CMU (Intro to database) - Lecture 1

Database

- It is an organized collection of interrelated data that models some aspects of the real world.

(Example: Albums of Artist)

Database Management Studio (DBMSs)

- It's a software program that manages databases.

Managing a database means providing the facility to create, manipulate, delete, and retrieve data from the database in the best performance way.

It also provides a security level, that allows only authorized users with specific permissions to deal with databases.

Flat File Issues

1. Data Integration

There's no guarantee that the data is consistent as name of objects in an entity differs from another, it can also store other datatypes in a column

2. Implementation

Searching for a specific record is so difficult accessing file from two developers 'd cause some problems such inconsistency and concurrence control

3. Durability

When manipulating data while the machine crashes down.

Data Model

- It is a high-level collection of concepts that describe the data stored in the database.

(Example: Relational Model, Key, Graph, Column-family, Array, Hierarchical, Network, Multi-Value)

Schema

- It is a Description of collection of data, using a data model.

Relational Model

- It is a type of data model that was introduced by Ted Codd in 1970 as he noticed that people were rewriting DBMSs every time, they want to change the physical layer.

The relational model has three keys:

- Store database in a simple data structure called relation.

Relation is an unordered set that contains the relationship of attributes that represent entities.

- Access database through high-level language.

Rather than telling DBMS explicitly how to retrieve data from a database, just tell it what you need, and DBMS will choose the best way to retrieve it.

- Physical storage of the data should be left up to DBMS for implementation

The relational data model defines three concepts:

• Structure: The definition of relations and their contents. These are the attributes the relations have and the values that those attributes can hold.

• Integrity: Ensure the database's contents satisfy constraints.

An example constraint would be that any value for the year attribute has to be a number.

• Manipulation: How to access and modify a database's contents.

Primary Key

- An attribute that uniquely identifies each tuple in the relation, some of DBMS automatically generate this attribute.

Foreign Key

Specifies that an attribute from one relation has to map to a tuple in another relation.

Data manipulation Languages (DML)

- Methods to store and retrieve information from the database.

Procedural: the query specifies the strategy that the DBMS should use to retrieve data from database (how to retrieve).

Non-procedural (Declarative):

The query only tells the DBMS what you want, and DBMS determines the best way to retrieve it.

Relational Algebra

- It is a set of fundamental operations to retrieve and manipulate tuples in a relation, each operator takes a relation or more than one as an input and retrieves a new relation as an output. In complex queries, you can use more than one operator as a chain.

- Select

Retrieve a subset of the tuples that satisfy a certain predicate. Maps to where clause in SQL.

- Projection

Generate a relation with tuples that contain only specified attributes which can rearrange the attributes order and manipulate its values. Maps to select clause in SQL.

- Union

Generate a new relation containing all tuples that appear in only one or more input relations. Maps to Union All in SQL.

- Intersection

Generate a new relation contains only tuples that appear in all input relations. Maps to Intersect clause in SQL.

- Difference

Generate a new relation containing tuples that only appear in the first input relation and not in the second input relation. (R - S). Maps to Except clause in SQL.

- Product

Generate a new relation contains all possible combinations of tuples from the input relations. (R * S). Maps to Cross Join in SQL.

- Join

Generate a new relation containing all tuples that are a combination of two tuples (one from each input relation). Maps to Join Clause in SQL.