Relational Algebra

- SQL is based on Relational Algebra
- SQL is “syntactically sugared” expressions of Relational Algebra
- The need for a query language
  - Simple
  - Can be optimized
  - Independent of programming languages

Algebra Vs Relational Algebra

- Algebra
  - Operands
  - Operators
  - Expressions
- Relational Algebra
  - Variables (Relations)
  - Constants (Finite Relations)
  - Operations
    - Set Operations (Union, Intersection, Difference)
    - Selection and Projection
    - Combining relations
    - Renaming relations

Set operations

- If R and S are two relations,
  - Union is denoted by R U S
  - Intersection is denoted by R ∩ S
  - Difference is denoted by R – S

- Conditions on R and S to apply set operations
  - R and S must have the same attributes that belong to the same domains.
  - Order of the attributes must be the same.

R U S

name	*address*	*gender*	*birthdate*
Carrie Fisher	123 Maple St., Hollywood	F	9/9/99
Mark Hamill	456 Oak Rd., Brentwood	M	8/8/88

Relation R

name	*address*	*gender*	*birthdate*
Carrie Fisher	123 Maple St., Hollywood	F	9/9/99
Harrison Ford	789 Palm Dr., Beverly Hills	M	7/7/77

Relation S

name	*address*	*gender*	*birthdate*
Carrie Fisher	123 Maple St., Hollywood	F	9/9/99
Mark Hamill	456 Oak Rd., Brentwood	M	8/8/88
Harrison Ford	789 Palm Dr., Beverly Hills	M	7/7/77

R ∩ S

name	*address*	*gender*	*birthdate*
Carrie Fisher	123 Maple St., Hollywood	F	9/9/99
Mark Hamill	456 Oak Rd., Brentwood	M	8/8/88

Relation R

name	*address*	*gender*	*birthdate*
Carrie Fisher	123 Maple St., Hollywood	F	9/9/99
Harrison Ford	789 Palm Dr., Beverly Hills	M	7/7/77

Relation S

name	*address*	*gender*	*birthdate*
Carrie Fisher	123 Maple St., Hollywood	F	9/9/99

R - S

name	*address*	*gender*	*birthdate*
Carrie Fisher	123 Maple St., Hollywood	F	9/9/99
Mark Hamill	456 Oak Rd., Brentwood	M	8/8/88

Relation R

name	*address*	*gender*	*birthdate*
Carrie Fisher	123 Maple St., Hollywood	F	9/9/99
Harrison Ford	789 Palm Dr., Beverly Hills	M	7/7/77

Relation S

name	*address*	*gender*	*birthdate*
Mark Hamill	456 Oak Rd., Brentwood	M	8/8/88

Projection

- Use projection to create a new relation with selected columns from a relation
- Analogous to SELECT in SQL
- Represented as $\prod_{a1, a2, ..}(R)$

title	*year*	*length*	*genre*	*studioName*	*producerC#*
Star Wars	1977	124	sciFi	Fox	12345
Galaxy Quest	1999	104	comedy	DreamWorks	67890
Wayne's World	1992	95	comedy	Paramount	99999

Figure 2.13: The relation Movies

title	*year*	*length*
Star Wars	1977	124
Galaxy Quest	1999	104
Wayne's World	1992	95

$\prod_{\text{title, year length}}(\text{Movies})$

Selection

- Selection operator returns the tuples that meet a condition.
- Analogous to WHERE in SQL
- Represented by $\sigma_c(R)$, where c is the condition (σ is pronounced sigma)

title	*year*	*length*	*genre*	*studioName*	*producerC#*
Star Wars	1977	124	sciFi	Fox	12345
Galaxy Quest	1999	104	comedy	DreamWorks	67890
Wayne's World	1992	95	comedy	Paramount	99999

Figure 2.13: The relation Movies

title	*year*	*length*	*genre*	*studioName*	*producerC#*
Star Wars	1977	124	sciFi	Fox	12345
Galaxy Quest	1999	104	comedy	DreamWorks	67890

$\sigma_{\text{length} \geq 100}$ (Movies)

Cartesian Product

- The Cartesian product (or cross-product , or just product) of two sets R and S is the set of pairs that can be formed by choosing the first element of the pair to be any element of R and the second any element of S.
- Denoted by R X S
- Notice the use of qualifier

A	B
1	2
3	4

(a) Relation R

B	C	D
2	5	6
4	7	8
9	10	11

(b) Relation S

A	$R.B$	$S.B$	C	D
1	2	2	5	6
1	2	4	7	8
1	2	9	10	11
3	4	2	5	6
3	4	4	7	8
3	4	9	10	11

(c) Result $R \times S$

Natural Join

- The simplest sort of match is the natural join of two relations R and S , denoted R ⋈ S , in which we pair only those tuples from R and S that agree in whatever attributes are common to the schemas of R and S.
- Same as inner join in SQL.
- The two relations may have more than one attribute in common.

A	B	C	D
1	2	5	6
3	4	7	8

R ⋈ S

A	B
1	2
3	4

(a) Relation R

B	C	D
2	5	6
4	7	8
9	10	11

(b) Relation S

Theta Join

- The natural join forces us to pair tuples using one specific condition .
- Theta join allows you to pair tuples from two relations on some other basis.
- Denoted by $R \bowtie_C S$
- Theta Join
  - Take the product of R and S .
  - Select from the product only those tuples that satisfy the condition C .

Theta Join Example - 1

A	B	C
1	2	3
6	7	8
9	7	8

(a) Relation U

B	C	D
2	3	4
2	3	5
7	8	10

(b) Relation V

A	$U.B$	$U.C$	$V.B$	$V.C$	D
1	2	3	2	3	4
1	2	3	2	3	5
1	2	3	7	8	10
6	7	8	7	8	10
9	7	8	7	8	10

Figure 2.17: Result of $U \bowtie_{A<D} V$

Theta Join Example - 2

A	$U.B$	$U.C$	$V.B$	$V.C$	D
1	2	3	7	8	10

$$U \bowtie_{A<D \text{ AND } U.B \neq V.B} V$$

A	B	C
1	2	3
6	7	8
9	7	8

(a) Relation U

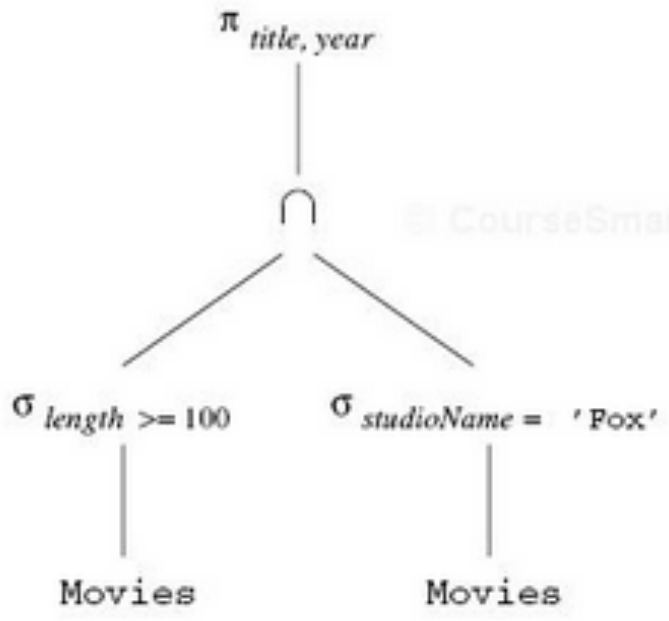
B	C	D
2	3	4
2	3	5
7	8	10

(b) Relation V

Expression Trees

- Combine expressions to solve problems
- Visualize using an expression tree

Example: What are the titles and years of movies made by Fox that are at least 100 minutes long ?



Alternate forms of expressions

$$\pi_{title,year}\left(\sigma_{length\geq 100}(\texttt{Movies}) \cap \sigma_{studioName=\texttt{'Fox'}}(\texttt{Movies})\right)$$

$$\pi_{title,year}\left(\sigma_{length\geq 100\ \texttt{AND}\ studioName=\texttt{'Fox'}}(\texttt{Movies})\right)$$

Naming and Renaming

- Attributes or relations can be renamed to avoid name conflicts and to simplify expressions.
- Denoted by $\rho_S(R)$ (ρ is pronounced rho)
- S is the name of the new relation

A	B	X	C	D
1	2	2	5	6
1	2	4	7	8
1	2	9	10	11
3	4	2	5	6
3	4	4	7	8
3	4	9	10	11

Figure 2.19: $R \times \rho_{S(X,C,D)}(S)$

A	B
1	2
3	4

(a) Relation R

B	C	D
2	5	6
4	7	8
9	10	11

(b) Relation S

$$\rho_{RS(A,B,X,C,D)}(R \times S)$$

← Alternate form

Relationships among operations

- Intersection can be expressed as difference $R \cap S = R - (R - S)$
- Theta - join can be expressed by product and selection $R \bowtie_C S = \sigma_C(R \times S)$
- The natural join of R and S can be expressed by starting with the product R X S , then applying the selection operator with a condition C and L is the list of attributes

$$R \bowtie S = \pi_L\left(\sigma_C(R \times S)\right)$$

Linear notation

- Another type of notation like the trees
- Invent names for the temporary relations that correspond to the interior nodes of the tree and write a sequence of assignments that create a value for each .
- Notation:
  - A relation name and parenthesized list of attributes for that relation
  - The assignment symbol : = .
  - Any algebraic expression on the right.

$R(t,y,l,i,s,p) := \sigma_{length \geq 100}(Movies)$
$S(t,y,l,i,s,p) := \sigma_{studioName='Fox'}(Movies)$
$T(t,y,l,i,s,p) := R \cap S$
$Answer(title, year) := \pi_{t,y}(T)$

$R(t,y,l,i,s,p) := \sigma_{length \geq 100}(Movies)$
$S(t,y,l,i,s,p) := \sigma_{studioName='Fox'}(Movies)$
$Answer(title, year) := \pi_{t,y}(R \cap S)$