

Databases

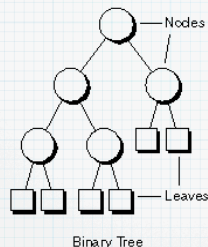
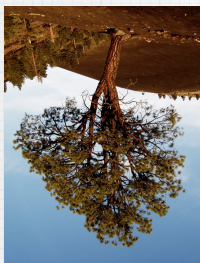
Relational Model, Algebra and operations

How do we **model** and **manipulate** complex data structures inside a computer system?

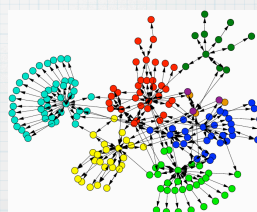
Until 1970 ..

Many different views or ways of doing this

Could use tree structures



Could use network structures



1970 ..

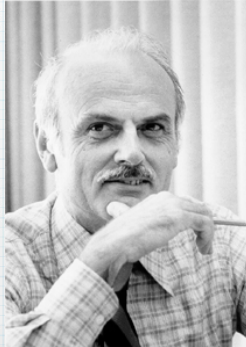
Many different views

Many different implementations

So if you bought some software your data was locked into the product

Until .. EF Codd

A Relational Model of Data for Large Shared Data Banks (1970)



Suggests

Base our data structures on a mathematical structure called a **relation**

Advantages

Everyone agrees on what a **relation** is

We get well defined (agreed) mathematical operations that work on these structures

(looks like set theory - Union, Intersection etc)

Definitions - Relation

The term **relation** is used here in its accepted mathematical sense. Given sets S_1, S_1, \dots, S_n (not necessarily distinct), R is a relation on these n sets if it is a set of n -tuples each of which has its first element from S_1 , its second element from S_1 , and so on.

From A Relational Model of Data for Large Shared Data Banks (1970)

Definitions - Relation

- A relation is a table with columns and rows

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

An abstraction

- Not really stored like this on the hard disk

Definitions - Attribute and Domain

An **attribute** is a named column in the relation

A **domain** is a set of allowable values for one or more attributes

Each attribute has a name

Each attribute has a value that comes from the allowable domain

hiredate must always be date values

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

Examples of attribute domains

Attribute	Domain name	Meaning	Domain defn.
Emp No	EMPLOYEE NO	The set of all possible employee numbers	Number, 3 digits
Ename	EMPLOYEE NAME	The set of all possible employee names	Character, size 15
Job	JOB TITLE	The set of all possible jobs	Character, size 15
Mgr	MANAGER NO	The set of all possible manager numbers	Number, 3 digits
Hiredate	HIRE DATES	The set of all possible hire dates	Date
Sal	SALARY	The set of all possible salary values	Number, 5 digits
Comm	COMMISSION	The set of all possible commission values	Number, 4 digits
Deptno	DEPTNO	The set of all possible department numbers	Number, 1 digit

Certain rules must be followed

Every relation (table) must have a distinct name

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

Certain rules must be followed

Every relation (table) must have a distinct name

dept table

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

grade table

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

Certain rules must be followed

Relation **attribute** names must be distinct

.. but names may be duplicated in other relations

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

Which is how we model **relationships**

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

Certain rules must be followed

Values are **atomic**

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

In the day to day manipulation of the structure(s),
we wouldn't normally split a value into smaller parts

Definitions - Tuple, cardinality and degree

Tuple is a row of a relation

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

Definitions - Tuple, cardinality and degree

Cardinality is the number of tuples in a relation

1	405	MARCH	ADMIN	938	13/06/1997	18000		2
2	535	BYRNE	SALES	734	15/08/1997	26000	300	3
3	557	BELL	SALES	734	26/03/2000	22500	500	3
4	602	BIRD	MANAGER	875	31/10/1997	39750		2
5	690	AHMAD	SALES	734	05/12/1997	22500	1400	3
6	734	COX	MANAGER	875	11/06/2002	38500		3
7	818	POLLARD	MANAGER	875	14/05/2000	34500		1
8	824	REES	ANALYST	602	05/03/2000	40000		2
9	875	PARKER	PRESIDENT		09/07/2002	60000		1
10	880	TURNER	SALES	734	04/06/2001	25000	0	3
11	912	HAYES	ADMIN	824	04/06/2001	21000		2
12	936	CASSY	ADMIN	734	23/07/2002	19500		3
13	938	GIBSON	ANALYST	602	05/12/1997	40000		2
14	970	BLACK	ADMIN	818	21/11/1997	23000		1

So the employee relation is cardinality 14

Definitions - Tuple, cardinality and degree

Degree is a number of attributes in a relation

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

1 2 3 4 5 6 7 8

So the employee relation is degree 8

Properties of Relations

As the relational model is based on set theory certain set theory properties apply

Ordering doesn't apply

{ Peas , Red Bull , Socks , Newspaper }

{ Red Bull , Peas , Socks , Newspaper }

{ Red Bull , Peas , Newspaper , Socks }

{ Newspaper , Socks , Red Bull , Peas }

Are all equivalent

Properties of Relations

hence .. attribute ordering makes no difference

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

Properties of Relations

also .. tuple ordering makes no difference

405	MARCH	ADMIN	938	13/06/1997	18000		2
535	BYRNE	SALES		15/08/1997	26000	300	3
557	BELL	SALES	734	26/03/2000	22500	500	3
602	BIRD	MANAGER	818	31/10/1997	39750		2
690	AHMAD	SALES	734	05/12/1997	22500	1400	3
734	COX	MANAGER	818	11/06/2002	38500		3
818	POLLARD	MANAGER	818	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

Properties of Relations

The concept of duplicates doesn't exist

{ Peas , Red Bull , Socks , Newspaper }
 { Red Bull , Peas, Socks, Peas , Newspaper }
 { Red Bull , Newspaper, Newspaper, Peas, Newspaper, Socks }
 { Newspaper, Socks, Red Bull, Peas, Newspaper,
 Socks, Socks, Red Bull, Peas, Newspaper, Socks, Red
 Bull, Peas, Newspaper, }

Are all equivalent

Properties of Relations

.. every tuple is distinct

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

We may need to introduce an attribute to make this true

Alternative terminology

Formal name	Some textbooks	Tradebooks
Relation	Table	File
Tuple	Row	Record
Attribute	Column	Field

Relational Algebra operations

Relations (tables) are manipulated by relational algebra expressions

These operations always produce more relations (tables) so that expressions can be nested

Can be thought of carving a table into rows, columns or merging tables together

Relational Algebra operations

There are 5 basic operations in relational algebra

Selection	σ	small sigma
Projection	π	small pi
Cartesian Product	\times	
Union	\cup	
Set Difference	$-$	

Relational Algebra operations

3 additional operations can be made up from the previous 5

Join	\bowtie
Intersection	\cap
Division	\div

Relational Algebra operations

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

Choose particular columns

PROJECTION

$$\pi_{\text{col 1, col 2, } \dots, \text{col n}}(R)$$

Projection operation works on a single relation R and defines a relation that contains a vertical subset of R , extracting the values of specified attributes and eliminating duplicates.

PROJECTION example

Produce a list of salaries for all staff, showing only the Empno, EName, Job, and 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		09/07/2002	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 example

Produce a list of salaries for all staff, showing only the Empno, EName, Job, and Salary

$$\pi_{\text{col 1, col 2, } \dots, \text{col n}}(R)$$

The attributes we require go here, in subscript separated by commas

Named relation goes here in brackets

PROJECTION example

Produce a list of salaries for all staff, showing only the Empno, EName, Job, and Salary

Π empno, ename, job, sal (emp)

The attributes we require go here, in subscript separated by commas

Named relation goes here in brackets

PROJECTION example

Produce a list of salaries for all staff, showing only the Empno, EName, Job, and Salary

Π empno, ename, job, sal (emp)

empno	ename	job	sal
405	MARCH	ADMIN	18000
535	BYRNE	SALES	26000
557	BELL	SALES	22500
602	BIRD	MANAGER	39750
690	AHMAD	SALES	22500
734	COX	MANAGER	38500
818	POLLARD	MANAGER	34500
824	REES	ANALYST	40000
875	PARKER	PRESIDENT	60000
880	TURNER	SALES	25000
912	HAYES	ADMIN	21000
936	CASSY	ADMIN	19500
938	GIBSON	ANALYST	40000
970	BLACK	ADMIN	23000

So how are relational operations implemented in a database?

Structured Query Language

SQL

All relational databases implement SQL

Access	mySQL	mSQL
Oracle	SQL server	postgresSQL

SQL

- Is a standard (ANSI and ISO)
- Various versions, typically known by the standardisation year
- SQL-86, SQL-89, SQL-92, SQL3
- Databases typically claim to support a particular version

SQL STATEMENTS

Always look like this:

```
select * or expression  
from relations  
[where expression]
```

PROJECTION SQL example

Produce a list of salaries for all staff, showing only the Empno, EName, Job, and Salary

```
select * or expression  
from relations
```

Note that we use the reserved word "select" here, even though its a projection

PROJECTION SQL example

Produce a list of salaries for all staff, showing only the Empno, EName, Job, and Salary

select * or expression
from relations

Attribute names
go here

table name goes
here

PROJECTION SQL example

Produce a list of salaries for all staff, showing only the Empno, EName, Job, and Salary

select empno, ename, job, sal
from emp

Attribute names
go here

table name goes
here

empno	ename	job	sal
405	MARCH	ADMIN	18000
535	BYRNE	SALES	26000
557	BELL	SALES	22500
602	BIRD	MANAGER	39750
690	AHMAD	SALES	22500
734	COX	MANAGER	38500
818	POLLARD	MANAGER	34500
824	REES	ANALYST	40000
875	PARKER	PRESIDENT	60000
880	TURNER	SALES	25000
912	HAYES	ADMIN	21000
936	CASSY	ADMIN	19500
938	GIBSON	ANALYST	40000
970	BLACK	ADMIN	23000

PROJECTION example (2)

Produce a list of department names and locations

Π dname, loc (dept)

select dname, loc
from dept

dname	loc
ACCOUNTING	LONDON
RESEARCH	YORK
SALES	BIRMINGHAM
OPERATIONS	LEEDS
*	

Records: 14 5 of 5

PROJECTION example (3)

Produce a list of all the columns in the **grade** table

Π **grade, losal, hisal** (**grade**)

OR

grade

select grade, losal, hisal
from grade

OR

select *
from grade

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

PROJECTION example (4)

Produce a list of employee names along with the salaries increased by 3%

Π **ename, sal/100*3+sal** (**emp**)

select ename, sal/100*3+sal
from emp

ename	Expr1001
MARCH	18540
BYRNE	26780
BELL	23175
BIRD	40942.5
AHMAD	23175
COX	39655
POLLARD	35535
REES	41200
PARKER	61800
TURNER	25750
HAYES	21630
CASSY	20085
GIBSON	41200
BLACK	23690

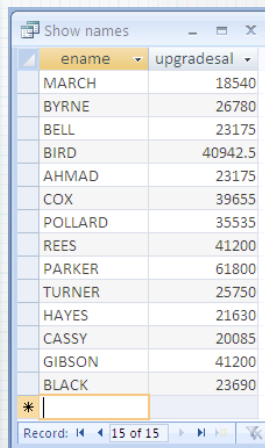
Note: SQL uses attribute names where possible, otherwise randomly generated names are used

select ename, sal/100*3+sal
from emp

ename	Expr1001
MARCH	18540
BYRNE	26780
BELL	23175
BIRD	40942.5
AHMAD	23175
COX	39655
POLLARD	35535
REES	41200
PARKER	61800
TURNER	25750
HAYES	21630
CASSY	20085
GIBSON	41200
BLACK	23690

Note: To force a name, use the **AS** reserved word

select **ename**, $sal/100*3+sal$ as **upgradesal**
from **emp**



ename	upgradesal
MARCH	18540
BYRNE	26780
BELL	23175
BIRD	40942.5
AHMAD	23175
COX	39655
POLLARD	35535
REES	41200
PARKER	61800
TURNER	25750
HAYES	21630
CASSY	20085
GIBSON	41200
BLACK	23690

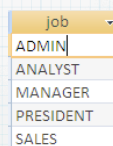
WATCH OUT!

- Database packages implement a **version** of the relational model
- You may get SQL results that are slightly different to the mathematical model

PROJECTION example (5) - RA ...

Produce a list of employee jobs

$\pi_{jobs}(emp)$



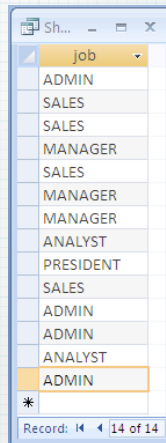
job
ADMIN
ANALYST
MANAGER
PRESIDENT
SALES

PROJECTION example (5) - SQL ...

Produce a list of employee jobs

```
select jobs  
from emp
```

Relational database don't
typically reduce duplicates - to do
this use the **distinct** keyword



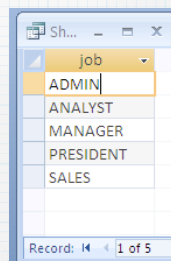
A screenshot of a database window titled 'Sh...' showing a list of employee jobs. The list contains 14 records, with many duplicates. The jobs listed are: ADMIN, SALES, SALES, MANAGER, SALES, MANAGER, MANAGER, ANALYST, PRESIDENT, SALES, ADMIN, ADMIN, ANALYST, and ADMIN. The status bar at the bottom indicates 'Record: 14' and '14 of 14'.

job
ADMIN
SALES
SALES
MANAGER
SALES
MANAGER
MANAGER
ANALYST
PRESIDENT
SALES
ADMIN
ADMIN
ANALYST
ADMIN

PROJECTION example (5) - SQL ...

Produce a list of employee jobs

```
select distinct jobs  
from emp
```



A screenshot of a database window titled 'Sh...' showing a list of distinct employee jobs. The list contains 5 records, with no duplicates. The jobs listed are: ADMIN, ANALYST, MANAGER, PRESIDENT, and SALES. The status bar at the bottom indicates 'Record: 14' and '1 of 5'.

job
ADMIN
ANALYST
MANAGER
PRESIDENT
SALES