions
- Selection of hardware and software (DBMS)
- Installing/upgrading DBMS
- Tuning database performance
- Improving query processing performance
- Managing data security, privacy, and integrity
- Data backup and recovery

Database Security
- **Database Security**: Protection of the data against accidental or intentional loss, destruction, or misuse
- Increased difficulty due to Internet access and client/server technologies

Figure 12-2: Possible locations of data security threats
Threats to Data Security

- Accidental losses attributable to:
  - Human error
  - Software failure
  - Hardware failure
- Theft and fraud.
- Improper data access:
  - Loss of privacy (personal data)
  - Loss of confidentiality (corporate data)
- Loss of data integrity
- Loss of availability (through, e.g. sabotage)

Views and Integrity Controls

- Views
  - Subset of the database that is presented to one or more users
  - User can be given access privilege to view without allowing access privilege to underlying tables
- Integrity Controls
  - Protect data from unauthorized use
  - Domains – set allowable values

Data Management Software Security Features

- Views or subschemas
- Integrity controls
- Authorization rules
- User-defined procedures
- Encryption
- Authentication schemes
- Backup, journalizing, and checkpointing

Authorization Rules

- Controls incorporated in the data management system
  - Restrict:
    - access to data
    - actions that people can take on data
- Authorization matrix for:
  - Subjects
  - Objects
  - Actions
  - Constraints

Views and Integrity Controls

- Views
  - Subset of the database that is presented to one or more users
  - User can be given access privilege to view without allowing access privilege to underlying tables
- Integrity Controls
  - Protect data from unauthorized use
  - Domains – set allowable values

Authorization Rules

- Controls incorporated in the data management system
  - Restrict:
    - access to data
    - actions that people can take on data
- Authorization matrix for:
  - Subjects
  - Objects
  - Actions
  - Constraints

Some DBMSs also provide capabilities for user-defined procedures to customize the authorization process.
**Authentication Schemes**
- **Goal** – obtain a positive identification of the user
- **Passwords are flawed:**
  - Users share them with each other
  - They get written down, could be copied
  - Automatic logon scripts remove need to explicitly type them in
  - Unencrypted passwords travel the Internet
- **Possible solutions:**
  - Biometric devices – use of fingerprints, retinal scans, etc. for positive ID
  - Third-party authentication – using secret keys, digital certificates

**Database Recovery**
- Mechanism for restoring a database quickly and accurately after loss or damage
- **Recovery facilities:**
  - Backup Facilities
  - Journalizing Facilities
  - Checkpoint Facility
  - Recovery Manager

**Backup Facilities**
- Automatic dump facility that produces backup copy of the entire database
- Periodic backup (e.g. nightly, weekly)
- Cold backup – database is shut down during backup
- Hot backup – selected portion is shut down and backed up at a given time
- Backups stored in secure, off-site location

**Journalizing Facilities**
- Audit trail of transactions and database updates
- Transaction log – record of essential data for each transaction processed against the database
- Database change log – images of updated data
  - Before-image – copy before modification
  - After-image – copy after modification
  - Produces an audit trail

**Recovery and Restart Procedures**
- Switch - Mirrored databases (keep exact copy simultaneously)
- Restore/Rerun - Reprocess transactions against the backup
- Transaction Integrity - Commit or abort all transaction changes
- Backward Recovery (Rollback) - Apply before images
- Forward Recovery (Roll Forward) - Apply after images (preferable to restore/rerun)
### Database Failure Responses

- **Aborted transactions**
  - Preferred recovery: rollback
  - Alternative: Rollforward to state just prior to abort

- **Incorrect data**
  - Preferred recovery: rollback
  - Alternative 1: re-run transactions not including inaccurate data updates
  - Alternative 2: compensating transactions

- **System failure (database intact)**
  - Preferred recovery: switch to duplicate database
  - Alternative 1: rollback
  - Alternative 2: restart from checkpoint

- **Database destruction**
  - Preferred recovery: switch to duplicate database
  - Alternative 1: rollforward
  - Alternative 2: reprocess transactions

### Concurrency Control

- **Problem** – in a multi-user environment, simultaneous access to data can result in interference and data loss

- **Solution – Concurrency Control**
  - The process of managing simultaneous operations against a database so that data integrity is maintained and the operations do not interfere with each other in a multi-user environment.

### Concurrency Control Techniques

- **Serializability** – 
  - Finish one transaction before starting another

- **Locking Mechanisms**
  - The most common way of achieving serialization
  - Data that is retrieved for the purpose of updating is locked for the updater
  - No other user can perform update until unlocked
This prevents the lost update problem

**Locking Mechanisms**

- **Locking level:**
  - Database – used during database updates
  - Table – used for bulk updates
  - Block or page – very commonly used
  - Record – only requested row; fairly commonly used
  - Field – requires significant overhead; impractical

- **Types of locks:**
  - Shared lock - Read but no update permitted. Used when just reading to prevent another user from placing an exclusive lock on the record
  - Exclusive lock - No access permitted. Used when preparing to update

**Deadlock**

- An impasse that results when two or more transactions have locked common resources, and each waits for the other to unlock their resources

**Versioning**

- Optimistic approach to concurrency control
- Instead of locking
- Assumption is that simultaneous updates will be infrequent
- Each transaction can attempt an update as it wishes
- The system will reject an update when it senses a conflict
- Use of rollback and commit for this

**Managing Deadlock**

- **Deadlock prevention:**
  - Lock all records required at the beginning of a transaction
  - Two-phase locking protocol
    - Growing phase
    - Shrinking phase
  - May be difficult to determine all needed resources in advance

- **Deadlock Resolution:**
  - Allow deadlocks to occur
  - Mechanisms for detecting and breaking them
    - Resource usage matrix

**Better performance than locking**
Managing Data Quality

- Five Data Quality Issues:
  - Security policy and disaster recovery
  - Personnel controls
  - Physical access controls
  - Maintenance controls (hardware & software)
  - Data protection and privacy

Database Performance Tuning

- DBMS Installation
  - Setting installation parameters
- Memory Usage
  - Set cache levels
  - Choose background processes
- Disk Input/Output (Read/Write) Contention
  - Use striping
  - Distribution of heavily accessed files
- CPU Usage
  - Monitor CPU load
- Application tuning
  - Modification of SQL code in applications

Data Warehouse Administration

- New role, coming with the growth in data warehouses
- Similar to DA/DBA roles
- Emphasis on integration and coordination of metadata/data across many data sources
- Specific roles:
  - Support decision support applications
  - Manage data warehouse growth
  - Establish service level agreements regarding data warehouses and data marts