Relational Algebra

- The relational algebra is a procedural language
- The relational algebra consists of six basic operators
  - select: σ
  - project: π
  - union: $\cup$
  - set difference: $-$
  - Cartesian product: $\times$
  - rename: ρ
- The operators take one or two relations as inputs and produce a new relation as a result.
- This property makes the algebra **closed** (i.e., all objects in the relational algebra are relations).

Select Operation

- **Notation**: $\sigma_p(r)$
- p is called the **selection predicate**
- **Definition**: $t \in \sigma_p(r) \Leftrightarrow t \in r \wedge p(t)$
- p is a condition in propositional calculus consisting of **terms** connected by : $\wedge$ (**and**), $\vee$ (**or**), $\neg$ (**not**)
- Example: $\sigma_{BranchName=`Perryridge`}(account)$
- Example: $\sigma_{A=B \wedge D>5}(r)$

r

A	B	C	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

$\sigma_{A=B \wedge D>5}(r)$

A	B	C	D
α	α	1	7
β	β	23	10

Project Operation

- **Notation**: $\pi_{A_1,\ldots,A_k}(r)$
- The result is defined as the relation of k columns obtained by deleting the columns that are not listed
- **Definition**: $t \in \pi_{A_1,\ldots,A_k}(r) \Leftrightarrow \exists x(x \in r \wedge t = x[A_1, \ldots, A_k])$
- There are no duplicate rows in the result since relations are sets
- Example: $\pi_{AccNr,Balance}(account)$
- Example: $\pi_{A,C}(r)$

r

A	B	C
α	10	1
α	20	1
β	30	1
β	40	2

$\pi_{A,C}(r)$

A	C
α	1
β	1
β	2

Union Operation

- **Notation**: $r \cup s$
- **Definition**: $t \in (r \cup s) \Leftrightarrow t \in r \vee t \in s$
- For $r \cup s$ to be valid r and s must have the same schema (i.e., attributes).
- Example: $\pi_{CustName}(depositor) \cup \pi_{CustName}(borrower)$
- Example: $r \cup s$

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

$r \cup s$

A	B
α	1
α	2
β	1
β	3

Set Difference Operation

- **Notation**: $r - s$
- **Definition**: $t \in (r - s) \Leftrightarrow t \in r \wedge t \notin s$
- Set differences must be taken between (union) compatible relations.
  - r and s must have the same **arity**
  - attribute domains of r and s must be compatible
- Example: $r - s$

r

A	B
α	1
α	2
β	1

s

A	B
α	2
β	3

$r - s$

A	B
α	1
β	1

Cartesian Product Operation

- **Notation**: $r \times s$
- **Definition**: $t \in (r \times s) \Leftrightarrow x \in r \wedge y \in s \wedge t = x \circ y$
- We assume that the attribute names of r and s are disjoint. If the attribute names are not disjoint, then renaming must be used.
- Example: $r \times s$

r

A	B
α	1
β	2

s

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

$r \times s$

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

Rename Operation

- Allows us to name the results of relational algebra expressions by setting relation and attribute names.
- The rename operator is also used if there are name clashes.
- Various flavors:
  - $\rho_r(E)$ changes the relation name to r.
  - $\rho_{r(A_1,\ldots,A_n)}(E)$ changes the relation name to r and the attribute names to $A_1, \ldots, A_k$.
  - $\rho_{(A_1,\ldots,A_n)}(E)$ changes attribute names to $A_1, \ldots, A_k$.
- Example: $\rho_{s(X,Y,U,V)}(r)$

r

A	B	C	D
α	α	1	7
β	β	23	10

s

X	Y	U	V
α	α	1	7
β	β	23	10

Natural Join

- **Notation**: $r \bowtie s$
- Let r and s be relations on schemas R and S, respectively.
- Attributes that occur in r and s must be identical.
- $r \bowtie s$ is a relation on a schema that includes all attributes from schema R and all attributes from schema S that do not occur in schema R.
- Example:
  - $r \bowtie s$ with $R(A, B, C, D)$ and $S(E, B, D)$
  - Schema of result is (A, B, C, D, E)
  - Equivalent to: $\pi_{A,B,C,D,E}(\sigma_{B=Y \wedge D=Z}(r \times \rho_{(E,Y,Z)}(s)))$

r

A	B	C	D
α	1	α	a
β	2	γ	a
γ	4	β	b
α	1	γ	a
δ	2	β	b

s

B	D	E
1	a	α
3	a	β
1	a	γ
2	b	δ
3	b	ϵ

$r \bowtie s$

A	B	C	D	E
α	1	α	a	α
α	1	α	a	γ
α	1	γ	a	α
α	1	γ	a	γ
δ	2	β	b	δ