

CMU 15-445 - Database Systems Lecture 1 - Introduction & Relational Model

Introduction & Relational Model

DBMSs

A database is an organized collection of related data that models some aspect of the real world (e.g., modeling the students in a class, or a digital music store). A DBMS is a software that allows applications to store and analyze the information in a database. It allows the definition, creation, querying, updating, and administration of databases.

DBMSs have 2 layers - the logical layer and the physical layer. The logical layer defines which entities and attributes the data stored in the database will have, and the physical layer is how those entities are actually stored (like the particular data structure used to store the data).

Before relational databases, the physical layer was defined in the application code itself, which made the database much harder to manage. If the structure of the data or the requirements of the application changed, the whole physical layer needed to be rewritten.

Relational Model

Ted Codd proposed the relational model in 1970. The relational model has 3 key points -

1. Store database in simple data structures.
2. Access data through a high-level query language.
3. The physical storage of data is left up to the implementation.

The relational model is an example of a *data model*. A data model is a collection of concepts for describing the data stored in a database. A *schema* is a description of a particular collection of data using a particular data model.

The relational model defines three concepts - 1. Structure - Data is distributed into

tables and tables can be related to each other. The structure also describes what kind or type of data each table can hold. 2. Integrity - The database's contents always satisfy some constraints, which can either be defined by the designer or the database itself. 3. Manipulation - Describes how the contents of the database can be modified.

Relational Algebra

Relational algebra gives a set of operations we can perform on relations to produce new relations. You can think of a relation as a table in a database. Relational algebra operators take a relation as an input and produce another relation.

The basic few relational algebra operators are -

1. Select ($\sigma_{\text{predicate}}(R)$) - Acts like a filter. Selects tuples from the relation that satisfy the predicate.
2. Projection ($\pi_{A1, A2, \dots}(R)$) - Returns the same relation but only with the attributes $A1, A2, \dots$ selected from the relation.
3. Union ($R \cup S$) - Produces a new relation that contains all the tuples in at least one of the relations (the relations need to have the same attributes).
4. Intersection ($R \cap S$) - Produces a new relation that contains only those tuples that exist in both R and S (the relations need to have the same attributes).
5. Difference ($R - S$) - Produces a new relation that contains all tuples that appear in the first relation but not in the second relation (the relations need to have the same attributes).
6. Product ($R \times S$) - Product takes in two relations and outputs a relation that contains all possible combinations for tuples from the input relations.
7. Join ($R \bowtie S$) - outputs a relation that contains all the tuples that are a combination of two tuples where for each attribute that the two relations share, the values for that attribute of both tuples is the same.