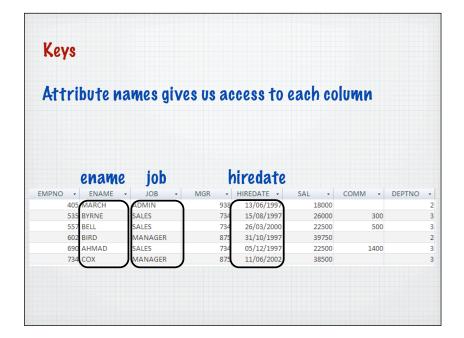# Databases - 2

More on the Relational Model and Selection

Keys

Codd requires us to have a distinct way of accessing every value in the relation

Having unique attribute names is a start..

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

Keys

Attribute names gives us access to each column

ename   job   hiredate

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

Keys

But how can we identify individual tuples?

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	
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

Keys

A **Superkey** is an attribute or a set of attributes that uniquely identifies a **tuple within a relation**

When we use a **key** with a single value, we should only get back one **tuple**

Example

empno is a superkey for the EMP relation

Every time we use **empno**, we only retrieve one **tuple**

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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

Example

Who is empno 405?

Who is empno 602?

Who is empno 912?

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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

Success!

We can now access every cell value by using the combination of the key and the attribute name

---

Example

What is empno 912s job?

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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

Example

What is 875s salary?

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		09/07/2001	60000		3
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

Keys

Keys can be superkeys, candidate keys or primary keys

superkeys	A key is an attribute or a set of attributes that uniquely identifies a tuple within a relation
candidate keys	A minimal superkey
primary keys	The best candidate key

Keys

A Superkey is an attribute or a set of attributes that uniquely identifies a tuple within a relation

A relation may have many superkeys..

We know that empno is a superkey

Who is empno 912?

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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

What other superkeys are there?

What about ename?

Who is ename MARCH?

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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

What about ename?

ename works (at the moment) but watch out - if another person joins with the same name it breaks

What about combinations of attributes?

A Superkey is an attribute or a set of attributes that uniquely identifies a tuple within a relation

Other superkeys could be ..

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

Note: There is no mathematical formula to calculate the number of superkeys - it all depends on the context

---

Keys

A Candidate key is a superkey which is minimal

If you can remove an attribute from a superkey and its still a superkey, its NOT a candidate key

Consider the SK - remove job ..

empno, job

its still a superkey so not a candidate key

Checking our list of superkeys ..

                        sk   ck

empno

ename

empno, ename

empno, job

ename, job

empno, ename, job

Keys

A **Primary key** is a Candidate key which is the best one

Depends on the context

Checking our list of candidate keys..

	sk	ck	pk
empno	✔	✔	
ename	✔	✔	
empno, ename	✔	✘	
empno, job	✔	✘	
ename, job	✔	✘	
empno, ename, job	✔	✘	

Foreign Keys

A **Foreign key** is an attribute or set of attributes within one relation that **matches** a candidate key of some (possibly same) relation

Best shown by example

EMP relation

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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

DEPT relation

Foreign key relationship

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

Relational Integrity

Other rules specified by Codd

Entity Integrity: A Primary key must never be null

Why not?

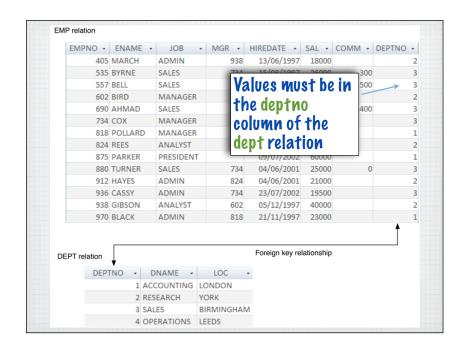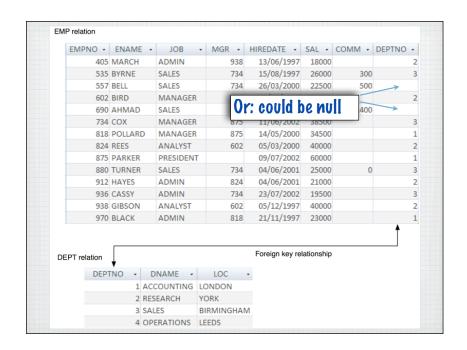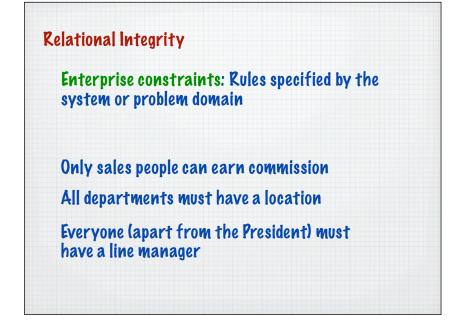| EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
405	MARCH	ADMIN	938	13/06/1997	18000		2
535	BYRNE	SALES	734	15/08/1997	26000	300	3
	BELL	SALES	734	26/03/2000	22500	500	3
602	BIRD	MANAGER	875	31/10/1997	39750		2
690	AHMAD	SALES	734	05/12/1997	22500	1400	3
734	COX	MANAGER	875	11/06/2002	38500		3
	POLLARD	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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

Relational Integrity

Referential Integrity: A Foreign key must either match a candidate key value in its home relation or be null

EMP relation

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				500	3
602	BIRD	MANAGER					2
690	AHMAD	SALES				400	3
734	COX	MANAGER					3
818	POLLARD	MANAGER					1
824	REES	ANALYST					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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

Values must be in the deptno column of the dept relation

Foreign key relationship

DEPT relation

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

EMP relation

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	
602	BIRD	MANAGER					2
690	AHMAD	SALES				400	
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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

Or: could be null

Foreign key relationship

DEPT relation

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

Relational Integrity

Enterprise constraints: Rules specified by the system or problem domain

Only sales people can earn commission

All departments must have a location

Everyone (apart from the President) must have a line manager

Relational Algebra operations

Selection	σ	Choose particular rows
Projection	π	
Cartesian Product	×	
Union	∪	
Set Difference	-	
Join	⋈	
Intersection	∩	
Division	÷	

SELECTION

$\sigma_{predicate}(R)$

Selection operation works on a single relation R and defines a relation that contains only those tuples (rows) of R that satisfy the specified condition (predicate)

SELECTION RA example

List all staff with a salary greater than £25,000

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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

SELECTION RA example

List all staff with a salary greater than £25,000

$$\sigma_{\text{predicate}} (\text{emp})$$

The condition or predicate goes in here

Named relation goes here in brackets

Predicates

Expressions that evaluate to **true** or **false** once all the names have been replaced with a value

expression	explanation
job = "manager"	is job equal to "manager"
sal > 30000	is salary greater than 30000
job="admin" and sal>25000	is job equal to "admin" and salary > 25000

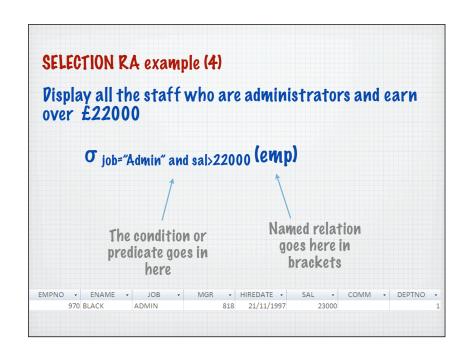## Predicates

hints	example
literal strings must be wrapped in " "	"Manager"
Use <, >, =, <=, >= and <>	sal>=25000 comm<=300 deptno<>3
Build more complex expressions with **and, or, not**	sal<25000 and deptno=5

SELECTION RA example

List all staff with a salary greater than £25,000

$$\sigma_{\text{sal}>25000}(emp)$$

The condition or predicate goes in here

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
535	BYRNE	SALES	734	15/08/1997	26000	300	3
602	BIRD	MANAGER	875	31/10/1997	39750		2
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
938	GIBSON	ANALYST	602	05/12/1997	40000		2

SELECTION RA example (2)

Show all the staff who are managers

$$\sigma_{\text{job}="manager"}(emp)$$

The condition or predicate goes in here

Named relation goes here in brackets

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
602	BIRD	MANAGER	875	31/10/1997	39750		2
734	COX	MANAGER	875	11/06/2002	38500		3
818	POLLARD	MANAGER	875	14/05/2000	34500		1

SELECTION RA example (3)

Show all the staff who are managed by employee no 734

$$\sigma_{\text{mgr}=734}(emp)$$

The condition or predicate goes in here

Named relation goes here in brackets

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
535	BYRNE	SALES	734	15/08/1997	26000	300	3
557	BELL	SALES	734	26/03/2000	22500	500	3
690	AHMAD	SALES	734	05/12/1997	22500	1400	3
880	TURNER	SALES	734	04/06/2001	25000	0	3
936	CASSY	ADMIN	734	23/07/2002	19500		3

SELECTION RA example (4)

Display all the staff who are administrators and earn over £22000

$$\sigma_{\text{job="Admin" and sal>22000}} (emp)$$

The condition or predicate goes in here

Named relation goes here in brackets

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
970	BLACK	ADMIN	818	21/11/1997	23000		1

So: how do we do SELECTION in SQL?

SQL always looks like this:

select * or expression
from relations
[where expression]

the 'where' part is optional

SELECTION SQL example

List all staff with a salary greater than £25,000

the particular columns we require go here

select *
from emp
where sal>25000

table name goes here

the predicate goes here

EMPNO	ENAME	JOB	MGR	HIREDATE	SAL	COMM	DEPTNO
535	BYRNE	SALES	734	15/08/1997	26000	300	3
602	BIRD	MANAGER	875	31/10/1997	39750		2
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
938	GIBSON	ANALYST	602	05/12/1997	40000		2

SELECTION SQL - other examples

Show all the staff who are managers

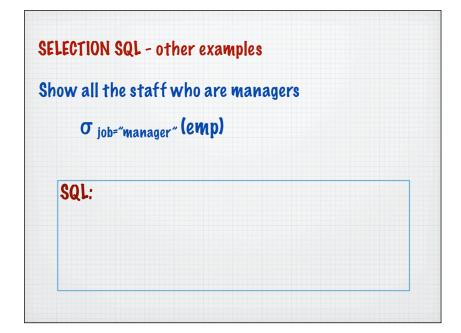$$\sigma_{job="manager"}(emp)$$

SQL:

SELECTION SQL - other examples

Show all the staff who are managed by employee no 734

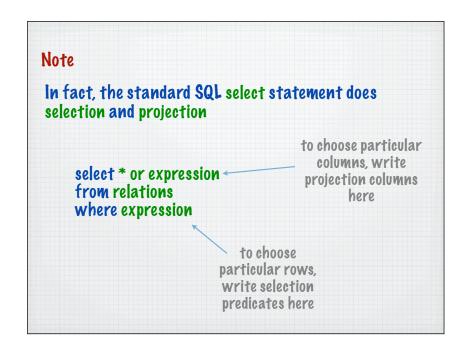$$\sigma_{mgr=734}(emp)$$

SQL:

SELECTION SQL - other examples

Display all the staff who are administrators and earn over £22000

$$\sigma_{job="Admin"\ and\ sal>22000}(emp)$$

SQL:

Note

In fact, the standard SQL **select** statement does **selection and projection**

select * or expression
from relations
where expression

to choose particular columns, write projection columns here

to choose particular rows, write selection predicates here

PROJECTION AND SELECTION SQL example

Produce a list of staff who earn over 25000, showing only the Empno, EName and Job

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	05/12/1997	40000		2
970	BLACK	ADMIN	818	21/11/1997	23000		1

PROJECTION AND SELECTION SQL example

Produce a list of staff who earn over 25000, showing only the Empno, EName and Job

select empno, ename, job
from emp
where sal>25000

particular columns go here

table name goes here

conditions go here

empno	ename	job
535	BYRNE	SALES
602	BIRD	MANAGER
734	COX	MANAGER
818	POLLARD	MANAGER
824	REES	ANALYST
875	PARKER	PRESIDENT
938	GIBSON	ANALYST

Note - even though we are using sal in the selection, it's not necessary in the projection