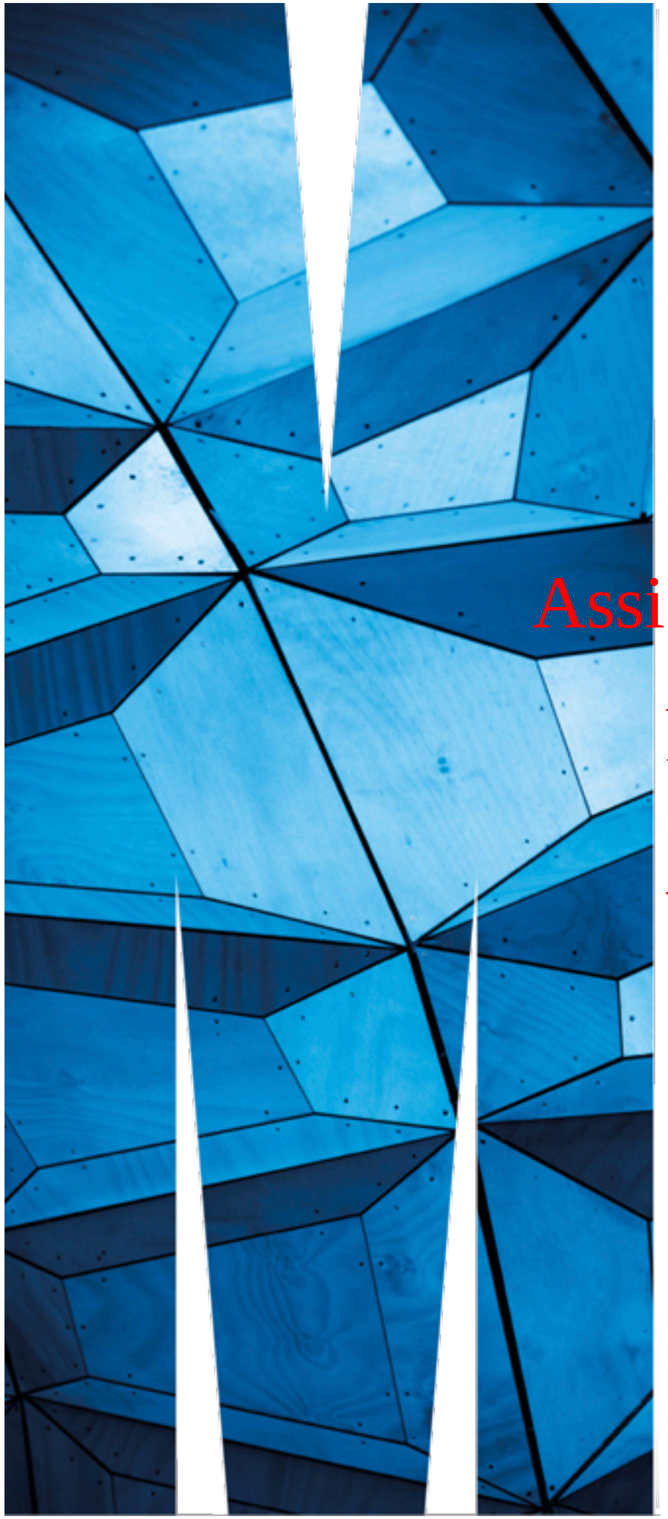


FIT9132 Introduction to Database

The Relational Database Model

Campbell Wilson
Faculty of Information Technology
Monash University

MONASH University

Agenda

- Data Redundancy
  - The motivation behind introducing relational model.

- Relational Model

	PROJECT_CODE	PROJECT_MANAGER	MANAGER_PHONE	MANAGER_ADDRESS	PROJECT_BID_PRICE
▶	21-5Z	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$16,833,460.00
	25-2D	Jane D. Grant	615-898-9909	218 Clark Blvd., Nashville, TN 36362	$12,500,000.00
	25-5A	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	$32,512,420.00
	25-9T	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$21,563,234.00
	27-4Q	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	$10,314,545.00
	29-2D	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$25,559,999.00
	31-7P	William K. Moor	904-445-2719	216 Morton Rd., Stetson, FL 30155	$56,850,000.00

Assume a database contains a single table depicted above. The table is used to record data regarding on-going projects and the project manager. Each row is uniquely identified by **project_code**. A project_code is assigned to a project when the project commences. A new project manager has been hired in an anticipation of a big project commencing in 2 months.

Q1. What would be a problem associated with using the above table if we try to add the details of the new manager to the database?

a. There will not be enough column to enter the data into the database.
b. The insertion of the project manager is not possible as no project code is available.
c. There will not be enough row to enter the data in the database.
d. None of the above

	PROJECT_CODE	PROJECT_MANAGER	MANAGER_PHONE	MANAGER_ADDRESS	PROJECT_BID_PRICE
▶	21-5Z	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$16,833,460.00
	25-2D	Jane D. Grant	615-898-9909	218 Clark Blvd., Nashville, TN 36362	$12,500,000.00
	25-5A	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	$32,512,420.00
	25-9T	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$21,563,234.00
	27-4Q	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	$10,314,545.00
	29-2D	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$25,559,999.00
	31-7P	William K. Moor	904-445-2719	216 Morton Rd., Stetson, FL 30155	$56,850,000.00

Assume Ms. Holly B. Parker needs to change her phone number.

Q2. What would be a potential issue associated with the changing of the phone number?

a. The database will not allow the changing of personal details.
b. The project details of Holy B Parker will be deleted.
c. The chance to make errors during the update increases as multiple rows with the same value need to be changed.
d. None of the above.

	PROJECT_CODE	PROJECT_MANAGER	MANAGER_PHONE	MANAGER_ADDRESS	PROJECT_BID_PRICE
▶	21-5Z	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$16,833,460.00
	25-2D	Jane D. Grant	615-898-9909	218 Clark Blvd., Nashville, TN 36362	$12,500,000.00
	25-5A	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	$32,512,420.00
	25-9T	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$21,563,234.00
	27-4Q	George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185	$10,314,545.00
	29-2D	Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123	$25,559,999.00
	31-7P	William K. Moor	904-445-2719	216 Morton Rd., Stetson, FL 30155	$56,850,000.00

Assume Mr. George F. Dorts leaves the company, hence his details need to be deleted. Assume a deletion can only be made to an entire row, not just a removal of cell's content.

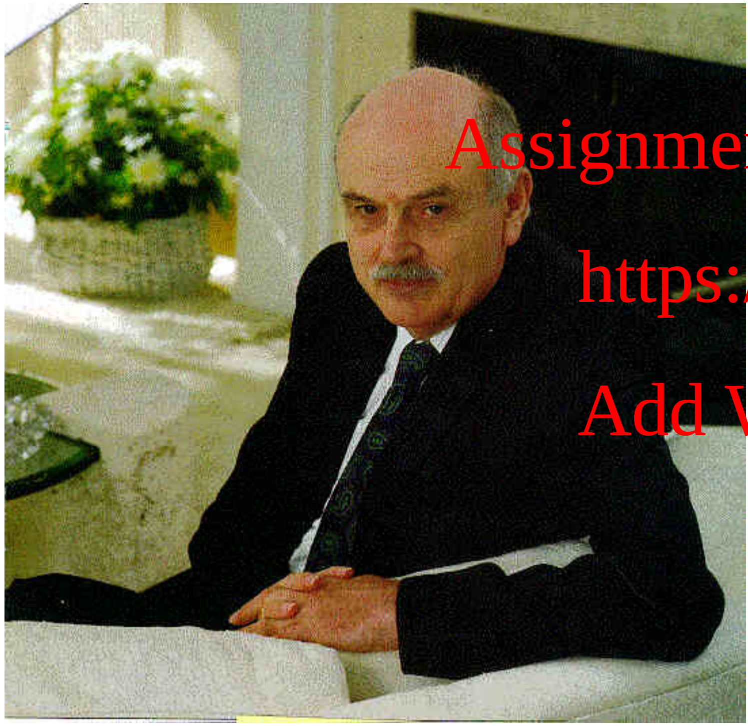
Q3. What would be the potential issue related to this deletion?

a. None.
b. All projects' details that managed by Dorts will be lost.
c. The database becomes smaller.
d. All the details of Dorts will be lost.
e. None of the above.

Introduction

- The Relational Data model was first proposed by E.F. Codd in 1970.
- Now the dominant model for commercial database implementations.
- Sound theoretical foundation.
- Examples of RDBMS products:
  - Oracle
  - INGRES
  - DB2
  - Microsoft Access

Edgar F. Codd



Dr. E. F. Codd

- 1923-2003.
- BA/MA (Maths) Oxford University
- PhD University of Michigan.

The Relational Oath:

"I promise to use the key, the whole key and nothing but the key, so help me Codd"

Basic Constructs

- The **relation** is a named table with columns and rows.

- An **attribute** is a named column of a relation.

- The **domain** of an attribute is the set of values the attribute may take.

- A **tuple** is a row of a relation.
- No. of attributes = **degree** of Relation
- No. of tuples = **cardinality** of Relation

Properties of Relations

- Relations exhibit several fundamental properties:
  - Each row (tuple) is unique - i.e. duplicate tuples are not allowed.
  - Each column has a (meaningful) name.
  - All the values in a column are values of a single attribute.
  - The order of attributes is immaterial.
  - The order of tuples is immaterial.
  - The entries are single-valued - each cell contains a single entry
  - Any value is addressable by specifying the name of the table, the primary key value for the relevant row, and the name of the column.

Q4. Which of the following statements is TRUE according the characteristics of relational table?

a. All values in a column need to be from the same domain.

b. Each column needs to have a distinct name.
c. The order of attribute(column) and tuple(row) matters.

d. Each intersection of a column and a row represent a single value.
e. More than one statement is TRUE.

surname	firstname	degree	DOB
Black	Sam	BBIS	02-02-1996
Brown	Jane	BITS	01-01-1995
Chen	Chan	BITS	09-02-1996
Grey	Maria	BCS	15-12-1995
Indigo	Jose	BITS	28-10-1995
Black	Jet	BCS	13-05-1996
Chen	Maria	BBIS	31-08-1995

Q5. Which of the following statement is TRUE when the concept of functional dependency is applied to the above data? **Assume this data represents the entire table and is not going to change in any way.**

a. Surname determines firstname.

b. (surname and firstname) determine degree.

c. DOB determines surname.

d. degree determines surname.

e. (firstname and degree) determine surname.

f. More than one statement is TRUE.

surname	firstname	degree	DOB
Black	Sam	BBIS	02-02-1996
Brown	Jane	BITS	01-01-1995
Chen	Chan	BITS	09-02-1996
Grey	Maria	BCS	15-12-1995
Indigo	Jose	BITS	28-10-1995
Black	Jet	BCS	13-05-1996
Chen	Maria	BBIS	31-08-1995

Q6. What attribute (column) or combination of columns that can be used to uniquely identify each row in the STUDENT table (relation). **Assume this data represents the entire table and is not going to change in any way.**

a. Combination of surname, firstname, degree, DOB.
b. Combination of surname and firstname
c. Combination of surname, firstname and degree
d. All of them.

Relational Keys

- A **candidate** key K of a relation R is an attribute or set of attributes which exhibits the following properties:
  - No two tuples of R have the same value for K (Uniqueness property)
  - No proper subset of K has the uniqueness property (Minimality or Irreducibility property)
- One candidate key is chosen to be the **primary key** of a relation. Remaining candidate keys are termed **alternate keys**.
- A **superkey** is an attribute or set of attributes which only exhibits the uniqueness property.

surname	firstname	degree	DOB
Black	Sam	BBIS	02-02-1996
Brown	Jane	BITS	01-01-1995
Chen	Chan	BITS	09-02-1996
Grey	Maria	BCS	15-12-1995
Indigo	Jose	BITS	28-10-1995
Black	Jet	BCS	13-05-1996
Chen	Maria	BBIS	31-08-1995

Q7. Based on the data provided in the above table, what is/are the candidate key(s) for the above relation(table)?

a. Combination of surname and firstname.
b. Surname.
c. Combination of surname, firstname and degree.
d. Firstname.

stu_no	surname	firstname	degree	DOB
1111	Black	Sam	BBIS	02-02-1996
1112	Brown	Jane	BITS	01-01-1995
1113	Chen	Chan	BITS	09-02-1996
1114	Grey	Maria	BCS	15-12-1995
1115	Indigo	Jose	BITS	28-10-1995
1116	Black	Jet	BCS	13-05-1996
1117	Chen	Maria	BBIS	31-08-1995

Q8. Based on the data provided in the above table, what is/are the candidate key(s) for the above relation(table)?

a. stu_no.
b. Combination of surname and firstname.
c. Combination of surname, firstname and stu_no.
d. More than one answer is correct.

stu_no	surname	firstname	degree	DOB	u_code
1111	Black	Sam	BBIS	02-02-1996	FIT1004
1112	Brown	Jane	BITS	01-01-1995	FIT1004
1113	Chen	Chan	BITS	09-02-1996	FIT1001
1114	Grey	Maria	BCS	15-12-1995	FIT1001
1115	Indigo	Jose	BITS	28-10-1995	FIT1004
1116	Black	Jet	BCS	13-05-1996	FIT1001
1117	Chen	Maria	BBIS	31-08-1995	FIT1004

Q9. A Primary Key is defined as “A candidate key that is selected to identify tuples uniquely within a relation”. How many primary keys does the above table have?

a.1
b.2
c.3
d.0

Writing Relations

- Relations may be represented using the following notation:
  - relation_name(attribute1, attribute2,…)

- The primary key is underlined.

Example:

staff(<u>staff-id</u>,surname,initials,address,phone)

Foreign Keys

- *A foreign key is an attribute or a set of attributes within one relation defined over the same domain as the primary key of another (possibly the same) relation.*

- Foreign keys implement relationships between tables (relations).

Where are the foreign keys in these two relations?

```
staff(staff-id,surname,initials,address,phone,dept-id,
                  supervisor-id)

department(dept-id,name)
```

MANAGER

PROJECT_MANAGER	MANAGER_PHONE	MANAGER_ADDRESS
Holly B. Parker	904-338-3416	3334 Lee Rd., Gainesville, FL 37123
Jane D. Grant	615-898-9909	218 Clark Blvd., Nashville, TN 36362
George F. Dorts	615-227-1245	124 River Dr., Franklin, TN 29185
~~Holly B. Parker~~	~~904-338-3416~~	~~3334 Lee Rd., Gainesville, FL 37123~~
~~George F. Dorts~~	~~615-227-1245~~	~~124 River Dr., Franklin, TN 29185~~
~~Holly B. Parker~~	~~904-338-3416~~	~~3334 Lee Rd., Gainesville, FL 37123~~
William K. Moor	904-445-2719	216 Morton Rd., Stetson, FL 30155

PROJECT

PROJECT_CODE	PROJECT_MANAGER	PROJECT_BID_PRICE
21-5Z	Holly B. Parker	$16,833,460.00
25-2D	Jane D. Grant	$12,500,000.00
25-5A	George F. Dorts	$32,512,420.00
25-9T	Holly B. Parker	$21,563,234.00
27-4Q	George F. Dorts	$10,314,545.00
29-2D	Holly B. Parker	$25,559,999.00
31-7P	William K. Moor	$56,850,000.00

10. In which table would you assign FK (and what attribute?) if the above two tables are to be created in a relational database?

a. MANAGER table on project_manager attribute.
b. PROJECT table on project_code attribute.
c. MANAGER table on manager_phone attribute.
d. PROJECT on project_manager.
e. None of the above.

Relational Database

- A **relational database** is a collection of **normalised** relations.

- Normalisation is part of the design phase of the database and will be discussed in a later lecture.

Example relational database:

```
order(order-id,date,)
order-line(order-id,product-id,quantity)
product-id(product-id, description, unit-price)
```

Data Integrity

- Original types:
  - Entity integrity
    - Primary Key should not be NULL.
  - Referential integrity
    - The values of FK should match the value of the PK in another relation (possibly the same) or be NULL.
  - Column/Domain integrity
    - All values in a given column has to come from the same domain (the same data type and range).
- Additional
  - NOT NULL
  - UNIQUE

Types of Tables (Relations)

- **Base table**. A stored table. A physically persistent table, stored as a file on disk when implemented in a relational database management system.
- **Derived table**. A temporary table produced as a result of a query on one or more base tables, or the invocation of a view. Exists for the duration of the operation that engenders it.
- **View**. A virtual table – Stored as a query which when invoked generates a derived table. This view can then be queried as though it were another base table (with some restrictions). Views will be discussed later in the course.

Relational Languages

- We need to have a method whereby we can specify the structure of tables and manipulate the data held in the tables.
- At a minimum we require:
  - A Data Definition Language (DDL), for specifying the structure of data.
  - A Data Manipulation Language (DML), for specifying the user's intent with respect to the use of the data – operations on data.

DDL

Create

- databases
- tables
- views
- integrity constraints
- indexes

Delete

- any of the above

Relational DMLs

- Relational Calculus
- Relational Algebra
- Transform Oriented Languages (e.g. SQL)
- Graphical Languages
- Fourth Generation Languages
- Fifth Generation Languages

- Exhibit the “closure” property - queries on relations produce relations.

Relational Calculus

- Based on mathematical logic.
- Non-procedural.
- Primarily of theoretical importance.
- May be used as a yardstick for measuring the power of other relational languages ("relational completeness").
- Operators may be applied to any number of relations.

Relational Algebra

- Relationally complete.
- Procedural.
- Operators only apply to at most two relations at a time.

8 basic operations:

- selection
- projection
- join
- Union
- Intersection
- difference
- cartesian product
- division

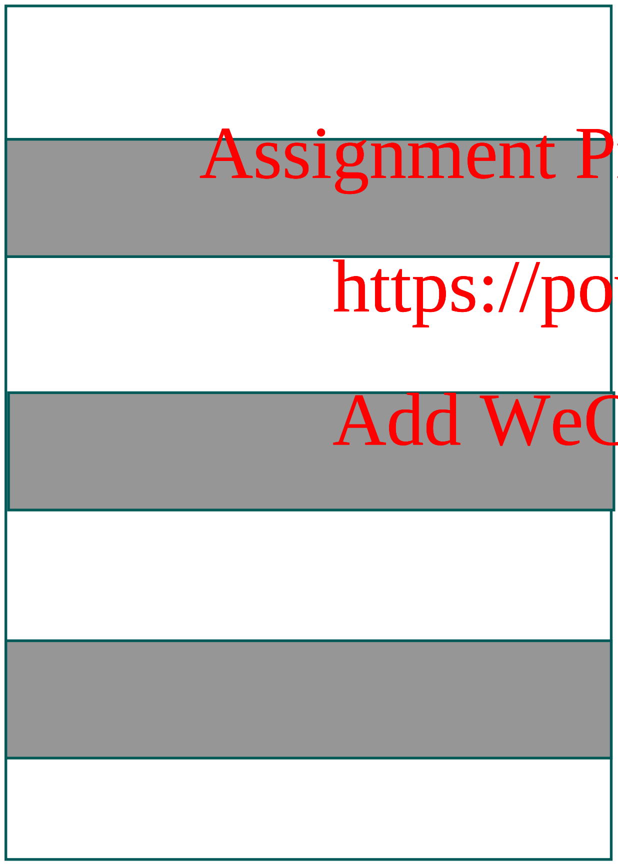
The ITEM Base Table

Item-Id	Description	Pack	Unit-Price
I26	Bolt	10	0.10
I35	Nut	100	0.05
I87	Washer	100	0.05
I22	Spanner	1	5.50
I98	Tool Box	1	21.75
I56	Hammer	1	14.95
I34	Nail	20	3.45

ITEM(Item-Id, Description, Pack, Unit-Price)

Selection

A predicate is a truth-valued function (i.e. it returns true or false)



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

- Produces a horizontal subset of R consisting of tuples which satisfy the predicate.

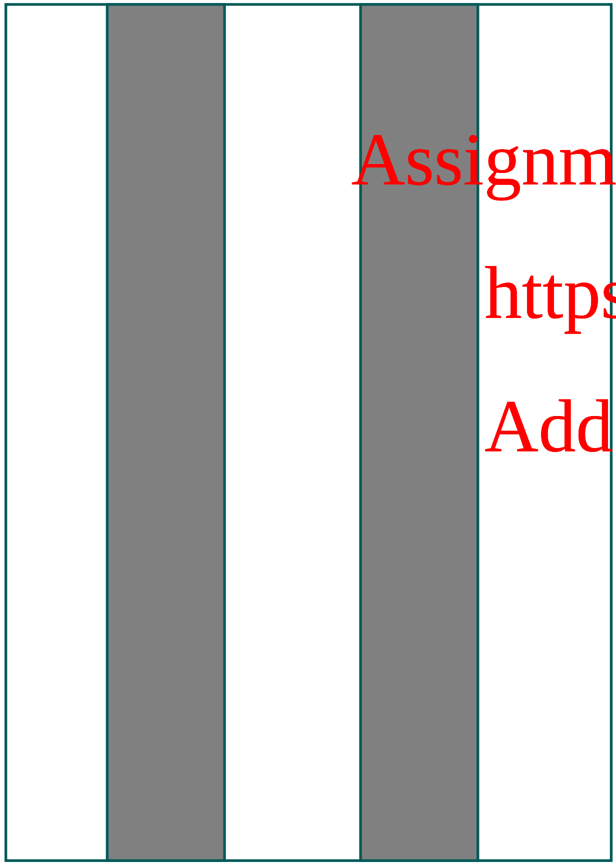
Selection Example

$$\sigma_{\text{Item-Id<“I35”}}(\text{ITEM})$$

Item-Id	Description	Pack	Unit-Price
I26	Bolt	10	0.10
I22	Spanner	1	5.50
I34	Nail	20	3.45

List items whose Item-Id is less than “I35”

Projection



$\pi_{\text{column list}}(R)$

- Produces a vertical subset of R consisting of columns in the column list

Projection Example

$$\pi_{\text{Item-Id,Description}}(\text{ITEM})$$

Item-Id	Description
I26	Bolt
I35	Nut
I87	Washer
I22	Spanner
I98	Tool Box
I56	Hammer
I34	Nail

List all Items by Item-Id and Description

Join

- Join operator used to combine data from two or more relations, based on a common attribute or attributes.
- Different types:
  - – theta-join
  - – equi-join
  - – natural join
  - – outer join

ITEM_SUPPLIER Base Table

Note that in this example not all items have associated suppliers.

We will illustrate joins between the two tables ITEM and ITEM_SUPPLIER.

Item-Id	Supp-Id
I26	S44
I87	S22
I22	S10
I98	S10
I56	S43

Theta-join

$$(\text{Relation\_1}) \bowtie_F (\text{Relation\_2})$$

F is a predicate (i.e. truth-valued function) which is of the form Relation_1.a_i θ Relation2.b_i

θ is one of the standard arithmetic comparison operators, *i.e.* $<, \leq, =, \geq, >$

Most commonly, θ is equals (=).

Equi-join example

$(\text{ITEM}) \bowtie_{\text{Item.Item-Id=Supplier.Item-Id}} (\text{ITEM\_SUPPLIER})$

Item-Id	Description	Pack	Unit-Price	Item-Id	Supp-Id
I26	Bolt	10	0.10	I26	S44
I87	Washer	100	0.05	I87	S22
I22	Spanner	1	5.50	I22	S10
I98	Tool Box	1	21.75	I98	S10
I56	Hammer	1	14.95	I56	S43

List all items and their suppliers id

MONASH University
Information Technology

www.it.monash.edu

Natural Join Example

(ITEM) $\bowtie$ (ITEM_SUPPLIER)

Item-Id	Description	Pack	Unit-Price	Supp-Id
I26	Bolt	10	0.10	S44
I87	Washer	100	0.05	S22
I22	Spanner	1	5.50	S10
I98	Tool Box	1	21.75	S10
I56	Hammer	1	14.95	S43

Duplicate attributes are eliminated

Outer join example

(ITEM) ⟕ (ITEM_SUPPLIER)

Item-Id	Description	Pack	Unit-Price	Item-Id	Supp-Id
I26	Bolt	10	0.10	I26	S44
I87	Washer	100	0.05	I87	S22
I22	Spanner	1	5.50	I22	S10
I98	Tool Box	1	21.75	I98	S10
I56	Hammer	1	14.95	I56	S43
I35	Nut	100	0.05		
I34	Nail	20	3.45		

List all items and suppliers including where there is no supplier

The (left) outer join is a join in which tuples from Relation_1 (ITEM) which do not have matching tuples in Relation_2 (ITEM_SUPPLIER) are included.

Union, Intersection, Difference

- These three operators require UNION-compatible tables.
- Two tables are said to be UNION-compatible if they have the same structure, that is, they have the same number of columns, and corresponding columns are defined over the same domain.
- UNION - an operator that results in a table containing all the rows from both tables, but with no duplicate rows.
- INTERSECTION - an operator that results in a table whose rows appear in both the contributing tables.
- DIFFERENCE - an operator that results in a table whose rows occur in the first table but not the second.

Student Table

Student-ID	Surname	Initials	Suburb	Year Mark
1005	Green	GG	Oakwood	56
1017	Brown	BB	Elmwood	87
1022	Gold	GG	Ashwood	68
1014	White	WW	Elmwood	59
1003	Violet	VV	Ashwood	77
1001	Blue	BB	Elmwood	76
1020	Black	BB	Oakwood	80
1021	Red	RR	Oakwood	89
1008	Orange	OO	Ashwood	75
1023	Yellow	YY	Oakwood	64

Teacher Table

Teacher-ID	Surname	Given Name	Suburb	Quals
T001	Silver	Sylvia	Ashwood	MSc
T002	Grey	Grace	Oakwood	BA
T003	Maroon	Mark	Elmwood	BSc
T004	Beige	Beryl	Elmwood	Bbus
T005	Violet	Vincent	Ashwood	BA

Cartesian Product

- The Cartesian product *R X S* of two relations *R* and *S* results in a relation consisting of every tuple from *R* concatenated with every tuple from *S*.
- Cardinality of the Cartesian product is the product of the respective cardinalities of *R* and *S*.
- Degree of the Cartesian product is the sum of the respective degrees of *R* and *S*.

Union Compatibility

- The STUDENT and TEACHER tables are not union compatible.

- However, *projections* (vertical subsets) of the two tables may be union-compatible.

- For example:

$$\pi_{Surname,Suburb}(STUDENT)$$

and

$$\pi_{Surname,Suburb}(TEACHER)$$

Union Example

$$\pi_{Surname,Suburb}(STUDENT) \cup \pi_{Surname,Suburb}(TEACHER)$$

Surname	Suburb
Beige	Elmwood
Black	Oakwood
Blue	Elmwood
Brown	Elmwood
Gold	Ashwood
Green	Oakwood
Grey	Oakwood
Maroon	Elmwood
Orange	Ashwood
Red	Oakwood
Silver	Ashwood
Violet	Ashwood
White	Elmwood
Yellow	Oakwood

List all students and teachers with their suburbs

Note that STUDENT had 10 rows, TEACHER had 5 rows but this UNION only has 14 rows (one duplicate row eliminated).

Intersection Example

$$\pi_{Surname,Suburb}(STUDENT) \cap \pi_{Surname,Suburb}(TEACHER)$$

Surname	Suburb
Violet	Ashwood

List all students and their suburbs who have the same surname and suburb as a teacher

Difference Example

$$\pi_{Surname,Suburb}(\text{STUDENT}) - \pi_{Surname,Suburb}(\text{TEACHER})$$

List all students with their suburbs who do not have the same surname as a teacher

Surname	Suburb
Black	Oakwood
Blue	Elmwood
Brown	Elmwood
Gold	Ashwood
Green	Oakwood
Orange	Ashwood
Red	Oakwood
White	Elmwood
Yellow	Oakwood

Cartesian Product Example

$$\sigma_{\text{Item-Id<“I35”}}(\text{ITEM}) \times \sigma_{\text{Supp-Id=“S10”}}(\text{ITEM\_SUPPLIER})$$

Item-Id	Description	Pack	Unit-Price
I26	Bolt	10	0.10
I22	Spanner	1	5.50
I34	Nail	20	3.45

x

Item-Id	Supp-Id
I22	S10
I98	S10

Item-Id	Description	Pack	Unit-Price	Item-Id	Supp-Id
I26	Bolt	10	0.10	I22	S10
I26	Bolt	10	0.10	I98	S10
I22	Spanner	1	5.50	I22	S10
I22	Spanner	1	5.50	I98	S10
I34	Nail	20	3.45	I22	S10
I34	Nail	20	3.45	I98	S10

List all item information for items with item-id less than “I35” and supplied by supplier S10

Division

- The division operation *R*÷ *S* of two relations *R* and *S* results in a relation consisting of tuples defined over the attributes in the set:

{attributes of R – attributes of S}

which match *every* tuple in *S.*

Division Example

Item-Id	Supp-Id
I26	S44
I87	S22
I22	S10
I98	S10
I56	S43
I11	S10
I22	S23

÷

Item-Id
I22
I11

=

Supp-Id
S10

Which individual suppliers can supply items with item-id = “I22” and “I11”?