

Join Operations

Theta, Equi, Natural, Outer and Semi

Relational Algebra operations

| | |
|-------------------|-----------|
| Selection | σ |
| Projection | π |
| Cartesian Product | \times |
| Union | \cup |
| Set Difference | $-$ |
| Join | \bowtie |
| Intersection | \cap |
| Division | \div |

We have seen Inner Join performed using CP and selection

Relational Algebra operations

In fact, there are 6 different types of Join

| | |
|--------------|-----------------|
| Theta join | \bowtie |
| Equi join | \bowtie |
| Natural join | \bowtie |
| Outer join | \ltimes |
| Semi join | \ltimes |
| Anti join | $\bar{\ltimes}$ |

Theta join

$R \bowtie_f S$

The resulting relation contains tuples that satisfy the predicate f from the cartesian product of relation R and relation S

$R \bowtie_f S$ is the same as $\sigma_f(R \times S)$

Often called an **inner join**

Emp / Dept example

$emp \bowtie_{emp.deptno=dept.deptno} dept$

| EMPNO | ENAME | JOB | MGR | HIREDATE | SAL | COMM | emp.DEPTNO | dept.DEPTNO | DNAME | LOC |
|-------|---------|-----------|-----|------------|-------|------|------------|-------------|------------|------------|
| 405 | MARCH | ADMIN | 938 | 13/06/1997 | 18000 | | 2 | 2 | RESEARCH | YORK |
| 535 | BYRNE | SALES | 734 | 15/08/1997 | 26000 | 300 | 3 | 3 | SALES | BIRMINGHAM |
| 557 | BELL | SALES | 734 | 26/03/2000 | 22500 | 500 | 3 | 3 | SALES | BIRMINGHAM |
| 602 | BIRD | MANAGER | 875 | 31/10/1997 | 39750 | | 2 | 2 | RESEARCH | YORK |
| 690 | AHMAD | SALES | 734 | 05/12/1997 | 22500 | 1400 | 3 | 3 | SALES | BIRMINGHAM |
| 734 | COX | MANAGER | 875 | 11/06/2002 | 38500 | | 3 | 3 | SALES | BIRMINGHAM |
| 818 | POLLARD | MANAGER | 875 | 14/05/2000 | 34500 | | 1 | 1 | ACCOUNTING | LONDON |
| 824 | REES | ANALYST | 602 | 05/03/2000 | 40000 | | 2 | 2 | RESEARCH | YORK |
| 875 | PARKER | PRESIDENT | | 09/07/2002 | 60000 | | 1 | 1 | ACCOUNTING | LONDON |
| 880 | TURNER | SALES | 734 | 04/06/2001 | 25000 | 0 | 3 | 3 | SALES | BIRMINGHAM |
| 912 | HAYES | ADMIN | 824 | 04/06/2001 | 21000 | | 2 | 2 | RESEARCH | YORK |
| 936 | CASSY | ADMIN | 734 | 23/07/2002 | 19500 | | 3 | 3 | SALES | BIRMINGHAM |
| 938 | GIBSON | ANALYST | 602 | 05/12/1997 | 40000 | | 2 | 2 | RESEARCH | YORK |
| 970 | BLACK | ADMIN | 818 | 21/11/1997 | 23000 | | 1 | 1 | ACCOUNTING | LONDON |

In SQL: Implicit Inner Join syntax

Can construct a theta join using **CP** and a predicate

```
select *
from emp, dept
where emp.deptno=dept.deptno
```

| EMPNO | ENAME | JOB | MGR | HIREDATE | SAL | COMM | emp.DEPTNO | dept.DEPTNO | DNAME | LOC |
|-------|---------|-----------|-----|------------|-------|------|------------|-------------|------------|------------|
| 405 | MARCH | ADMIN | 938 | 13/06/1997 | 18000 | | 2 | 2 | RESEARCH | YORK |
| 535 | BYRNE | SALES | 734 | 15/08/1997 | 26000 | 300 | 3 | 3 | SALES | BIRMINGHAM |
| 557 | BELL | SALES | 734 | 26/03/2000 | 22500 | 500 | 3 | 3 | SALES | BIRMINGHAM |
| 602 | BIRD | MANAGER | 875 | 31/10/1997 | 39750 | | 2 | 2 | RESEARCH | YORK |
| 690 | AHMAD | SALES | 734 | 05/12/1997 | 22500 | 1400 | 3 | 3 | SALES | BIRMINGHAM |
| 734 | COX | MANAGER | 875 | 11/06/2002 | 38500 | | 3 | 3 | SALES | BIRMINGHAM |
| 818 | POLLARD | MANAGER | 875 | 14/05/2000 | 34500 | | 1 | 1 | ACCOUNTING | LONDON |
| 824 | REES | ANALYST | 602 | 05/03/2000 | 40000 | | 2 | 2 | RESEARCH | YORK |
| 875 | PARKER | PRESIDENT | | 09/07/2002 | 60000 | | 1 | 1 | ACCOUNTING | LONDON |
| 880 | TURNER | SALES | 734 | 04/06/2001 | 25000 | 0 | 3 | 3 | SALES | BIRMINGHAM |
| 912 | HAYES | ADMIN | 824 | 04/06/2001 | 21000 | | 2 | 2 | RESEARCH | YORK |
| 936 | CASSY | ADMIN | 734 | 23/07/2002 | 19500 | | 3 | 3 | SALES | BIRMINGHAM |
| 938 | GIBSON | ANALYST | 602 | 05/12/1997 | 40000 | | 2 | 2 | RESEARCH | YORK |
| 970 | BLACK | ADMIN | 818 | 21/11/1997 | 23000 | | 1 | 1 | ACCOUNTING | LONDON |

SQL 92: Explicit Inner Join syntax

Can construct a theta join using specific inner join syntax

select * or expression
from relation inner join relation
on field operator field

table names go here

the predicate goes here

The on clause must be included

SQL 92: Explicit Inner Join syntax

Can construct a theta join using specific inner join syntax

select *
from emp inner join dept
on emp.deptno=dept.deptno

| EMPNO | ENAME | JOB | MGR | HIREDATE | SAL | COMM | emp.DEPTNO | dept.DEPTNO | DNAME | LOC |
|-------|---------|-----------|-----|------------|-------|------|------------|-------------|------------|------------|
| 405 | MARCH | ADMIN | 938 | 13/06/1997 | 18000 | | 2 | 2 | RESEARCH | YORK |
| 535 | BYRNE | SALES | 734 | 15/08/1997 | 26000 | 300 | 3 | 3 | SALES | BIRMINGHAM |
| 557 | BELL | SALES | 734 | 26/03/2000 | 22500 | 500 | 3 | 3 | SALES | BIRMINGHAM |
| 602 | BIRD | MANAGER | 875 | 31/10/1997 | 39750 | | 2 | 2 | RESEARCH | YORK |
| 690 | AHMAD | SALES | 734 | 05/12/1997 | 22500 | 1400 | 3 | 3 | SALES | BIRMINGHAM |
| 734 | COX | MANAGER | 875 | 11/06/2002 | 38500 | | 3 | 3 | SALES | BIRMINGHAM |
| 818 | POLLARD | MANAGER | 875 | 14/05/2000 | 34500 | | 1 | 1 | ACCOUNTING | LONDON |
| 824 | REES | ANALYST | 602 | 05/03/2000 | 40000 | | 2 | 2 | RESEARCH | YORK |
| 875 | PARKER | PRESIDENT | | 09/07/2002 | 60000 | | 1 | 1 | ACCOUNTING | LONDON |
| 880 | TURNER | SALES | 734 | 04/06/2001 | 25000 | 0 | 3 | 3 | SALES | BIRMINGHAM |
| 912 | HAYES | ADMIN | 824 | 04/06/2001 | 21000 | | 2 | 2 | RESEARCH | YORK |
| 936 | CASSY | ADMIN | 734 | 23/07/2002 | 19500 | | 3 | 3 | SALES | BIRMINGHAM |
| 938 | GIBSON | ANALYST | 602 | 05/12/1997 | 40000 | | 2 | 2 | RESEARCH | YORK |
| 970 | BLACK | ADMIN | 818 | 21/11/1997 | 23000 | | 1 | 1 | ACCOUNTING | LONDON |

Theta join

$R \bowtie_f S$ The predicate f may use any of the comparison operators $=, <, >, <=, >=, \neq$

Example: Theta join with $>=$ and $<=$

A high earner, is defined as someone who is earning salary in the top 10% of the grade they are in.

Show the $ename$, $grade$, sal and $hisal$ for all high earners

Example: Theta join with >= and <=

A high earner, is defined as someone who is earning salary in the top 10% of the grade they are in.

Show the *ename*, *grade*, *sal* and *hisal* for all high earners

Requires information in *emp* and *grade* tables to be combined

emp table

| EMPNO | ENAME | JOB | MGR | HIREDATE | SAL | COMM | DEPTNO |
|-------|---------|-----------|-----|------------|-------|------|--------|
| 405 | MARCH | ADMIN | 938 | 13/06/1997 | 18000 | | 2 |
| 535 | BYRNE | SALES | 734 | 15/08/1997 | 26000 | 300 | 3 |
| 557 | BELL | SALES | 734 | 26/03/2000 | 22500 | 500 | 3 |
| 602 | BIRD | MANAGER | 875 | 31/10/1997 | 39750 | | 2 |
| 690 | AHMAD | SALES | 734 | 05/12/1997 | 22500 | 1400 | 3 |
| 734 | COX | MANAGER | 875 | 11/06/2002 | 38500 | | 3 |
| 818 | POLLARD | MANAGER | 875 | 14/05/2000 | 34500 | | 1 |
| 824 | REES | ANALYST | 602 | 05/03/2000 | 40000 | | 2 |
| 875 | PARKER | PRESIDENT | | 09/07/2002 | 60000 | | 1 |
| 880 | TURNER | SALES | 734 | 04/06/2001 | 25000 | 0 | 3 |
| 912 | HAYES | ADMIN | 824 | 04/06/2001 | 21000 | | 2 |
| 936 | CASSY | ADMIN | 734 | 23/07/2002 | 19500 | | 3 |
| 938 | GIBSON | ANALYST | 602 | 05/03/2000 | | | 2 |
| 970 | BLACK | ADMIN | 818 | 21/07/2002 | | | 1 |

grade table

| GRADE | LOSAL | HISAL |
|-------|-------|-------|
| 1 | 17000 | 21999 |
| 2 | 22000 | 23999 |
| 3 | 24000 | 29999 |
| 4 | 30000 | 49999 |
| 5 | 50000 | 99999 |

Example: Theta join with >= and <=

Combine the *emp* relation with the *grade* relation

emp ⋈ *grade*

Example: Theta join with \geq and \leq

Add the predicate

losal

hisal

| DATE | SAL | COM |
|----------|-------|-----|
| '06/1997 | 18000 | |
| '08/1997 | 26000 | |
| '03/2000 | 22500 | |
| '10/1997 | 39750 | |
| '12/1997 | 22500 | |
| '06/2002 | 38500 | |
| '05/2000 | 34500 | |

90% 100%

90% of hisal \leq sal \leq hisal

| GRADE | LOSAL | HISAL |
|-------|-------|-------|
| 1 | 17000 | 21999 |
| 2 | 22000 | 23999 |
| 3 | 24000 | 29999 |
| 4 | 30000 | 49999 |
| 5 | 50000 | 99999 |

Example: Theta join with \geq and \leq

Combine the emp relation with the grade relation

Π ename, grade, sal, hisal (emp \bowtie grade)
 hisal/100*90 \leq sal and sal \leq hisal

| ename | grade | sal | hisal |
|-------|-------|-------|-------|
| BELL | 2 | 22500 | 23999 |
| AHMAD | 2 | 22500 | 23999 |
| HAYES | 1 | 21000 | 21999 |
| BLACK | 2 | 23000 | 23999 |

SQL

With implicit join syntax

```
select ename, grade, sal, hisal
from emp, grade
where hisal/100*90 <= emp.sal and emp.sal <= hisal
```

| ename | grade | sal | hisal |
|-------|-------|-------|-------|
| BELL | 2 | 22500 | 23999 |
| AHMAD | 2 | 22500 | 23999 |
| HAYES | 1 | 21000 | 21999 |
| BLACK | 2 | 23000 | 23999 |

SQL

With explicit join syntax

```
select ename, grade, sal, hisal  
from emp inner join grade  
on hisal/100*90 <= emp.sal and emp.sal <= hisal
```



Fails in many DBs- can't do explicit join syntax
without common columns

Equi join

$$R \bowtie_f S$$

A particular type of Theta join that only uses
equality in the predicate/condition

$$\text{emp} \bowtie \text{dept}$$

$\text{emp.deptno}=\text{dept.deptno}$

Natural join

$$R \bowtie S$$

The equi-join of relation **R** and **S** over all common
attributes. One occurrence of each common
attribute is eliminated from the result.

No condition required as the join occurs across
ALL common attributes

SQL

With explicit **natural** join syntax (rare support)

```
select *  
from emp natural join dept
```

No Access
support

| EMPNO | ENAME | JOB | MGR | HIREDATE | SAL | COMM | emp.DEPTNO | DNAME | LOC |
|-------|---------|-----------|-----|------------|-------|------|------------|------------|------------|
| 405 | MARCH | ADMIN | 938 | 13/06/1997 | 18000 | | 2 | RESEARCH | YORK |
| 535 | BYRNE | SALES | 734 | 15/08/1997 | 26000 | 300 | 3 | SALES | BIRMINGHAM |
| 557 | BELL | SALES | 734 | 26/03/2000 | 22500 | 500 | 3 | SALES | BIRMINGHAM |
| 602 | BIRD | MANAGER | 875 | 31/10/1997 | 39750 | | 2 | RESEARCH | YORK |
| 690 | AHMAD | SALES | 734 | 05/12/1997 | 22500 | 1400 | 3 | SALES | BIRMINGHAM |
| 734 | COX | MANAGER | 875 | 11/06/2002 | 38500 | | 3 | SALES | BIRMINGHAM |
| 818 | POLLARD | MANAGER | 875 | 14/05/2000 | 34500 | | 1 | ACCOUNTING | LONDON |
| 824 | REES | ANALYST | 602 | 05/03/2000 | 40000 | | 2 | RESEARCH | YORK |
| 875 | PARKER | PRESIDENT | | 09/07/2002 | 60000 | | 1 | ACCOUNTING | LONDON |
| 880 | TURNER | SALES | 734 | 04/06/2001 | 25000 | 0 | 3 | SALES | BIRMINGHAM |
| 912 | HAYES | ADMIN | 824 | 04/06/2001 | 21000 | | 2 | RESEARCH | YORK |
| 936 | CASSY | ADMIN | 734 | 23/07/2002 | 19500 | | 3 | SALES | BIRMINGHAM |
| 938 | GIBSON | ANALYST | 602 | 05/12/1997 | 40000 | | 2 | RESEARCH | YORK |
| 970 | BLACK | ADMIN | 818 | 21/11/1997 | 23000 | | 1 | ACCOUNTING | LONDON |

Outer join

Often when joining relations there are no matching values in the join columns. The condition used will exclude these values. To include values that don't match, use an **Outer Join**

$$R \underset{f}{\times} S$$

Right Outer Join

$$R \underset{f}{\times} S$$

Left Outer Join

Best explained by example

Show a list of all departments and the employees who work in them

First attempt using implicit join syntax

```
select dept.deptno, dname, loc, ename  
from emp, dept  
where emp.deptno=dept.deptno
```

Best explained by example

Show a list of all departments and the employees who work in them

| deptno | dname | loc | ename |
|--------|------------|------------|---------|
| 1 | ACCOUNTING | LONDON | BLACK |
| 1 | ACCOUNTING | LONDON | PARKER |
| 1 | ACCOUNTING | LONDON | POLLARD |
| 2 | RESEARCH | YORK | GIBSON |
| 2 | RESEARCH | YORK | HAYES |
| 2 | RESEARCH | YORK | REES |
| 2 | RESEARCH | YORK | BIRD |
| 2 | RESEARCH | YORK | MARCH |
| 3 | SALES | BIRMINGHAM | CASSY |
| 3 | SALES | BIRMINGHAM | TURNER |
| 3 | SALES | BIRMINGHAM | |
| 3 | SALES | BIRMINGHAM | |
| 3 | SALES | BIRMINGHAM | |

| DEPTNO | DNAME | LOC |
|--------|------------|------------|
| 1 | ACCOUNTING | LONDON |
| 2 | RESEARCH | YORK |
| 3 | SALES | BIRMINGHAM |
| 4 | OPERATIONS | LEEDS |

Fails to show deptno 4

Best explained by example

Show a list of all departments and the employees who work in them

```
select dept.deptno, dname, loc, ename  
from emp, dept  
where emp.deptno=dept.deptno
```

We need to include unmatched rows from the dept relation

A Right Outer Join

Right Outer join

Show a list of all departments and the employees who work in them

```
Emp  $\times$  Dept  
emp.deptno=dept.deptno
```

Includes unmatched tuples from the right

Explicit Outer Join syntax

Can construct an Outer join using explicit right or left join syntax

select * or expression
from relation left join relation
on field operator field

$$R \underset{f}{\bowtie} S$$

select * or expression
from relation right join relation
on field operator field

$$R \underset{f}{\bowtie} S$$

SQL

With explicit right outer join syntax

select dept.deptno, dname, loc, ename
from emp right join dept
on emp.deptno = dept.deptno

| deptno | dname | loc | ename |
|--------|------------|------------|---------|
| 1 | ACCOUNTING | LONDON | POLLARD |
| 1 | ACCOUNTING | LONDON | PARKER |
| 1 | ACCOUNTING | LONDON | BLACK |
| 2 | RESEARCH | YORK | MARCH |
| 2 | RESEARCH | YORK | BIRD |
| 2 | RESEARCH | YORK | REES |
| 2 | RESEARCH | YORK | HAYES |
| 2 | RESEARCH | YORK | GIBSON |
| 3 | SALES | BIRMINGHAM | BYRNE |
| 3 | SALES | BIRMINGHAM | BELL |
| 3 | SALES | BIRMINGHAM | AHMAD |
| 3 | SALES | BIRMINGHAM | COX |
| 3 | SALES | BIRMINGHAM | TURNER |
| 3 | SALES | BIRMINGHAM | CASSY |
| 4 | OPERATIONS | LEEDS | |

Full Outer join

Includes tuples from both sides where there are null values

$$R \underset{f}{\bowtie} S$$

Full Outer Join

Full Outer join

Change to emp table to demonstrate this - the president doesn't have a department

| EMPNO | ENAME | JOB | MGR | HIREDATE | SAL | COMM | DEPTNO |
|-------|---------|-----------|-----|------------|-------|------|--------|
| 405 | MARCH | ADMIN | 938 | 13/06/1997 | 18000 | | 2 |
| 535 | BYRNE | SALES | 734 | 15/08/1997 | 26000 | 300 | 3 |
| 557 | BELL | SALES | 734 | 26/03/2000 | 22500 | 500 | 3 |
| 602 | BIRD | MANAGER | 875 | 31/10/1997 | 39750 | | 2 |
| 690 | AHMAD | SALES | 734 | 05/12/1997 | 22500 | 1400 | 3 |
| 734 | COX | MANAGER | 875 | 11/06/2002 | 38500 | | 3 |
| 818 | POLLARD | MANAGER | 875 | 14/05/2000 | 34500 | | 1 |
| 824 | REES | ANALYST | 602 | 05/03/2000 | 40000 | | 2 |
| 875 | PARKER | PRESIDENT | | 09/07/2002 | 60000 | | . |
| 880 | TURNER | SALES | 734 | 04/06/2001 | 25000 | 0 | 3 |
| 912 | HAYES | ADMIN | 824 | 04/06/2001 | 21000 | | 2 |
| 936 | CASSY | ADMIN | 734 | 23/07/2002 | 19500 | | 3 |
| 938 | GIBSON | ANALYST | 602 | 05/12/1997 | 40000 | | 2 |
| 970 | BLACK | ADMIN | 818 | 21/11/1997 | 23000 | | 1 |

Full Outer join

Show a list of all departments and all the employees

select dept.deptno, dname, loc, ename
from emp, dept
where emp.deptno=dept.deptno

We lose both
department 4 and
the President

| deptno | dname | loc | ename |
|--------|------------|------------|---------|
| 1 | ACCOUNTING | LONDON | BLACK |
| 1 | ACCOUNTING | LONDON | PARKER |
| 1 | ACCOUNTING | LONDON | POLLARD |
| 2 | RESEARCH | YORK | GIBSON |
| 2 | RESEARCH | YORK | HAYES |
| 2 | RESEARCH | YORK | REES |
| 2 | RESEARCH | YORK | BIRD |
| 2 | RESEARCH | YORK | MARCH |
| 3 | SALES | BIRMINGHAM | CASSY |
| 3 | SALES | BIRMINGHAM | TURNER |
| 3 | SALES | BIRMINGHAM | COX |
| 3 | SALES | BIRMINGHAM | AHMAD |
| 3 | SALES | BIRMINGHAM | BELL |
| 3 | SALES | BIRMINGHAM | BYRNE |

Show a list of all departments and all the employees

select dept.deptno, dname, loc, ename
from emp full outer join dept
on emp.deptno = dept.deptno

emp ⋈ dept
emp.deptno=dept.deptno

| deptno | dname | loc | ename |
|--------|------------|------------|---------|
| | | | PARKER |
| 1 | ACCOUNTING | LONDON | BLACK |
| 1 | ACCOUNTING | LONDON | POLLARD |
| 2 | RESEARCH | YORK | BIRD |
| 2 | RESEARCH | YORK | GIBSON |
| 2 | RESEARCH | YORK | HAYES |
| 2 | RESEARCH | YORK | MARCH |
| 2 | RESEARCH | YORK | REES |
| 3 | SALES | BIRMINGHAM | AHMAD |
| 3 | SALES | BIRMINGHAM | BELL |
| 3 | SALES | BIRMINGHAM | BYRNE |
| 3 | SALES | BIRMINGHAM | CASSY |
| 3 | SALES | BIRMINGHAM | COX |
| 3 | SALES | BIRMINGHAM | TURNER |
| 4 | OPERATIONS | LEEDS | |

Full Outer join

However - not many databases support the full outer join syntax

Emulate it with a Union

$$R \underset{f}{\times} S \cup R \underset{f}{\times} S$$

Show a list of all departments and all the employees

```
select dept.deptno, dname, loc, ename  
from emp left join dept  
on emp.deptno = dept.deptno
```

UNION

```
select dept.deptno, dname, loc, ename  
from emp right join dept  
on emp.deptno = dept.deptno
```

| deptno | dname | loc | ename |
|--------|------------|------------|---------|
| | | | PARKER |
| 1 | ACCOUNTING | LONDON | BLACK |
| 1 | ACCOUNTING | LONDON | POLLARD |
| 2 | RESEARCH | YORK | BIRD |
| 2 | RESEARCH | YORK | GIBSON |
| 2 | RESEARCH | YORK | HAYES |
| 2 | RESEARCH | YORK | MARCH |
| 2 | RESEARCH | YORK | REES |
| 3 | SALES | BIRMINGHAM | AHMAD |
| 3 | SALES | BIRMINGHAM | BELL |
| 3 | SALES | BIRMINGHAM | BYRNE |
| 3 | SALES | BIRMINGHAM | CASSY |
| 3 | SALES | BIRMINGHAM | COX |
| 3 | SALES | BIRMINGHAM | TURNER |
| 4 | OPERATIONS | LEEDS | |

Semi join

Theta join, but only includes columns from one side

$$R \underset{f}{\times} S$$

Only includes
columns from R

Same symbol as
an outer join

Semi join

Show all staff details for staff who work in London

$emp \bowtie (\sigma_{loc='London'}(dept))$
 $emp.deptno=dept.deptno$

| EMPNO | ENAME | JOB | MGR | HIREDATE | SAL | COMM | DEPTNO |
|-------|---------|-----------|-----|------------|-------|------|--------|
| 818 | POLLARD | MANAGER | 875 | 14/05/2000 | 34500 | | 1 |
| 875 | PARKER | PRESIDENT | | 09/07/2002 | 60000 | | 1 |
| 970 | BLACK | ADMIN | 818 | 21/11/1997 | 23000 | | 1 |

No explicit SQL implementation

Anti join

Left for you to find out

Hint: Sometimes described as the opposite of a semi join