Database Management System

MSAJCE

Database Management System

Database: A collection of related data. It should support

- Definition
- Construction
- Manipulation

Database Management System: A collection of programs that enable the users to create and maintain a database.

Features of DBMS

1. Data storage, retrieval, and update: The ability to store, retrieve, and update the data that are in the database.

2. User-accessible catalog: where descriptions of database components are stored and are accessible to the users

3. Shared update support: A mechanism to ensure accuracy when several users are updating the database at the same time

4. Backup and Recovery Services: Mechanisms for recovering the database in the event that a database is damaged somehow.

5. Security Services: Mechanisms to ensure that certain rules are followed with regard to data in the database and any changes that are made in the data

Features of DBMS

5. Integrity services: Mechanisms to ensure that certain rules are followed with regard to data in the database and any changes that are made in the data.

6. Data Independence: Facilities to support the independence of programs from the structure of the database.

7. Replication support: A facility to manage copies of the same data at multiple locations.

8. Utility Services: DBMS provided services that assist in the general maintenance of the database.

Shared Updates

- Multiple users are making updates to the database at the same time.

Problem:

- Multiple people updating the database simultaneously can override each other

Example:

- Agents T1 & T2 simultaneously read the seats reserved on Flight 890 i.e. 80
- T1 cancels 5 seats updating the seats reserved on Flight 890 to 75
- T2 reserves 4 additional seats on the flight and updates the seats reserved on Flight 890 to 84.
- If T1 updates the database before T2. T2 will override T1's change and make reservations to 84 rather than getting the correct value of 79.
- Similarly if T2 updates before T1 the seats reserved will be 75

Shared Updates: Solution

- Batch Processing
  - Allow multiple users to retrieve data simultaneously
  - Updates are added to a batch file which does the appropriate processing
  - Does not work for real time situations
- Locking
  - Restrict access to the record being updated by a user till the transaction is complete.

Two Phase Lock

- Required when multiple records are updated as a result of a user action (e.g. filling form etc.)
- All the records accessed are locked progressively till the required updates are completed
  - Growing Phase: More and more locks are added without releasing locks
  - After all locks are placed the database is updated
  - Shrinking Phase: All locks are removed and no new ones are added

Deadlock

- When two transactions require a common set of records.
- Both of them are in growing phase and each locks some of the records
- None of the records are released and they wait for each other to release the locked records

They will wait forever!!!

Breaking Deadlock

Facilities

- Programs can lock entire tables or an individual row
- Programs can release any or all of the locks they currently hold
- Programs can inquire whether a given row or table is locked

Rules

- If more than one row is required then the entire table must be locked
- Limit the amount of wait for a lock to be released beyond which a transaction is aborted
- A well designed transaction should lock all the rows and tables before starting the transaction
- Users should release locks as soon as possible to improve the efficiency of the database

Security

- Protection against unauthorized access: either intentional or accidental.
- Three main features for protection
  - **Passwords:** Allows only authorized users to access the database. Access privileges can be provided based on access needs
  - **Encryption:** Encodes data to non-decipherable. Data decoded on demand to prevent hackers from accessing data
  - **Views:** Different snapshot of the data ensures that users only get access to data they need

Integrity

- Integrity Constraints are the conditions that data must satisfy during initial input & updates.
- There are four categories of constraints
  - Data Type
  - Legal Values
  - Format
  - Key Constraints
    - Entity Integrity Constraints (Primary Key)
      - Enforces the uniqueness of the primary key
    - Referential Integrity Constraints (Foreign Key)
      - Value of foreign key must match the value of primary key for some row in another table

Integrity: Solutions

- Ignore constraint
  - Undesirable as it can lead to inconsistent data
- Let user enforce the constraint
  - Undesirable since user mistakes can be disastrous
- Let programmer build the logic of constraints in the programs
  - Makes programs complex: harder to write, harder to maintain, and expensive
- Place burden on the DBMS.
  - Preferred way: Cost of DBMS development amortized over large user base, hence economical

Replication

- Duplication of data at multiple physical locations
- Each replica of the data can be changed independently
- Periodically the replicas update their data to the master database – this process is called synchronization

Disaster Planning: Backup & Recovery

- Database can be damaged in a number of ways
  - Power outage, disk crashes, floods, user errors
- Periodic backups limit the loss due to sudden failures
- Data can be recovered from the latest backup and the changes since the backup need to be done in either of two ways
  - Manually
  - From a catalog (if exists) recording all updates to the database since the last backup.

Catalog/Data Dictionary

- Contains information describing the database
  - Schema for the database
  - Characteristic for each field
  - Possible values for each field
  - Description of the data
  - Relationships
  - Description of the programs
- Data Dictionary is same as catalog but may contain wider set of information than catalog