Relational Algebra

(Reference: Chapter 4 of Ramakrishnan & Gehrke)

Example Database

Movies

title	director	myear	rating
Fargo	Coen	1996	8.2
Raising Arizona	Coen	1987	7.6
Spiderman	Raimi	2002	7.4
Wonder Boys	Hanson	2000	7.6

Actors

actor	ayear
Cage	1964
Hanks	1956
Maguire	1975
McDormand	1957

Acts

actor	title
Cage	Raising Arizona
Maguire	Spiderman
Maguire	Wonder Boys
McDormand	Fargo
McDormand	Raising Arizona
McDormand	Wonder Boys

Directors

director	dyear
Coen	1954
Hanson	1945
Raimi	1959

Some Queries

- Find movies made after 1997
- Find movies made by Hanson after 1997
- Find all movies and their ratings
- Find all actors and directors
- Find Coen's movies with McDormand
- Find movies with Maguire but not McDormand
- Find actors who have acted in some Coen's movie
- Find (director, actor) pairs where the director is younger than the actor
- Find actors who have acted in all of Coen's movies

Relational Algebra

- A formal query language for asking questions
- A query is composed of a collection of operators called **relational operators**
- *Unary operators*: selection, projection, renaming
- *Binary operators*: union, intersect, difference, cartesian product, join
- Relations are closed under relational operators
- Operators can be *composed* to form **relational algebra expressions**

Selection: σ

- $\sigma_c(R)$ selects rows from relation R that satisfy *selection condition* c
- **Example:** Find movies made after 1997

Movies

title	director	myear	rating
Fargo	Coen	1996	8.2
Raising Arizona	Coen	1987	7.6
Spiderman	Raimi	2002	7.4
Wonder Boys	Hanson	2000	7.6

$\sigma_{myear>1997}$**(Movies)**

title	director	myear	rating
Spiderman	Raimi	2002	7.4
Wonder Boys	Hanson	2000	7.6

Selection Condition

- **Selection condition** is a boolean combination of *terms*
- A **term** is one of the following forms:
  1. attribute **op** constant
  2. attribute$_1$ **op** attribute$_2$
  3. term$_1$ $\wedge$ term$_2$
  4. term$_1$ $\vee$ term$_2$
  5. $\neg$ term$_1$
  6. (term$_1$)

op $\in \{=, \neq, <, \leq, >, \geq\}$

- Operator precedence: (), *op*, $\neg$, $\wedge$, $\vee$
- **Examples**:

Selection Condition (cont.)

- **Example:** Find movies made by Hanson after 1997

Movies

title	director	myear	rating
Fargo	Coen	1996	8.2
Raising Arizona	Coen	1987	7.6
Spiderman	Raimi	2002	7.4
Wonder Boys	Hanson	2000	7.6

$\sigma_{myear>1997 \ \wedge \ director='Hanson'}$**(Movies)**

title	director	myear	rating
Wonder Boys	Hanson	2000	7.6

Projection: π

- $\pi_L(R)$ projects columns given by list L from relation R
- **Example**: Find all movies and their ratings

Movies

title	director	myear	rating
Fargo	Coen	1996	8.2
Raising Arizona	Coen	1987	7.6
Spiderman	Raimi	2002	7.4
Wonder Boys	Hanson	2000	7.6

$\pi_{title, \ rating}$**(Movies)**

title	rating
Fargo	8.2
Raising Arizona	7.6
Spiderman	7.4
Wonder Boys	7.6

Renaming: ρ

- Given relation $R(A, B, C)$, $\rho_{S(X,Y,Z)}(R)$ renames it to $S(X, Y, Z)$

Actors

actor	ayear
Cage	1964
Hanks	1956
Maguire	1975
McDormand	1957

$\rho_{Stars(name,yob)}$**(Actors)**

Stars

name	yob
Cage	1964
Hanks	1956
Maguire	1975
McDormand	1957

Set Operations

- **Union**: $R \cup S$ returns a relation containing all tuples that occur in R or S (or both)
- **Intersection**: $R \cap S$ returns a relation containing all tuples that occur in both R and S
- **Set-difference**: $R - S$ returns a relation containing all tuples in R but not in S
- Two relations are **union compatible** if
  - they have the same arity, and
  - the corresponding attributes have same domains
- union ($\cup$), intersection ($\cap$), and set-difference ($-$) operators require input relations to be union compatible

Actors

actor	ayear
Cage	1964
Hanks	1956
Maguire	1975
McDormand	1957

π_{actor}**(Actors)**

actor
Cage
Hanks
Maguire
McDormand

Directors

director	dyear
Coen	1954
Hanson	1945
Raimi	1959

$\pi_{director}$**(Directors)**

director
Coen
Raimi
Hanson

$\pi_{actor}(Actors) \cup \pi_{director}(Directors)$

actor
Cage
Hanks
Maguire
McDormand
Coen
Raimi
Hanson

Union Example:

Find all actors & directors

$\pi_{actor}(Actors) \cup \pi_{director}(Directors)$

Acts

actor	title
Cage	Raising Arizona
Maguire	Spiderman
Maguire	Wonder Boys
McDormand	Fargo
McDormand	Raising Arizona
McDormand	Wonder Boys

Movies

title	director	myear	rating
Fargo	Coen	1996	8.2
Raising Arizona	Coen	1987	7.6
Spiderman	Raimi	2002	7.4
Wonder Boys	Hanson	2000	7.6

e_1

title
Fargo
Raising Arizona
Wonder Boys

e_2

title
Fargo
Raising Arizona

$e_1 \cap e_2$

title
Fargo
Raising Arizona

Intersection Example:

Find Coen's movies with McDormand

$e_1 = \pi_{title}(\sigma_{actor='McDormand'}(Acts))$

$e_2 = \pi_{title}(\sigma_{director='Coen'}(Movies))$

result $= e_1 \cap e_2$

Acts

actor	title
Cage	Raising Arizona
Maguire	Spiderman
Maguire	Wonder Boys
McDormand	Fargo
McDormand	Raising Arizona
McDormand	Wonder Boys

$\sigma_{actor='McDormand'}$ **(Acts)**

$\sigma_{actor='Maguire'}$ **(Acts)**

title
Spiderman
Wonder Boys

title
Fargo
Raising Arizona
Wonder Boys

$\sigma_{actor='Maguire'}$ **(Acts) -** $\sigma_{actor='McDormand'}$**(Acts)**

title
Spiderman

Set-difference Example:

Find movies with Maguire but not McDormand

$\sigma_{actor='Maguire'}$**(Acts)** - $\sigma_{actor='McDormand'}$**(Acts)**

Set Operations (cont.)

- Consider $R(A, B, C)$ and $S(X, Y)$
- **Cross-product**: $R \times S$ returns a relation with attribute list (A, B, C, X, Y) defined as follows:

$$R \times S = \{(a, b, c, x, y) \mid (a, b, c) \in R,\ (x, y) \in S\}$$

- Cross-product operation is also known as **cartesian product**

Cross-product Example

- Find actors who have acted in some Coen's movies
- $e_1 = \rho_{T(title2)}(\pi_{title}(\sigma_{director='Coen'}(Movies)))$

Movies

title	director	myear	rating
Fargo	Coen	1996	8.2
Raising Arizona	Coen	1987	7.6
Spiderman	Raimi	2002	7.4
Wonder Boys	Hanson	2000	7.6

e_1 →

T

title2
Fargo
Raising Arizona

Cross-product Example (cont.)

$e_2 =$

Acts

actor	title
Cage	Raising Arizona
Maguire	Spiderman
Maguire	Wonder Boys
McDormand	Fargo
McDormand	Raising Arizona
McDormand	Wonder Boys

$\times$

T

title2
Fargo
Raising Arizona

$=$

actor	title	title2
Cage	Raising Arizona	Fargo
Cage	Raising Arizona	Raising Arizona
Maguire	Spiderman	Fargo
Maguire	Spiderman	Raising Arizona
Maguire	Wonder Boys	Fargo
Maguire	Wonder Boys	Raising Arizona
McDormand	Fargo	Fargo
McDormand	Fargo	Raising Arizona
McDormand	Raising Arizona	Fargo
McDormand	Raising Arizona	Raising Arizona
McDormand	Wonder Boys	Fargo
McDormand	Wonder Boys	Raising Arizona

Cross-product Example (cont.)

e_2

actor	title	title2
Cage	Raising Arizona	Fargo
Cage	Raising Arizona	Raising Arizona
Maguire	Spiderman	Fargo
Maguire	Spiderman	Raising Arizona
Maguire	Wonder Boys	Fargo
Maguire	Wonder Boys	Raising Arizona
McDormand	Fargo	Fargo
McDormand	Fargo	Raising Arizona
McDormand	Raising Arizona	Fargo
McDormand	Raising Arizona	Raising Arizona
McDormand	Wonder Boys	Fargo
McDormand	Wonder Boys	Raising Arizona

$e_3 = \sigma_{title=title2}(e_2)$

e_3

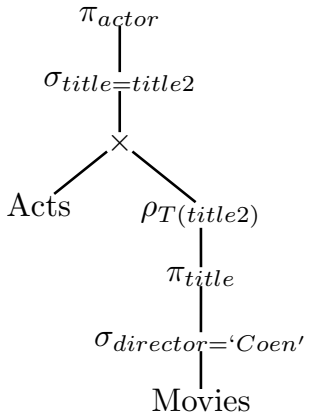
actor	title	title2
Cage	Raising Arizona	Raising Arizona
McDormand	Fargo	Fargo
McDormand	Raising Arizona	Raising Arizona

$\pi_{actor}(e_3)$

actor
Cage
McDormand

Cross-product Example (cont.)

- **Query**: Find actors who have acted in some Coen's movie
- **Answer**: $\pi_{actor}($ $\sigma_{title=title2}$ (Acts $\times$ $\rho_{T(title2)}($ $\pi_{title}($ $\sigma_{director='Coen'}$(Movies)))))



Join

- Combines cross-product, selection, and projection
- Join operator is more useful than the plain cross-product operator
- Three types of join:
  - Condition join
  - Equijoin
  - Natural join

Condition Join: $R \bowtie_c S$

- **Condition join** = Cross-product followed by selection

$$R \bowtie_c S = \sigma_c(R \times S)$$

- **Example**: Find (director,actor) pairs where the director is younger than the actor
- **Answer**: $\pi_{director,actor}($ Directors $\bowtie_{dyear>ayear}$ Actors$)$

Directors

director	dyear
Coen	1954
Hanson	1945
Raimi	1959

Actors

actor	ayear
Cage	1964
Hanks	1956
Maguire	1975
McDormand	1957

e_1 = Directors $\bowtie_{dyear>ayear}$ Actors

director	dyear	actor	ayear
Raimi	1959	Hanks	1956
Raimi	1959	McDormand	1957

$\pi_{director,actor}(e_1)$

director	actor
Raimi	Hanks
Raimi	McDormand

Equijoin: $R \bowtie_c S$

- **Equijoin** = Condition join of the form

$$R \bowtie_c S = \pi_L(\sigma_c(R \times S))$$

where

- c is a conjunction of equality conditions of the form $R.A_i = S.A_j$
- L is a sequence of attributes consisting of L_1 followed by L_2
- L_1 is a sequence of attributes in schema of R
- L_2 is a sequence of attributes in schema of S that are not referenced in c

Equijoin (cont.)

- **Example**: Find actors who have acted in some Coen's movie
- $\pi_{actor}(\ \sigma_{director='Coen'}\ (\ \text{Acts} \bowtie_{Acts.title\ =\ Movies.title} \text{Movies}\ )\ )$

e_1 = **Acts** $\bowtie_{Acts.title\ =\ Movies.title}$ **Movies**

actor	title	director	myear	rating
Cage	Raising Arizona	Coen	1987	7.6
Maguire	Spiderman	Raimi	2002	7.4
Maguire	Wonder Boys	Hanson	2000	7.6
McDormand	Fargo	Coen	1996	8.2
McDormand	Raising Arizona	Coen	1987	7.6
McDormand	Wonder Boys	Hanson	2000	7.6

$\pi_{actor}(\sigma_{director='Coen'}((e_1))$

actor
Cage
McDormand

Natural Join: $R \bowtie S$

- **Natural join** = Equijoin of the form

$$R \bowtie S = R \bowtie_c S$$

  where c is specified for all attributes having the same name in R and S

- **Example**: Find actors who have acted in some Coen's movie

$$\pi_{actor}(\sigma_{director='Coen'} (\text{Acts} \bowtie \text{Movies}))$$

- **Example**: Find the name and the year of birth of all actors who were in some Coen's movie

$$\pi_{actor,ayear}(\sigma_{director='Coen'} (\text{Movies}) \bowtie \text{Acts} \bowtie \text{Actors})$$

Example: Condition, Equi-, Natural Joins

R

A	B	X
0	6	x_1
1	9	x_2
2	7	x_3

S

A	B	Y
0	8	y_1
1	5	y_2
2	7	y_3

- $R \bowtie_{A=A' \wedge B<B'} \rho_{S'(A',B',Y)}(S)$

A	B	X	A'	B'	Y
0	6	x_1	0	8	y_1

- $R \bowtie_{A=A'} \rho_{S'(A',B',Y)}(S)$

A	B	X	B'	Y
0	6	x_1	8	y_1
1	9	x_2	5	y_2
2	7	x_3	7	y_3

- $R \bowtie S$

A	B	X	Y
2	7	x_3	y_3

Quiz

- **Query**: Find actors who have acted in all Coen's movies
- **CMovies** = $\pi_{title}(\sigma_{director='Coen'}(Movies))$

Movies

title	director	myear	rating
Fargo	Coen	1996	8.2
Raising Arizona	Coen	1987	7.6
Spiderman	Raimi	2002	7.4
Wonder Boys	Hanson	2000	7.6

CMovies

title
Fargo
Raising Arizona

Acts

actor	title
Cage	Raising Arizona
Maguire	Spiderman
Maguire	Wonder Boys
McDormand	Fargo
McDormand	Raising Arizona
McDormand	Wonder Boys

Summary

- **Relational algebra**: simple and powerful query language
- Basic operators: σ, π, $\cup$, $-$, $\times$
- Additional operators: ρ, $\cap$, $\bowtie$, $\div$
- Relational algebra is closed: operator's output is a relation
- Relational operators can be composed to form complex **relational algebra expressions**