Chapter 7: Database Administration and Security

Objectives

- Define terms
- List functions and roles of data/database administration
- Describe role of data dictionaries and information repositories
- Compare optimistic and pessimistic concurrency control
- Describe problems and techniques for data security
- Understand role of databases in Sarbanes-Oxley compliance
- Describe problems and facilities for data recovery
- Describe database tuning issues and list areas where changes can be done to tune the database
- Describe importance and measures of data availability

Data and Database Administration

- **Data administration** is 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** is 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.

Data Administration Functions/Roles

1. Set data policies, procedures, standards

- **Data policies** are statements of the goals of data administration.

  Examples: - All users must have passwords.

  - Passwords must be changed every six months.

- **Data procedures** are written instructions that describe the steps when perform certain activities. Example: backup and recovery procedures
- **Data standards** are rules to be followed when performing database activities.

  Examples: - A password must have a minimum of 8 characters.

  - IC numbers, names and birth dates cannot be used as passwords.

Data Administration Functions/Roles

2. Develop information architecture

Data administrators must understand the data and information needed for the organizations and able to lead the development of the information architecture to meet the diverse needs of the organization.

3. Resolve data conflict

Databases are shared and involve data from several departments. A data conflict occurs when two or more departments claims the ownership of the same data. Establishing procedures for resolving such conflicts is essential.

4. Managing the data repository

A data repository is the data storage used by DBMS to store data generated by application programs.

Examples of ineffective data administration that leads to poor data quality

- Multiple data definitions of the same data elements in separate databases can cause data integration problems
- Missing key data elements can eliminate the value of existing data
- Inappropriate data sources and timing can reduce data reliability
- Poor query response time and excessive database downtime
- Lack of access to data due to damaged, sabotaged, or stolen data files or due to hardware failures
- Unauthorized access to data can cause embarrassment to organization

Database Administration Functions/Roles

- Analyzing and designing databases
- Selecting DBMS and related software tools
- Installing/upgrading DBMS
- Tuning database performance
- Improving database query processing performance
- Managing data security, privacy, and integrity
- Performing data backup and recovery

Data Warehouse Administration

- The significant growth in data warehousing has caused a new role, **data warehouse administrator** (DWA) to emerge.
- Similar to DA/DBA roles
- The role of a DWA emphasizes on integration and coordination of metadata and data across many data sources.
- Specific roles:
  - Build and manage decision support applications
  - Manage data warehouse growth
  - Develop service-level agreements with suppliers and consumers of data for the data warehouse.

Open Source Database Management

- An alternative to proprietary packages such as Oracle, Microsoft SQL Server, or Microsoft Access
- MySQL is an example of an open source DBMS
- Less expensive than proprietary packages
- Source code available, for modification
- Absence of complete documentation
- Ambiguous licensing concerns
- Not as feature-rich as proprietary DBMSs
- Vendors may not have certification programs

Database Recovery

- **Database recovery** is the mechanism for restoring a database quickly and accurately after loss or damage.
- A DBMS should provide four basic facilities for backup and recovery of a database:
  a) Backup facilities
  b) Journalizing facilities
  c) Checkpoint facility
  d) Recovery manager

a) Backup Facilities

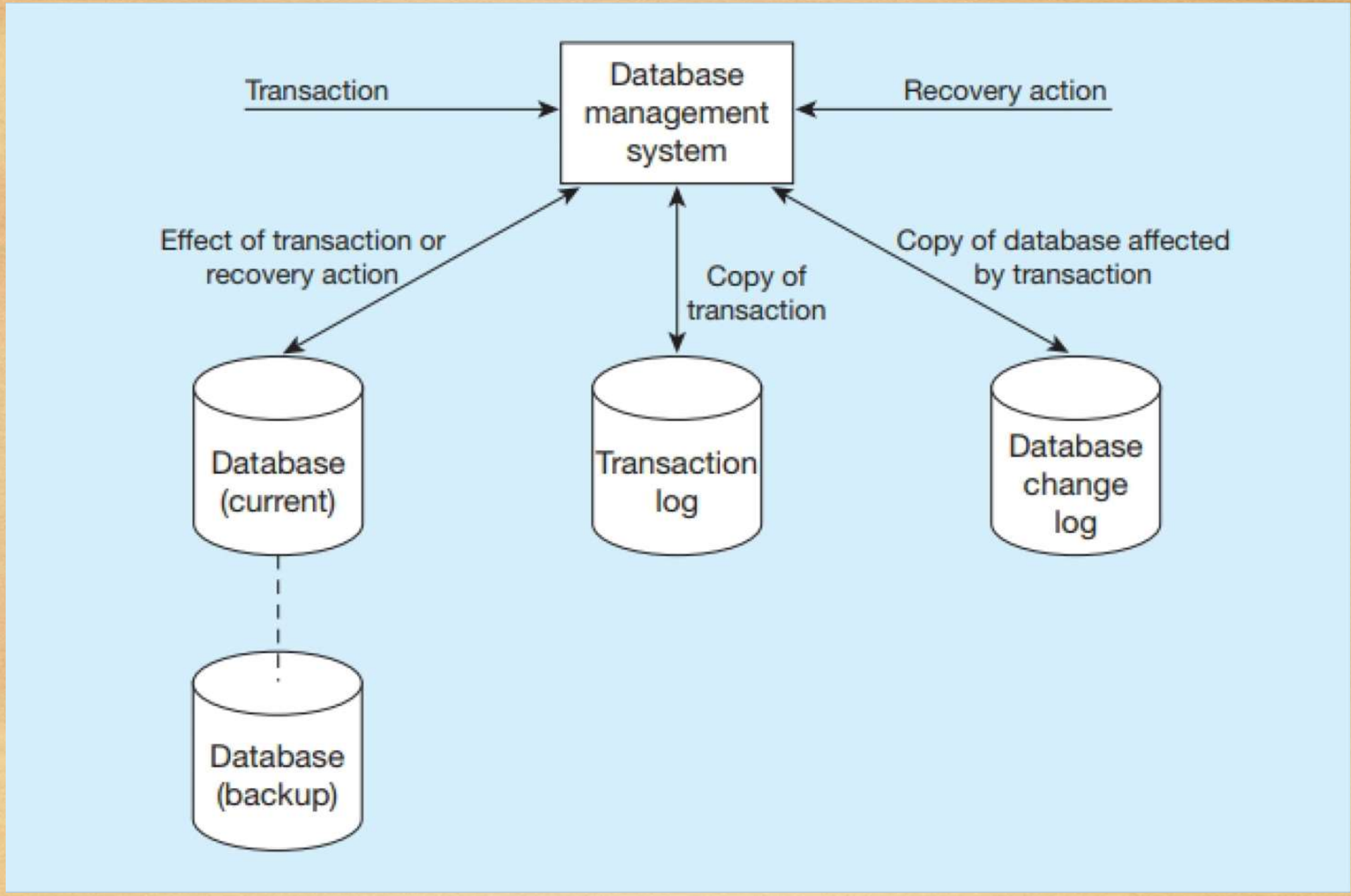
- A **backup facility** provide a periodic backup (nightly or weekly) that produces a backup copy of the portion of or the entire database. Each DBMS provides a COPY utility for this purpose.
- Cold backup – database is shut down during backup.
- Hot backup – selected portion is shut down and backed up at a given time.
- Backups stored in secure, off-site location.

b) Journalizing Facilities

- A DBMS must provide **journalizing facilities** to produce an audit trail of transactions and database changes.
- In the event of a failure, a consistent database state can be reestablished, using the information in the journals together with the most recent complete backup.
- There are two basic journals, or logs.

1. **Transaction log**, which contains a record of essential data for each transaction that is processed against the database.
2. **Database change log**, which contains before and after image of records that have been modified by transactions. A **before image** is simply a copy of a record before it has been modified, and an **after image** is a copy of the same record after it has been modified.

Database audit trail



c) Checkpoint Facilities

- A **checkpoint facility** in a DBMS periodically refuses to accept any new transactions. All transactions in progress are completed, and the journal files are brought up to date. At this point, the system is in a quiet state, and the database and transaction logs are synchronized.
- A DBMS may perform checkpoints automatically (which is preferred) or in response to commands in user application programs.
- Checkpoints should be taken frequently (several times an hour).
- When failures occur, it is often possible to resume processing from the most recent checkpoint. Thus, only a few minutes of processing work must be repeated, compared with several hours for a complete restart of the day's processing.

d) Recovery Manager

- The recovery manager is a module of a DBMS that restores the database to a correct condition when a failure occurs and then resumes processing user requests.
- The recovery manager uses the *transaction log* and the *database change log* to restore the database.

Recovery and Restart Procedures

Techniques used in recovery procedures include:

a) **Disk Mirroring** – switch between identical copies of databases

- To be able to switch to an existing copy of a database, the database must be mirrored. That is, at least two copies of the database must be kept and updated simultaneously. When a media failure occurs, processing is switched to the duplicate copy of the database.
- This strategy allows for the fastest recovery and has become increasingly popular for applications requiring high availability as the cost of long-term storage has dropped.

Recovery and Restart Procedures

Techniques used in recovery procedures include:

b) **Restore/Rerun** – reprocess transactions against the backup copy

- First, the database is shut down. Then the most recent backup copy of the database is mounted, and all transactions that have occurred since that copy (which are stored on the transaction log) are rerun.

Recovery and Restart Procedures

Techniques used in recovery procedures include:

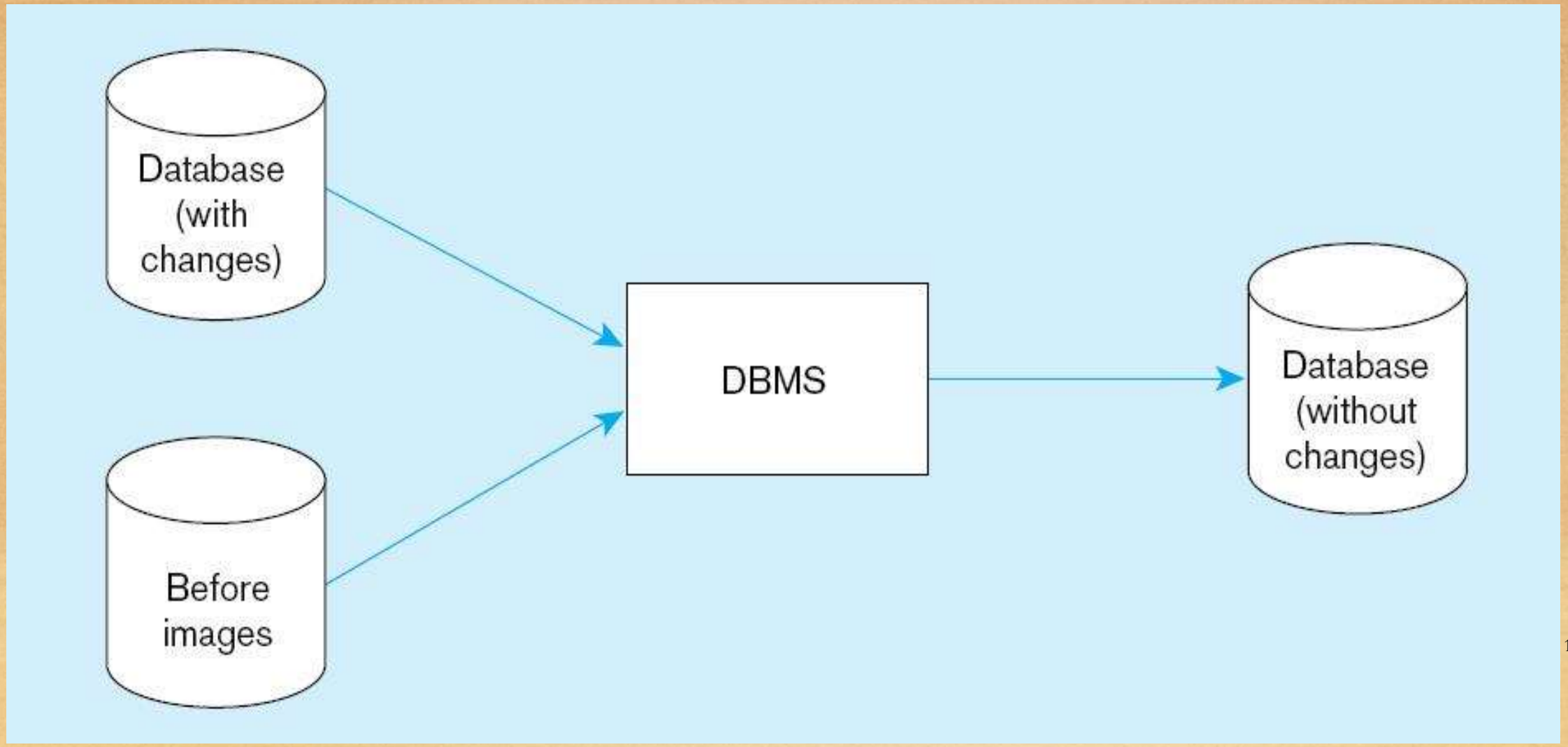
c) Backward Recovery (Rollback) – apply *before images*

- With backward recovery (also called rollback), the DBMS undo unwanted changes to the database by applying *before images* of the records to the database. As a result, the database is returned to an earlier state and the unwanted changes are eliminated.

d) Forward Recovery (Rollforward) – apply *after images* (better than restore/rerun)

With forward recovery (also called rollforward), the DBMS starts with an earlier copy of the database. Applying *after images* (the results of good transactions) quickly moves the database forward to a later state.

Basic recovery techniques: Rollback



Basic recovery techniques: Rollforward

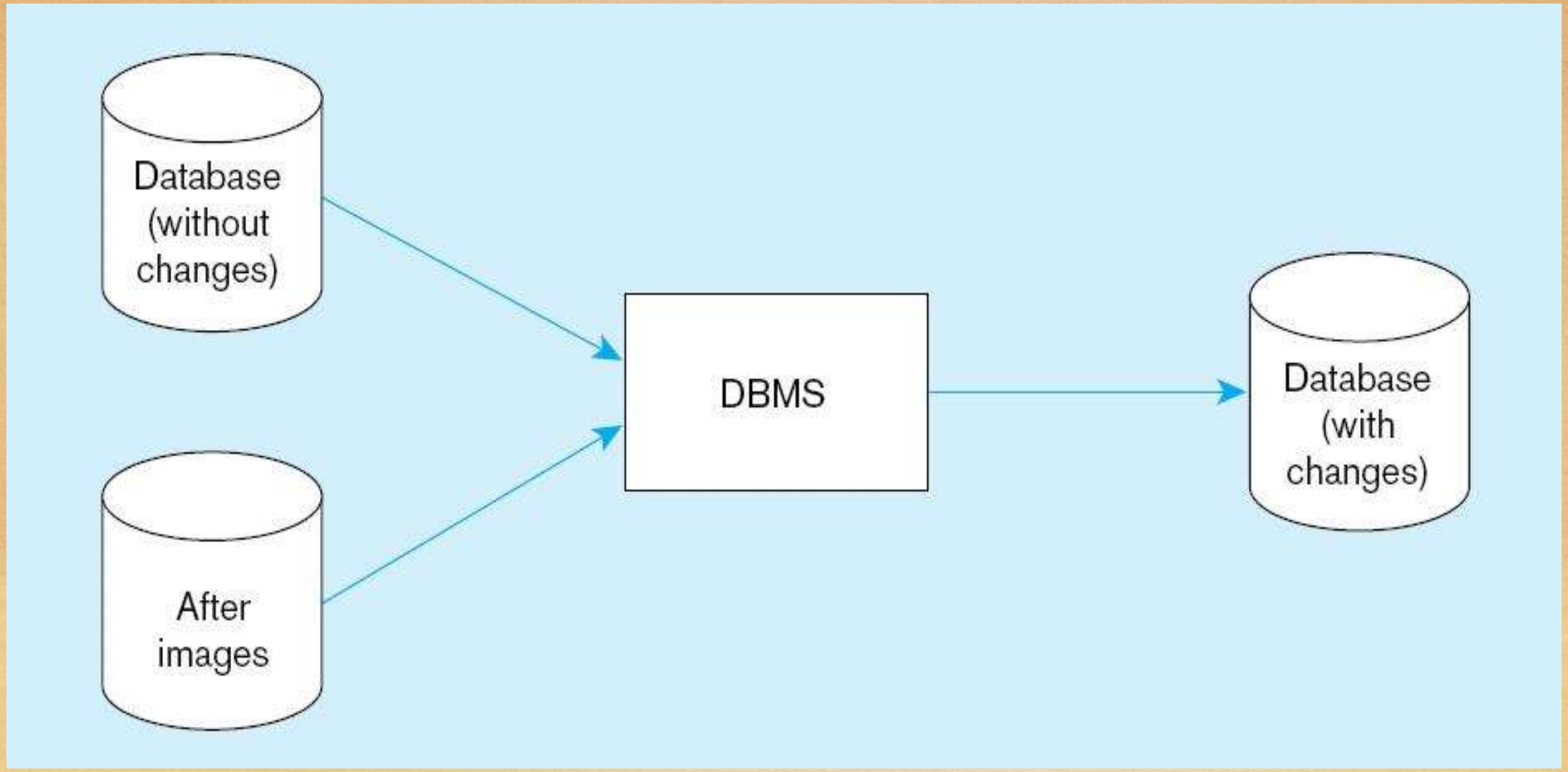


TABLE 11-1 Responses to Database Failure

Type of Failure	Recovery Technique
Aborted transaction	Rollback (preferred)
	Rollforward/return transactions to state just prior to abort
Incorrect data (update inaccurate)	Rollback (preferred)
	Reprocess transactions without inaccurate data updates
	Compensating transactions
System failure (database intact)	Switch to duplicate database (preferred)
	Rollback
	Restart from checkpoint (rollforward)
Database destruction	Switch to duplicate database (preferred)
	Rollforward
	Reprocess transactions

Data Availability

- How to ensure data availability?
  - Hardware failures – provide redundancy or standby components to replace a failing system
  - Loss of data – database mirroring
  - Human error – standard operating procedures, training, documentation
  - Maintenance downtime – automated and non-disruptive maintenance utilities
  - Network problems – careful traffic monitoring, firewalls, and routers

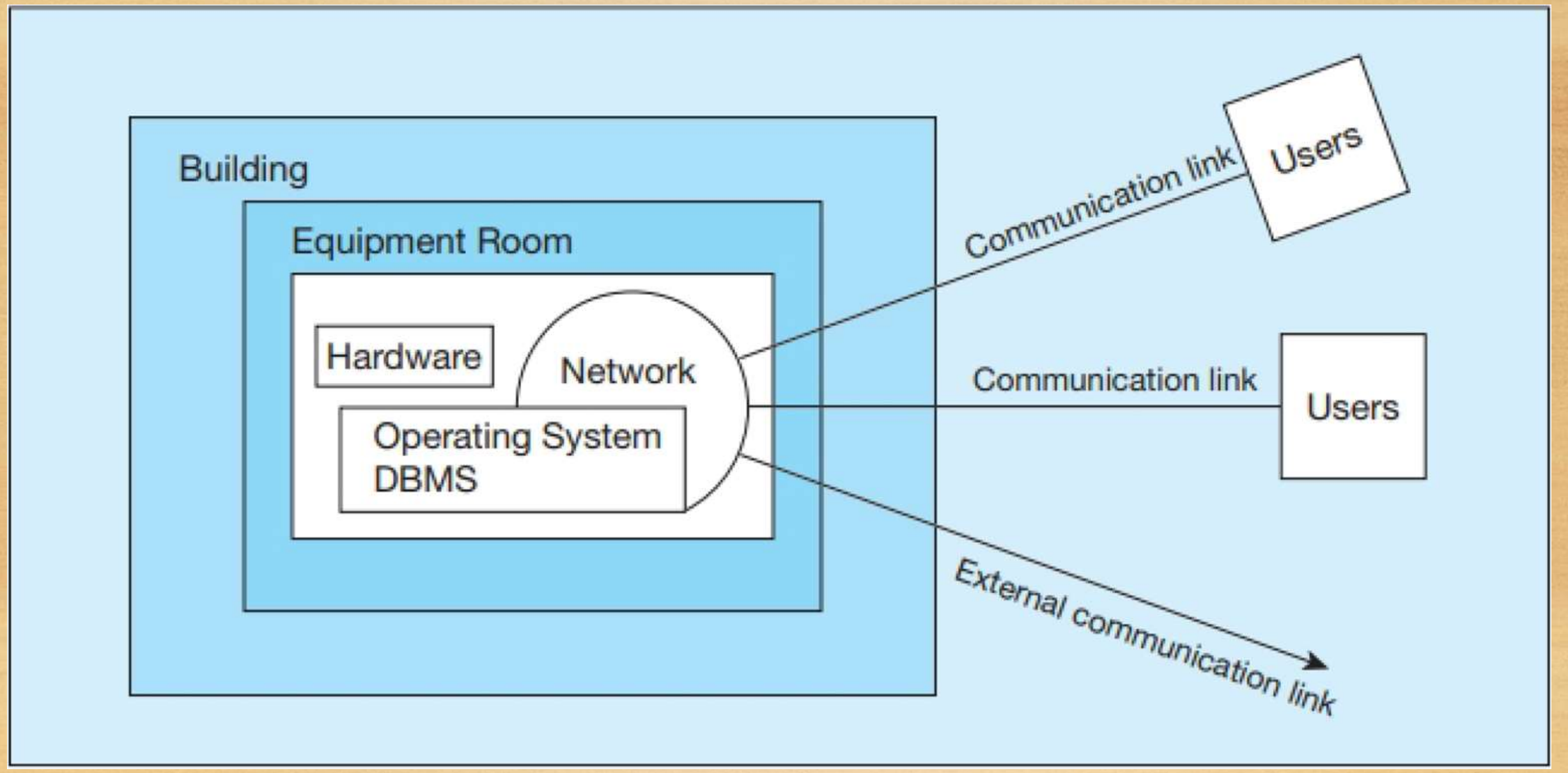
Data Security

- Database security refers to the protection of the data against accidental or intentional loss, destruction, or misuse
- Access to data has become more open through the Internet and corporate intranets and from mobile computing devices. As a result, managing data security has become more difficult and time consuming.

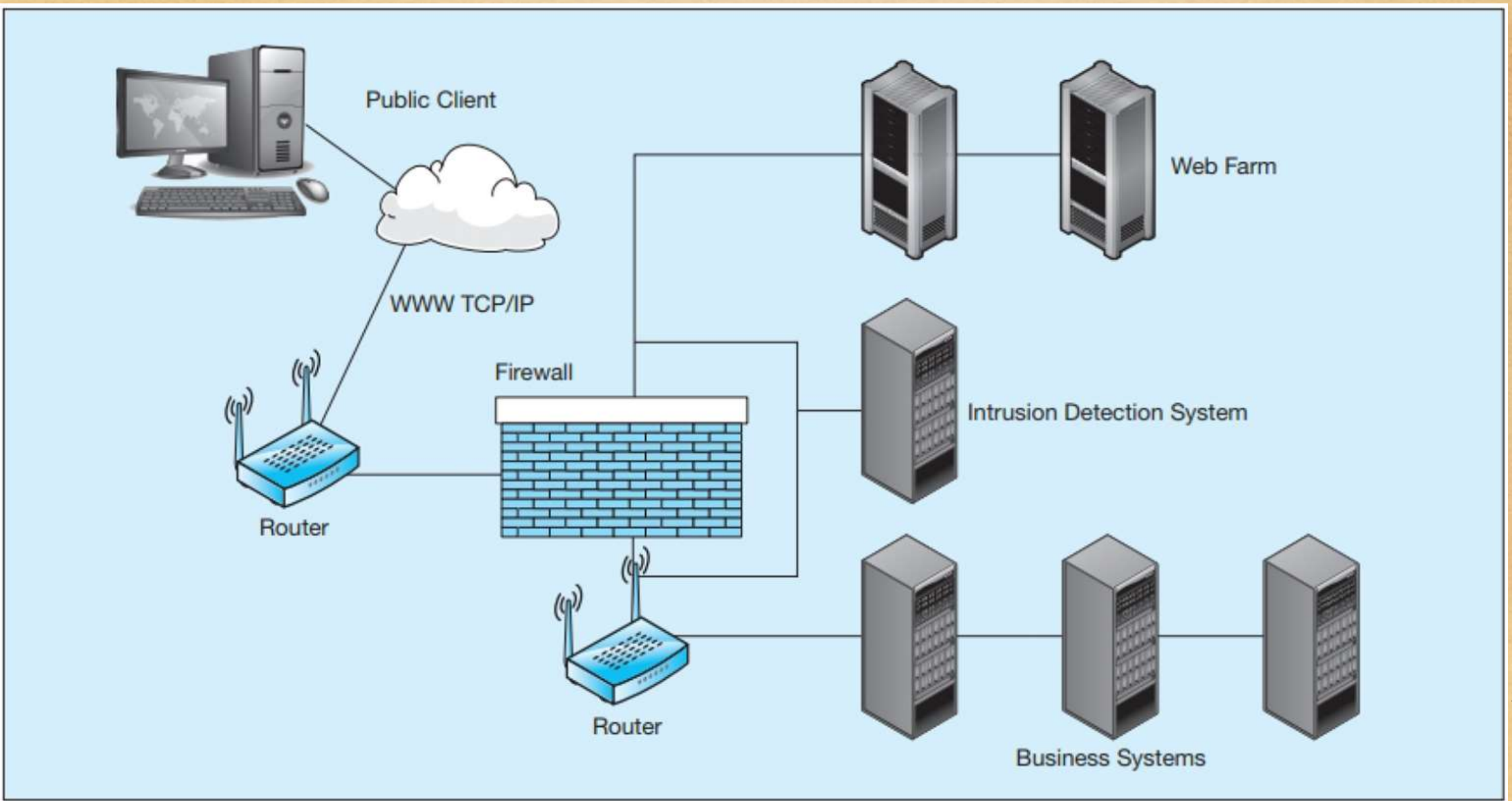
Threats to Data Security

- Accidental losses due to human error, software and hardware-caused breaches
- Theft and fraud (see the figure on next slide)
- Loss of privacy (personal data) or confidentiality (corporate data)
- Loss of data integrity (data will be invalid or corrupted)
- Loss of availability (e.g., through sabotage of hardware, networks, or applications)

Possible locations of data security threats



Establishing Internet Security



Client – Server Application security

- Static HTML files are easy to secure
  - Standard database access controls
  - Place Web files in protected directories on server
- Dynamic pages are harder to secure
  - User authentication
  - Session security
  - SSL for encryption
  - Restrict number of users and open ports
  - Remove unnecessary programs

SSL (Secure Sockets Layer)

- Standard security technology for establishing an encrypted link between a web server and a browser.
- This link ensures that all data passed between the web server and browsers remain private and secure.

Database Software Security Features

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

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
  - Assertions – enforce database conditions
  - Triggers – prevent inappropriate actions, invoke special handling procedures, write to log files

Authorization Rules

- **Authorization rules** are controls incorporated in the data management system restrict access to data and also restrict actions that people can take on data.

- **Authorization matrix**

Subject	Object	Action	Constraint
Sales Dept.	Customer record	Insert	Credit limit LE $5000
Order trans.	Customer record	Read	None
Terminal 12	Customer record	Modify	Balance due only
Acctg. Dept.	Order record	Delete	None
Ann Walker	Order record	Insert	Order amt LT $2000
Program AR4	Order record	Modify	None

FIGURE 8-10 Implementing authorization rules

(a) Authorization table for subjects (salespersons)

	Customer records	Order records
Read	Y	Y
Insert	Y	Y
Modify	Y	N
Delete	N	N

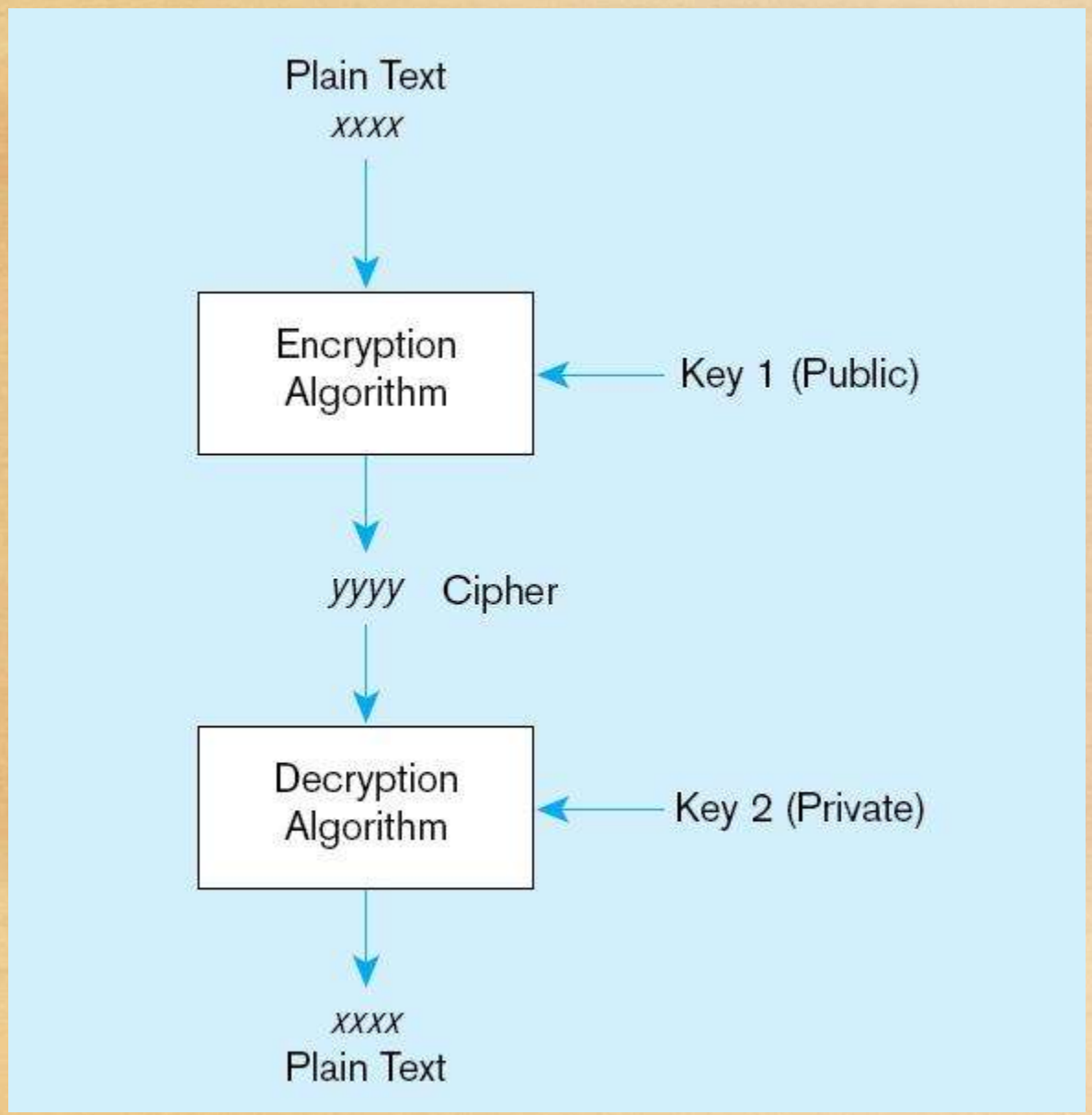
(b) Authorization table for objects (order records)

	Salespersons (password BATMAN)	Order entry (password JOKER)	Accounting (password TRACY)
Read	Y	Y	Y
Insert	N	Y	N
Modify	N	Y	Y
Delete	N	N	Y

FIGURE 8-11 Oracle privileges

Privilege	Capability
SELECT	Query the object.
INSERT	Insert records into the table/view. Can be given for specific columns.
UPDATE	Update records in table/view. Can be given for specific columns.
DELETE	Delete records from table/view.
ALTER	Alter the table.
INDEX	Create indexes on the table.
REFERENCES	Create foreign keys that reference the table.
EXECUTE	Execute the procedure, package, or function.

Basic two-key encryption



Encryption – the coding or scrambling of data so that humans cannot read them

Secure Sockets Layer (SSL) is a popular encryption scheme for TCP/IP connections.

34

Authentication Schemes

- Goal – obtain a *positive* identification of the user
- Passwords: First line of defense
  - Should be at least 8 characters long
  - Should combine alphabetic and numeric data
  - Should not be complete words or personal information
  - Should be changed frequently

Authentication Schemes (cont.)

- Strong Authentication
  - 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:
  - Two factor – e.g., smart card plus PIN
  - Three factor – e.g., smart card, biometric, PIN

Logical Access to Data

- Personnel controls
  - Hiring practices, employee monitoring, security training, separation of duties
- Physical access controls
  - Swipe cards, equipment locking, check-out procedures, screen placement, laptop protection