Data and Its Applications (CS4.301)

Monsoon 2021, IIIT Hyderabad
16 September, Thursday (Lecture 8)

Taught by Prof. Kamal Karlapalem

Relational Data Model (contd.)

Relational Algebra

The relational algebra describes a set of operators used to manipulate relations. There are two kinds of operators: *set operators* (union, intersection, difference and cartesian product) and *database-specific operators* (select, project and join).

The select operator preserves the columns (attributes) of its input; it filters only the rows (tuples). If we write $\sigma_c(R(A_1, \dots, A_n)) = R'(A_1, \dots, A_n)$, then the rows of R' are exactly those rows of R which satisfy the condition c. In SQL, this is expressed as `SELECT * FROM <table> WHERE c`.

The project operator, conversely, preserves the rows (tuples) and returns only some of columns (attributes). For example, $\pi_{(A_{i_1},\dots,A_{i_k})}(R(A_1, \dots, A_n)) = R'(A_{i_1}, \dots, A_{i_k})$. This is expressed as the SQL command `SELECT Ai1, ..., Aik FROM <table>`.
Note that in the general case, only distinct rows are included in the projection. It remains a proper set, like the input.
We write $\pi_k(R)$ to denote the projection of only the key attribute of R.

The join operator is equivalent to the cartesian product, followed by select. For example, $R \bowtie_c S = \sigma_c(R \times S)$. If the condition c involves equality of some attributes, it is called an equijoin.
For an equijoin, if the condition is $A = B$, then a natural join (denoted $R\ S$) does not include the attribute B.

The set operations of union, intersection and difference require that the operands be *compatible*, *i.e.*, if the operation is performed on $R(A_1, \dots, A_n)$ and $S(B_1, \dots, B_n)$, then $\text{dom}(A_i) = \text{dom}(B_i)$ for all i.
The cartesian product combines the tuples of two relations. It returns a relation with all the attributes of both its operands, *i.e.*, $R_1(A_1, \dots, A_n) \times R_2(B_1, \dots, B_n) = R(A_1, \dots, A_n, B_1, \dots, B_n)$.