Chapter Seven

Advanced Concepts in Database Systems

- **Database Security and Integrity**
- **Distributed Database Systems**
- **Data warehousing**

1. Database Security and Integrity

A database represents an essential corporate resource that should be properly secured using appropriate controls.

- Database security encompasses hardware, software, people and data

Multi-user database system - DBMS must provide a database security and authorization subsystem to enforce limits on individual and group access rights and privileges.

Database security and integrity is about protecting the database from being inconsistent and being disrupted. We can also call it database misuse.

Database misuse could be Intentional or accidental, where accidental misuse is easier to cope with than intentional misuse.
Accidental inconsistency could occur due to:

- System crash during transaction processing
- Anomalies due to concurrent access
- Anomalies due to redundancy
- Logical errors

Like wise, even though there are various threats that could be categorized in this group, intentional misuse could be:

- Unauthorized reading of data
- Unauthorized modification of data or
- Unauthorized destruction of data

Most systems implement good **Database Integrity** to protect the system from accidental misuse while there are many computer based measures to protect the system from intentional misuse, which is termed as **Database Security** measures.

- Database security is considered in relation to the following situations:
  - Theft and fraud
  - Loss of confidentiality (secrecy)
  - Loss of privacy
  - Loss of integrity
  - Loss of availability

Security Issues and general considerations

- **Legal**, **ethical** and **social** issues regarding the right to access information
- Physical control
- **Policy** issues regarding privacy of individual level at enterprise and national level
- **Operational** consideration on the techniques used (password, etc)
- **System** level security including operating system and hardware control
- Security levels and security policies in enterprise level

- **Database security** - the mechanisms that protect the database against intentional or accidental ***threats.*** And Database security encompasses hardware, software, people and data

- **Threat –** any situation or event, whether intentional or accidental, that may adversely affect a system and consequently the organization
- A threat may be caused by a situation or event involving a person, action, or circumstance that is likely to bring harm to an organization
- The harm to an organization may be ***tangible*** or ***intangible***
  Tangible – loss of hardware, software, or data
  Intangible – *loss of credibility or client confidence*

Examples of threats:

- ✓ Using another persons' means of access
- ✓ Unauthorized amendment/modification or copying of data
- ✓ Program alteration
- ✓ Inadequate policies and procedures that allow a mix of confidential and normal out put
- ✓ Wire-tapping
- ✓ Illegal entry by hacker
- ✓ Blackmail
- ✓ Creating 'trapdoor' into system
- ✓ Theft of data, programs, and equipment
- ✓ Failure of security mechanisms, giving greater access than normal
- ✓ Staff shortages or strikes
- ✓ Inadequate staff training
- ✓ Viewing and disclosing unauthorized data
- ✓ Electronic interference and radiation
- ✓ Data corruption owing to power loss or surge
- ✓ Fire (electrical fault, lightning strike, arson), flood, bomb
- ✓ Physical damage to equipment
- ✓ Breaking cables or disconnection of cables
- ✓ Introduction of viruses

Levels of Security Measures

Security measures can be implemented at several levels and for different components of the system. These levels are:

1. **Physical Level:** concerned with securing the site containing the computer system should be physically secured. The backup systems should also be physically protected from access except for authorized users.
2. **Human Level:** concerned with authorization of database users for access the content at different levels and privileges.
3. **Operating System:** concerned with the weakness and strength of the operating system security on data files. Weakness may serve as a means of unauthorized access to the database. This also includes protection of data in primary and secondary memory from unauthorized access.
4. **Database System:** concerned with data access limit enforced by the database system. Access limit like password, isolated transaction and etc.

Even though we can have different levels of security and authorization on data objects and users, ***who access which data is a policy matter rather than technical.***

These policies

- should be known by the system: should be encoded in the system
- should be remembered: should be saved somewhere (the catalogue)

- An organization needs to identify the types of threat it may be subjected to and initiate appropriate plans and ***countermeasures***, bearing in mind the costs of implementing them

Countermeasures: Computer based controls

- The types of countermeasure to threats on computer systems range from **physical controls to administrative procedures**
- Despite the range of computer-based controls that are available, it is worth noting that, generally, ***the security of a DBMS is only as good as that of the operating system***, owing to their close association
- The following are computer-based security controls for a multi-user environment:

➢ Authorization

- The granting of a right or privilege that enables a subject to have legitimate access to a system or a system's object
- Authorization controls can be built into the software, and govern not only what system or object a specified user can access, but also what the user may do with it
- Authorization controls are sometimes referred to as ***access controls***
- The process of authorization involves authentication of ***subjects*** (i.e. a user or program) requesting access to ***objects*** (i.e. a database table, view, procedure, trigger, or any other object that can be created within the system)

➢ Views

- A view is the dynamic result of one or more relational operations operation on the base relations to produce another relation
- A view is a virtual relation that does not actually exist in the database, but is produced upon request by a particular user
- The view mechanism provides a powerful and flexible security mechanism by hiding parts of the database from certain users
- Using a view is more restrictive than simply having certain privileges granted to a user on the base relation(s)

➢ Integrity

- Integrity constraints contribute to maintaining a secure database system by preventing data from becoming invalid and hence giving misleading or incorrect results
- Domain Integrity
- Entity integrity
- Referential integrity
- Key constraints

➢ Backup and recovery

- Backup is the process of periodically taking a copy of the database and log file (and possibly programs) on to offline storage media
- A DBMS should provide backup facilities to assist with the recovery of a database following failure
- Database recovery is the process of restoring the database to a correct state in the event of a failure
- Journaling is the process of keeping and maintaining a log file (or journal) of all changes made to the database to enable recovery to be undertaken effectively in the event of a failure
- The advantage of journaling is that, in the event of a failure, the database can be recovered to its last known consistent state using a backup copy of the database and the information contained in the log file
- If no journaling is enabled on a failed system, the only means of recovery is to restore the database using the latest backup version of the database
- However, without a log file, any changes made after the last backup to the database will be lost

➢ Encryption

- The encoding of the data by a special algorithm that renders the data unreadable by any program without the decryption key
- If a database system holds particularly sensitive data, it may be deemed necessary to encode it as a precaution against possible external threats or attempts to access it
- The DBMS can access data after decoding it, although there is a degradation in performance because of the time taken to decode it
- Encryption also protects data transmitted over communication lines
- To transmit data securely over insecure networks requires the use of a ***Cryptosystem***, which includes:

Authentication

- All users of the database will have different access levels and permission for different data objects, and authentication is the process of checking whether the user is the one with the privilege for the access level.
- Is the process of checking the users are who they say they are.
- Each user is given a unique identifier, which is used by the operating system to determine who they are
- Thus the system will check whether the user with a specific username and password is trying to use the resource.
- Associated with each identifier is a password, chosen by the user and known to the operation system, which must be supplied to enable the operating system to authenticate who the user claims to be

Any database access request will have the following three major components

1. **Requested Operation:** what kind of operation is requested by a specific query?
2. **Requested Object:** on which resource or data of the database is the operation sought to be applied?
3. **Requesting User:** who is the user requesting the operation on the specified object?

The database should be able to check for all the three components before processing any request. The checking is performed by the security subsystem of the DBMS.

Forms of user authorization

There are different forms of user authorization on the resource of the database. These forms are privileges on what operations are allowed on a specific data object.

User authorization on the data/extension

1. **Read Authorization**: the user with this privilege is allowed only to read the content of the data object.
2. **Insert Authorization**: the user with this privilege is allowed only to insert new records or items to the data object.
3. **Update Authorization**: users with this privilege are allowed to modify content of attributes but are not authorized to delete the records.
4. **Delete Authorization**: users with this privilege are only allowed to delete a record and not anything else.

- Different users, depending on the power of the user, can have one or the combination of the above forms of authorization on different data objects.

Role of DBA in Database Security

The database administrator is responsible to make the database to be as secure as possible. For this the DBA should have the most powerful privilege than every other user. The DBA provides capability for database users while accessing the content of the database.

The major responsibilities of DBA in relation to authorization of users are:

1. ***Account Creation:*** involves creating different accounts for different **USERS** as well as **USER GROUPS.**
2. ***Security Level Assignment***: involves in assigning different users at different categories of access levels.
3. ***Privilege Grant:*** involves giving different levels of privileges for different users and user groups.
4. ***Privilege Revocation:*** involves denying or canceling previously granted privileges for users due to various reasons.
5. ***Account Deletion:*** involves in deleting an existing account of users or user groups. Is similar with denying all privileges of users on the database.

2. Distributed Database Systems

- Database development facilitates the integration of data available in an organization and enforces security on data access. But it is not always the case that organizational data reside in one site. This demand databases at different sites to be integrated and synchronized with all the facilities of database approach. This leads to Distributed Database Systems.
- In a distributed database system, the database is stored on several computers. The computers in a distributed system communicate with each other through various communication media, such as high speed buses or telephone line.
- A distributed database system consists of a collection of sites, each of which maintains a local database system and also participates in global transaction where different databases are integrated together.
- Even though integration of data implies centralized storage and control, in distributed database systems the intention is different. Data is stored in different database systems in a decentralized manner but act as if they are centralized through development of computer networks.
- A distributed database system consists of loosely coupled sites that share no physical component and database systems that run on each site are independent of each other.
- Transactions may access data at one or more sites
- Organization may implement their database system on a number of separate computer system rather than a single, centralized mainframe. Computer Systems may be located at each local branch office.

The functionalities of a DDBMS will include: Extended Communication Services, Extended Data Dictionary, Distributed Query Processing, Extended Concurrency Control and Extended Recovery Services.

Concepts in DDBMS

- **Replication:** System maintains multiple copies of data, stored in different sites, for faster retrieval and fault tolerance.
- **Fragmentation:** Relation is partitioned into several fragments stored in distinct sites
- **Data transparency**: Degree to which system user may remain unaware of the details of how and where the data items are stored in a distributed system

Advantages of DDBMS

1. **Data sharing and distributed control:**
   - User at one site may be able access data that is available at another site.
   - Each site can retain some degree of control over local data
   - We will have local as well as global database administrator

2. **Reliability and availability of data**
   - If one site fails the rest can continue operation as long as transaction does not demand data from the failed system and the data is not replicated in other sites

3. **Speedup of query processing**
   - If a query involves data from several sites, it may be possible to split the query into sub-queries that can be executed at several sites which is parallel processing

Disadvantages of DDBMS

1. **Software development cost**
2. **Greater potential for bugs (parallel processing may endanger correctness)**
3. **Increased processing overhead (due to communication jargons)**
4. **Communication problems**

Homogeneous and Heterogeneous Distributed Databases

- **In a homogeneous distributed database**
  - All sites have identical software
  - Are aware of each other and agree to cooperate in processing user requests.
  - Each site surrenders part of its autonomy in terms of right to change schemas or software
  - Appears to user as a single system
- **In a heterogeneous distributed database**
  - Different sites may use different schemas and software
    - Difference in schema is a major problem for query processing
    - Difference in software is a major problem for transaction processing
  - Sites may not be aware of each other and may provide only limited facilities for cooperation in transaction processing

3. Data warehousing

- ***Data warehouse is an integrated, subject-oriented, time-variant, non-volatile database that provides support for decision making.***

  - ✓ ***Integrated*** → centralized, consolidated database that integrates data derived from the entire organization.
    - ➢ Consolidates data from multiple and diverse sources with diverse formats.
    - ➢ Helps managers to better understand the company's operations.
  - ✓ ***Subject-Oriented*** → Data warehouse contains data organized by topics. Eg. Sales, marketing, finance, etc.
  - ✓ ***Time variant***: In contrast to the operational data that focus on current transactions, the warehouse data represent the flow of data through time.
    - ➢ Data warehouse contains data that reflect what happened last week, last month, past five years, and so on.
  - ✓ ***Non volatile*** → Once data enter the data warehouse, they are never removed. Because the data in the warehouse represent the company's entire history.

Differences between database and data warehouse

- ✓ Because data is added all the time, warehouse is growing.
- ✓ The data warehouse and operational environments are separated. Data warehouse receives its data from operational databases.
- ✓ Data warehouse environment is characterized by read-only transactions to very large data sets.
- ✓ Operational environment is characterized by numerous update transactions to a few data entities at a time.
- ✓ Data warehouse contains historical data over a long time horizon.

- Ultimately Information is created from data warehouses. Such Information becomes the basis for rational decision making.

- The data found in data warehouse is analyzed to discover previously unknown data characteristics, relationships, dependencies, or trends.