# Relations and Relational Algebra

CS1F IM Lecture 8 Craig Macdonald

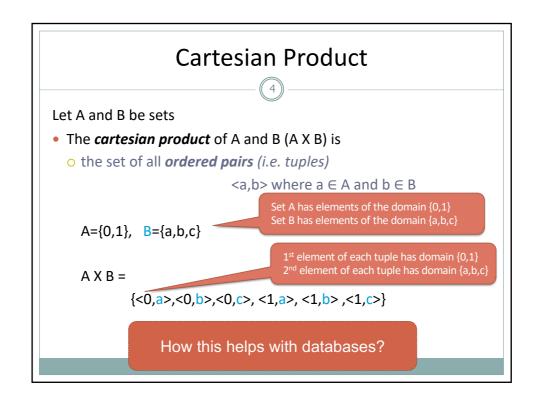
## Where are we now......

- Set Theory (Tuesday)
- Relations and Relational Algebra (today)
- SQL and querying a database (lecture 9)
- Populate and Query your database (Next weeks Lab (week 5))
- Database Assessment (week 6)
  - Verbally in labs with queries
  - O Report (intro, description, rationale, conclusion)
  - o TOTAL: 24 marks

## Recap on Sets



- What are sets? {1,3,5}
- Set builder notation for making sets; comparing sets
- Operators: making new sets from other sets
  - ∪ union
  - ∩ intersection
  - – difference
  - A complement
  - ⊕ symmetric difference
- And... the Cartesian Product...



## Cartesian Products & Relations



- Consider a relation, where the two attributes have the same domain restrictions: R(A,B)
  - R({0,1}, {a,b,c})
- All possible rows in R:
  - Same as {0,1} x {a,b,c}
- If we had less rows in our relation R, it would be a subset of {0,1} x {a,b,c}

A ={0,1}	B ={a,b,c}
0	а
0	b
0	С
1	а
1	b
1	С

## Relations



- Relationships between elements of sets are represented using a structure called a *relation*
- "A relation R is a subset of the Cartesian product of the domains that define R"
  - i.e. of the domains of its attributes
- Relations are the fundamental data structure used to store information in databases

# Relations and their properties



A and B are sets

#### a binary relation R from A to B

is a subset of the Cartesian product A X B

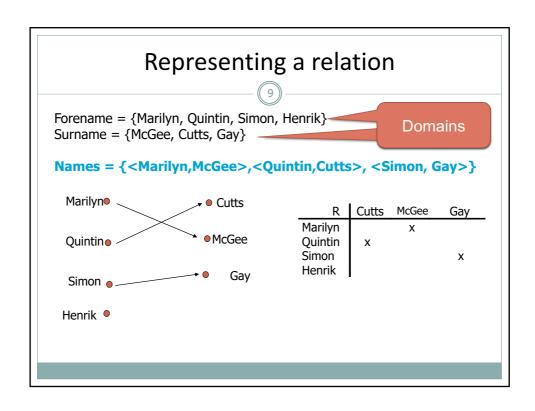
- That is a set R of ordered pairs where the first element of each ordered pair comes from A and the second element comes from B
- Notation
  - o a R b denotes <a,b> ∈ R
  - o a is related to b by R

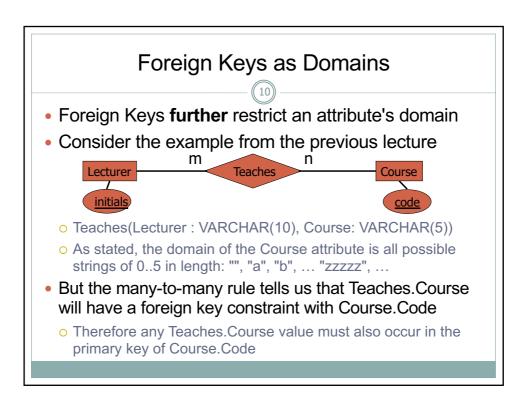
# An example relation

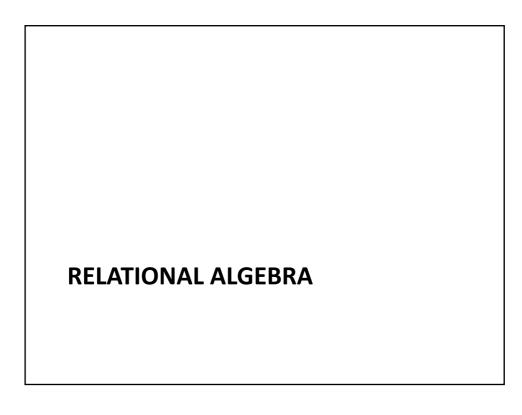


- Let A =  $\{3,4,5\}$  and B =  $\{x,y\}$
- Cartesian Product A x B =

- A Relation R between A and B is a **<u>subset</u>** of this Cartesian Product
- E.g. R= {<3,x>, <4,x>, <5,y>} is a relation from A to B
  - o 3 R x True
  - o 3 R y False







# Relational Algebra



- Sets and relations help us understand how the underlying data elements are structured in a database
- <u>Relational Algebra</u> helps us understand how the DBMS extracts the required data from the DB

# Querying a Database



#### A query...

...is performed when we wish to extract from the database a *subset* of the information that answers some question:

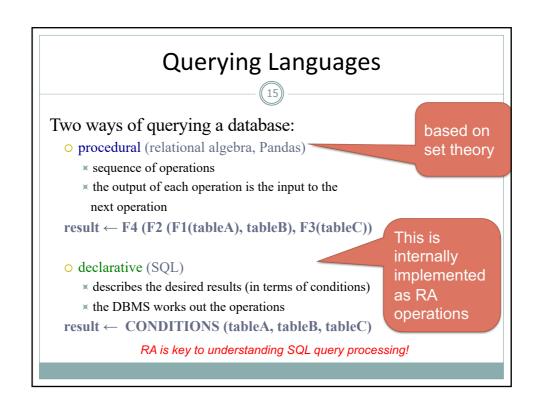
- o "What are the department names?"
- o "Tell me all the data held about male employees."
- "What are the names of the employees in the R&D department?"

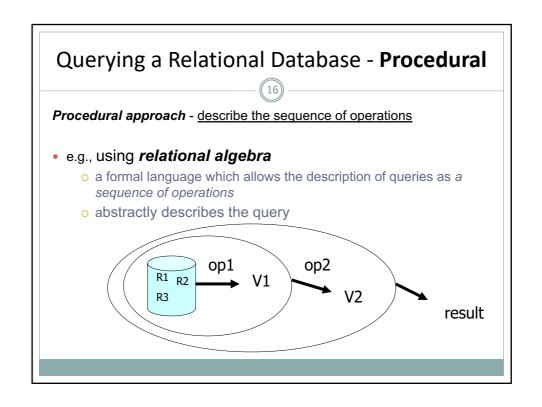
13

# Querying a Database



- Extraction of results consists of 'programs' built out of
  - o retrieving data as a subset of some relation
  - o combining two relations together in a meaningful way





## **Preliminaries**



- A query is a sequence of <u>operations</u> applied to relation instances, and the result of a query is also a relation instance
  - Schema of relations are fixed (c.f. previous lectures)
  - The query will then execute over any valid relation instance...
  - o ... and return a relation
  - The schema of the result can also be determined

## Relational Algebra



- Relational algebra is a set of <u>operations</u> which can be combined to <u>provide the result of a query in the form of a relation</u>
- The algebra consists of:
  - o a collection of operations which fall into two categories:
    - x special relational operations
    - × traditional set operations
  - o a form of "relational assignment" statement so that partial results can be assigned a name and passed to subsequent operations

# Relational Assignment



- A query is made up of a sequence of operations of the form:
- Format:
  - o newRelation := UnaryOperation parameter (inputRelation)
    - × Parameters can include column names or conditions

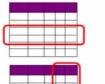
Or (for a binary operator)

newRelation := inputRelation<sub>1</sub> Operator parameter inputRelation<sub>2</sub>

# **Relation Operations**



- Principal relation operations are:
  - Select pick rows from a relation by some condition
  - o Project pick columns by name



select

project

o Join - connect two relations (usually by a Foreign Key)

## **Set Operations**



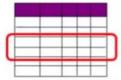
- Set operations include:
  - union make a relation containing all the rows of two relations;
  - intersection pick the tuples which are common to two relations;
  - difference pick the tuples which are in one relation but not another;
  - Cartesian Product pair off each of the tuples in one relation with those in another - creating a double sized row for each pair.

21

# Selection (σ)



Extract the tuples (rows) of a relation (table)
which satisfy some condition on the values of their rows
and return these as a relation (table view)



select

# Selection ( $\sigma$ )



## **Syntax**

 $\sigma_{\text{ Condition}}$  ( RelationName )

Locals :=  $\sigma_{city = "Glasgow"}$  (Employee)

- where the condition can contain:
  - o Literals, e.g. "Glasgow"
  - o column names, e.g. city
  - o comparison operators ( =, >, etc. )
  - o boolean operators ( and, not, or )

23

## Selection

## **Employee**

<u>name</u>	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	17	75	My St	Stirling
Black	541	5	F	44	28	77	Ur St	Edinburgh

"Who are the employees that live in Glasgow?"

Locals :=  $\sigma_{\text{city} = "Glasgow"}$  (Employee)

#### Locals

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow

# Selection

## **Employee**

name	NT	Dept	sex	age	salarv	house	street	citv
Harrie			30%	uge	Jaiai	110000		Cicy
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	1/	75	My St	Stirling
Black	541	5	F	44	28	77	Ur St	Edinburgh

"Who are the employees that live in Glasgow?"

Locals :=  $\sigma_{\text{city} = "Glasgow"}$  (Employee)

#### Locals

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow

25

# Projection ( $\Pi$ )



- Extracts some of the columns from a relation, given the names
  - O E.g.

GenderSalary :=  $\Pi$  (gender, salary) (Employee)

- o no attribute may occur more than once
- o duplicates will be removed



project

# Projection

## **Employee**

name	NI	Dept	aena	lor	age	salar	,	house	street	citv
<u>Harric</u>	IAT	Dept	gene	CI	age	Salai	<u>Y</u> _	House	Succi	City
Smith	123	7	F		52	34		5	High St	Glasgow
Jones	456	5	М		35	23		21	Lo St	Glasgow
McSomething	789	6	М		24	17		75	My St	Stirling
Black	541	5	F		44	28		77	Ur St	Edinburgh
*			\			\ /				-

GenderSalary :=  $\Pi$  (Gender, salary) (Employee)

## GenderSalary

gender	salary
F	34
М	23
М	17
F	28

27

# Projection ( $\Pi$ )



• Projection and selection can be combined

 $\Pi$  (house, street) (  $\sigma$  city = "Glasgow" (Employee))

# Projection ( $\Pi$ )



· Projection and selection can be combined

```
\Pi (house, street, city) (\sigma city = "Glasgow" (Employee))
```

• does a selection

29

# Projection ( $\Pi$ )



• Projection and selection can be combined

```
\Pi (house, street, city) (\sigma city = "Glasgow" (Employee))
```

- does a selection followed by a projection
- For this query, the following RA is equivalent

 $\sigma_{\text{city} = \text{"Glasgow"}}$  (  $\Pi_{\text{(house, street, city)}}$  (Employee))

• A DBMS can re-organise RA expressions for faster retrieval

# Union ( $\cup$ )



- Produces a relation which combines two relations into a new relation containing **all** of the tuples from each (removing duplicates)
- The two relations must be "union compatible" i.e. have the same number of attributes drawn from the same domain
- E.g.

GlasgowOrStirling :=

 $\sigma_{\text{city = "Glasgow"}}$  (Employee)  $\cup \sigma_{\text{city = "Stirling"}}$  (Employee)

31

## Union

## **Employee**

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	17	75	My St	Stirling
Black	541	5	F	44	28	77	Ur St	Edinburgh

GlasgowOrStirling : =  $\sigma_{\text{city = "Glasgow"}}$  (Employee)  $\sigma_{\text{city = "Stirling"}}$  (Employee)

## GlasgowOrStirling

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	17	75	My St	Stirling

## Union

## Employee

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	9	ĪΫĪ	24	17	<b>7</b> 5	My St	Stirling
Black	541	5	F	44	28	77	Ur St	Edinburgh

GlasgowOrStirling : =  $\sigma$  <sub>city = "Glasgow"</sub> (Employee)  $\cup$   $\sigma$  <sub>city = "Stirling"</sub> (Employee)

## GlasgowOrStirling

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	17	75	Mv St	Stirling

33

# Union

## **Employee**

n	ame_	NI	Dept	sex	age	salary	house	street	city
S	mith	123	7	F	52	34	5	High St	Glasgow
J	ones	456	7	М	35	23	21	Lo St	Glasgow
\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\	1cSomething	789	6	М	24	17	75	My St	Stirlina
В	lack	541	5	F	44	28	77	Ur St	Edinburgh

GlasgowOrStirling : =  $\sigma$  <sub>city = "Glasgow"</sub> (Employee)  $\cup$   $\sigma$  <sub>city = "Stirling"</sub> (Employee)

## GlasgowOrStirling

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	17	75	My St	Stirling

# **Union Compatibility**



# • E.g. S1∪S2 is valid, but S1∪R1 is not

**R**1

sid	bid	day
22	101	101001
99	103	111201

**S**1

I				
	sid	sname	rating	age
	11	Sue	7	26
	22	Tim	8	26
	33	Bob	9	28
	55	Kim	10	28

S2

sid	sname	rating	age
10	Myleene	6	23
22	Tim	8	26
99	Julia	100	20
88	Gavin	100	21

# Intersection ( $\cap$ )



Similar to union but returns tuples that are in both relations

E.g.

FemalesInGlasgow :=

 $\sigma_{\text{city = "Glasgow"}}$  (Employee)  $\cap$   $\sigma_{\text{gender = "F"}}$  (Employee)

## Intersection

## Employee

<u>name</u>	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	M	24	17	75	My St	Stirling
Black	541	5	F	44	28	77	Ur St	Edinburgh

FemalesInGlasgow :=

 $\sigma$  city = "Glasgow" (Employee)  $\cap$ 

σ gender = "F" (Employee)

#### FemalesInGlasgow

<u>name</u>	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow

37

## Intersection

## Employee

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Clasgow
McSomething	789	6	М	24	17	75	My St	Stirling
Black	541	5	F	44	28	77	Ur St	Edinburah

FemalesInGlasgow :=

 $\sigma$  city = "Glasgow" (Employee)  $\cap$ 

 $\sigma$  gender = "F" (Employee)

## FemalesInGlasgow

<u>name</u>	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow

## Intersection

## **Employee**

name	NT	Dept	SOX	age	salary	house	street	city
Smith	123	7	F	52	34	5	Hiah St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	17	75	My St	Stirling
Plack	541	5	F	44	28	77	Ur St	Edinburdi

FemalesInGlasgow :=

 $\sigma$  city = "Glasgow" (Employee)  $\cap$ 

σ gender = "F" (Employee)

#### FemalesInGlasgow

<u>name</u>	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow

39

# Difference ( - )



 Similar to union but returns tuples that are in the first relation but not the second

E.g.

NonLocals := Employee – Locals

- Intersection and difference both require union compatibility
- Both use column names from the first relation

# More Examples of Union compatible operations

Customers

FN	LN
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert

Em	pΙ	O	/e	es

,0,0						
FN	LN					
John	Smith					
Ricardo	Browne					
Susan	Yao					
Francis	Johns					
Ramesh	Shah					

**Customers** U Employees

FN	LN
Susan	Yao
Ramesh	Shah
Johnny	Kohler
Barbara	Jones
Amy	Ford
Jimmy	Wang
Ernest	Gilbert
John	Smith
Ricardo	Browne
Francis	Johns

# Customers∩ Employees FN

FN	LN
Susan	Yao
Ramesh	Shah

FN LN
John Smith
Ricardo Browne
Francis Johns

4

# Operations on any two relations

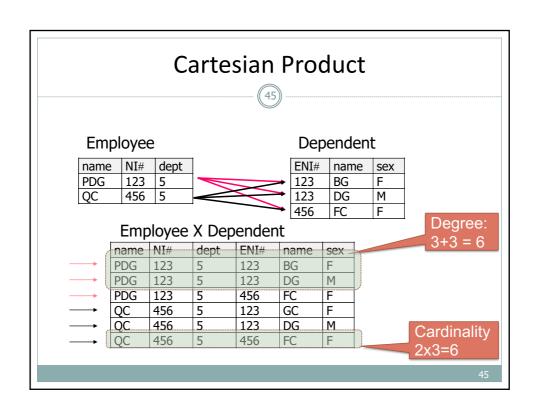


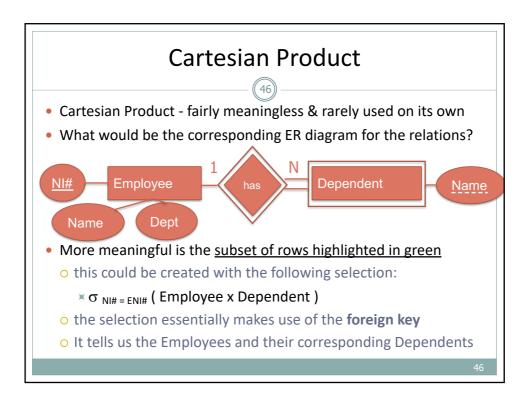
- Relations may have different schema
- Conditions indicate how to join the relations around common attributes

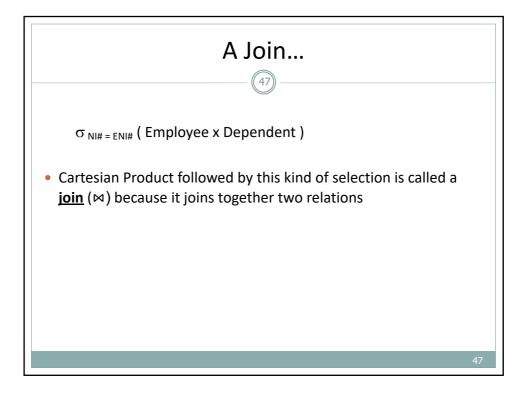
# Cartesian Product of Relations (X)



- Cartesian Product A X B of two relations A and B, which have attributes A1 ... Am and B1 ... Bn.....
  - o ....is the relation with m + n attributes containing a row for every pair of rows, one from A and one from B
- Thus if A has a tuples and B has b tuples then the result has a x b tuples







## **Join** = Cartesian Product followed by Selection



## **Employee**

name	NI#	dept
PDG	123	5
QC	456	5

## Dependent

ENI#	name	sex
123	BG	F
123	DG	М
456	FC	F

Deps :=  $\sigma_{\text{NI\#} = \text{ENI\#}}$  ( Employee x Dependent )

#### Deps \_\_\_

name	NI#	dept	ENI#	name	sex
PDG	123 🝧	5	123	BG	F
PDG	123 🥌	5	123	DG	М
QC	456	5	456	FC	F
•					

48

# Join operators



- Join operators takes two relations and makes a new one, based on the pairing of rows between those two relations
- There are a wide variety of join operators
- This particular form is called an equi-join:

Relation<sub>1</sub>  $\bowtie$  A=B Relation<sub>2</sub>

 $\Leftrightarrow$ 

 $\sigma_{A=B}$  (Relation<sub>1</sub> x Relation<sub>2</sub>)

## **Join** = Cartesian Product followed by Selection



Deps :=  $\sigma_{\text{NI\#} = \text{ENI\#}}$  ( Employee x Dependent )

#### Deps

name	NI#	dept	ENI#	name	sex
PDG	123	5	123	BG	F
PDG	123	5	123	DG	М
QC	456	5	456	FC	F

 $\Leftrightarrow$ 

 $\textbf{Deps} \coloneqq \texttt{Employee} \ \bowtie \ \texttt{NI\#} = \texttt{ENI\#} \ \texttt{Dependent}$ 

50

# Natural Join (⋈)



- In its simplest form, the join of relations A and B pairs off the tuples of A and B so that **identically named** attributes from the relations have the same value
- We now have two columns holding the same value, so we eliminate the duplicated common attributes to form the natural join or inner join
- Natural Join is simply written as

Relation1 ⋈ Relation2

# Example



B1

R1

sid	bid	day
22	101	101001
99	103	111201

bid	colour
101	red
102	blue
103	green

R1 ⋈ B1

sid	R1.bid	day	colour
22	101	101001	red
99	103	111201	green

# Join Summary



- Cartesian product A X B : every row in A with every row in B
- Join ⋈: Cartesian product followed by a selection
  - o Equi-join ⋈<sub>X=Y</sub>: Selection is on some attributes being equal
  - Natural join ⋈: Selection is on attributes with the same name being equal, and duplicate attributes in the resulting schema removed

## Summary of Relational Algebra Operators



Three basic types of Relational Algebra operators:

- 1. Applying to one relation
  - projection ☐ (attributes), selection σ (conditions)
- 2. Applying to two relations of identical structure
- union U, intersection ∩, difference (no conditions)

  3. Applying to two relations of different structure

(Cartesian) product **×** (no conditions) joins **⋈** (conditions)

## Example



Employee and Department are two tables in a relational database.

- Employee (Name, NI-Number, Email, Phone-No, Works In)
- Department (DeptName, <u>Code</u>, BuildingName)

The attribute **Works\_In** is a foreign key in Employee relating to the primary key **Code** in Department.

Express in relational algebra:

(a) the names of all departments based in the Main building

DeptsInMain :=  $\pi_{\text{DeptName}}$  ( $\sigma_{\text{BuildingName}} = \text{"Main"}$  (Departments))

## Example



Employee and Department are two tables in a relational database.

- Employee (Name, NI-Number, Email, Phone-No, Works\_In)
- Department (DeptName, Code, BuildingName)

#### Schema:

yee relating to the

#### Schema

AccountsEmployeesNames (Name)

(b) the nanother the employees of the "Accounts" department. AllEm, eDepts := Employee  $\bowtie$  Works\_In = Code Department Account Employees :=  $\sigma$  DeptName = "Accounts" (AllEmployeeDepts ) AccountsEmployeesNames :=  $\pi$  Name (AccountsEmployees)

- 1. How much of this RA material (as a percentage) do you understand?
- 2. (If the answer to Q1 is less than 100%), what are you going to do to raise this percentage?
- 3. When (exactly) are you going to do it?

# Where to go for more info...

- Rosen, Discrete Mathematics and Its Applications
  - o sets sections 1.4 & 1.5
  - o relations sections 6.1 & 6.2
  - http://www.mhhe.com/math/advmath/rosen/
- OR
- Garcia Molina, Chapter 5, Section 1 (pgs 199-205)
- OR
- https://www.tutorialspoint.com/dbms/relational\_algebra.htm
- OB
- more worked examples in the handouts at end of these slides...!
- OR
- Relational Algebra quiz on Moodle

## On Moodle...



IM Labs and Tutorials

Relational Algebra Quiz

The solutions will be placed online once enough students have viewed the quiz!

Some helpful external resources for Information Management

what is a database?

What is a relational database?

what is an ER diagram?

drawing ER diagrams

tutorial on relational algebra

## Relational Algebra Queries – Worked Examples



#### **Employee**

. ,								
name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	17	75	My St	Stirling
Black	541	5	F	44	28	77	Ur St	Edinburah

#### ResDept

Dnum	Dname	manager
5	R&D	456

61

# Relational Algebra Queries: Example 1



- Give the name and addresses of all employees who work for the R&D department
- ResDept :=  $\sigma_{\text{Dname} = \text{"R&D"}}$  (Department)
  - o this should return just one tuple
- ResDeptEmps := ResDept M Dnumber = Dept# Employee
  - o this picks out the employees in that department.
- Result :=  $\Pi$  (Name, Address) ResDeptEmps
- Notes:
  - o the order could be changed e.g. to make it faster.
  - o the sub-steps could be merged into one:
  - $\Pi$  (Name, Address) ( ( $\sigma$  Dname = "R&D" (Department)  $\bowtie$  Dnumber = Dept# Employee
  - o but this is not so readable

# Example 1: Breakdown (1/3)



Give name and addresses of employees who work for the R&D dept

## Department

Dnum	Dname	manager
5	R&D	456
6	Production	111
7	Admin	345

ResDept :=  $\sigma$  Dname = "R&D" (Department)

## ResDept

Dnum	Dname	manager		
5	R&D	456		

63

# Example 1: Breakdown (2/3)



## **Employee**

name	NI	Dept	sex	age	salary	house	street	city
Smith	123	7	F	52	34	5	High St	Glasgow
Jones	456	5	М	35	23	21	Lo St	Glasgow
McSomething	789	6	М	24	17	75	My St	Stirling
Black	541	5	F	44	28	77	Ur St	Edinburgh

#### ResDept

Dnum	Dname	manager
5	R&D	456

## ResDeptEmps

Dnum	Dname	Manager	name	NI	sex	age	salary	house	street	city
5	R&D	456	Jones	456	М	35	23	21	Lo St	Glasgow
5	R&D	456	Black	541	F	44	28	77	Ur St	Edinburah

# Example 1: Breakdown (3/3)



## ResDeptEmps

Dnum	Dname	Manager	name	NI	Dept	sex	age	salary	house	street	city
5	R&D	456	Jones	456	5	М	35	23	21	Lo St	Glasgow
5	R&D	456	Black	541	5	F	44	28	77	Ur St	Edinburgh

 $\Pi$  (Name, House, Street, City) ResDeptEmps

#### Result

1	name	house	street	city
	Jones	21	Lo St	Glasgow
	Black	77	Ur St	Edinburah

65

# Example 2



- List the name, controlling department name, the department manager's name, address and date of birth for every project in Stafford
- StaffordProjs :=  $\sigma_{\text{PLocation}} = "Stafford"$  (Project)
  - o it is common to start by restricting to an area of interest
- StaffordProjDepts:= StaffordProjs ⋈ Dnum=Department Department
  - o this brings together the department information
- StaffordProjDeptMngrs:= StaffordProjDepts  $\bowtie$  MgrNI#=NI# Employee
  - this brings in the manager information
- Result :=  $\Pi_{\text{(Pname, Dname, Name, Address, DateOfBirth)}}$  StaffordProjDeptMngrs