**Unary Relational Operations: SELECT and PROJECT**

**The SELECT Operation**

The SELECT operation is used to choose a subset of the tuples from a relation that satisfies a selection condition.

The SELECT operation can also be visualized as a horizontal partition of the relation into two sets of tuples. Those tuples that satisfy the condition and are selected, and those tuples that do not satisfy the condition and are filtered out.

For example, to select the EMPLOYEE tuples whose department is 4, or those whose salary is greater than $30,000, we can individually specify each of these two conditions with a SELECT operation as follows:

σDno=4(EMPLOYEE)

σSalary>30000(EMPLOYEE)

In general, the SELECT operation is denoted by

σ<selection condition>(R)

where the symbol σ (sigma) is used to denote the SELECT operator and the selection condition is a Boolean expression (condition) specified on the attributes of relation R.

The relation resulting from the SELECT operation has the same attributes as R.

The Boolean expression specified in <selection condition> is made up of a number of clauses of the form

<attribute name> <comparison op> <constant value>

or

<attribute name> <comparison op> <attribute name>

where <attribute name> is the name of an attribute of R, <comparison op> is normally one of the operators {=, <, ≤, >, ≥, ≠}, and <constant value> is a constant value from the attribute domain.

Clauses can be connected by the standard Boolean operators and, or, and not to form a general selection condition.

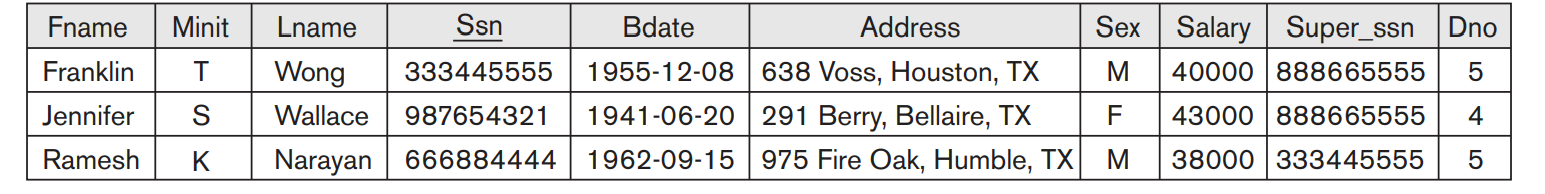
For example, to select the tuples for all employees who either work in department 4 and make over

$25,000 per year, or work in department 5 and make over $30,000,

we can specify the following SELECT operation:

σ(Dno=4 AND Salary>25000) OR (Dno=5 AND Salary>30000)(EMPLOYEE)

The result is shown as follows



The Boolean conditions AND, OR, and NOT have their normal interpretation, as follows:

■ (cond1 AND cond2) is TRUE if both (cond1) and (cond2) are TRUE; otherwise, it is FALSE.

■ (cond1 OR cond2) is TRUE if either (cond1) or (cond2) or both are TRUE; otherwise, it is FALSE.

■ (NOT cond) is TRUE if cond is FALSE; otherwise, it is FALSE

The SELECT operation is commutative;

that is, σ<cond1>(σ<cond2>(R)) = σ<cond2>(σ<cond1>(R))

Hence, a sequence of SELECTs can be applied in any order.

In SQL, the SELECT condition is typically specified in the WHERE clause of a query.

For example, the following operation:

σDno=4 AND Salary>25000 (EMPLOYEE)

would correspond to the following SQL query:

SELECT \* FROM EMPLOYEE WHERE Dno=4 AND Salary>25000;

**The PROJECT Operation**

If we think of a relation as a table, the SELECT operation chooses some of the rows from the table while discarding other rows. The PROJECT operation, on the other hand, selects certain columns from the table and discards the other columns.

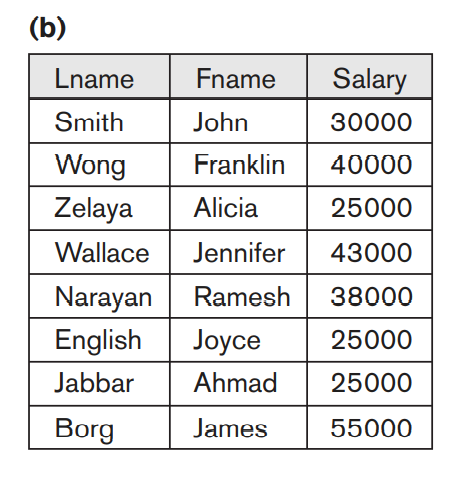
If we are interested in only certain attributes of a relation, we use the PROJECT operation to project the relation over these attributes only.

Therefore, the result of the PROJECT operation can be visualized as a vertical partition of the relation into two relations: one has the needed columns (attributes) and contains the result of the operation, and the other contains the discarded columns.

For example, to list each employee’s first and last name and salary, we can use the PROJECT operation as follows:

π **Lname, Fname, Salary**(EMPLOYEE)

The resulting relation is as follows



The general form of the PROJECT operation is

π<attribute list>(R)

where π (pi) is the symbol used to represent the PROJECT operation, and <attribute list> is the desired sub list of attributes from the attributes of relation R.

The result of the PROJECT operation has only the attributes specified in <attribute list> in the same order as they appear in the list. Hence, its degree is equal to the number of attributes in <attribute list>.

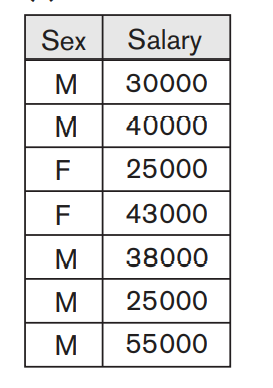
**Duplicate elimination**

The PROJECT operation removes any duplicate tuples, so the result of the PROJECT operation is a set of distinct tuples, and hence a valid relation. This is known as duplicate elimination.

For example, consider the following PROJECT operation:

πSex, Salary(EMPLOYEE)

The result will be as follows



In SQL, the PROJECT attribute list is specified in the SELECT clause of a query.

For example, the following operation:

πSex, Salary(EMPLOYEE)

would correspond to the following SQL query:

SELECT DISTINCT Sex, Salary FROM EMPLOYEE

If we remove the keyword DISTINCT from this SQL query, then duplicates will not be eliminated.