

Relational algebra operators - projection

Relational algebra is the mathematical basis for performing queries against a relational database. Operations are performed against relations – resulting in relations. Because the result of relational algebra operation is a relation, operations can be stacked up against each other. More on this as we go forward.

For the following examples, we are going to be using a relation Employees with the following schema:

Employee (Emp_id:int, Emp_name:string, Emp_office:int)

Projection

Let's start with projection, which is represented by the symbol π (pi). With the projection operation, you can select attributes to be pulled out of the relation. Let's look at the following relation instance of Employee:

Employee		
Emp_id	Emp_name	Emp_office
1001	Bob	10
1002	Alice	11
1003	Sandy	10
1004	Larry	11
1005	Susan	11

If we apply the projection operation to this instance of Employee, we are saying we want a specific set of attributes out of Employee relation. The statement:

$\Pi_{\text{Emp_id}}(\text{Employee})$ - which is also written as: $\Pi_{\text{Emp_id}}\text{Employee}$

is saying, project out the field Emp_id from this instance of the relation Employee. The result of this operation is a relation that looks like this:

Emp_id
1001
1002
1003
1004
1005

The operation basically said "Build a new relation that consists of only the employee ids from the original relation called Employee.

If the operation had looked like this:

$\Pi_{\text{Emp_id, Emp_name}} \text{Employee}$

The resulting relation would look like:

Emp_id	Emp_name
1001	Bob
1002	Alice
1003	Sandy
1004	Larry
1005	Susan

Duplicates

One thing to remember about a relation is that each set of tuples is unique – i.e. there are no duplicates. In the above two examples this isn't an issue – however if we were to execute the following operation:

$\Pi_{\text{Emp_office}} \text{Employee}$

The resulting relation is:

Emp_office
10
11

Because the resulting relation cannot have duplicates. Note that in SQL, when you perform the select statement (which is where projection manifests itself), this is not usually the case. You must use the DISTINCT clause to eliminate the duplicates. But for relational algebra – elimination of duplicates is assumed.