資料庫系統
Class 2: Relational Model

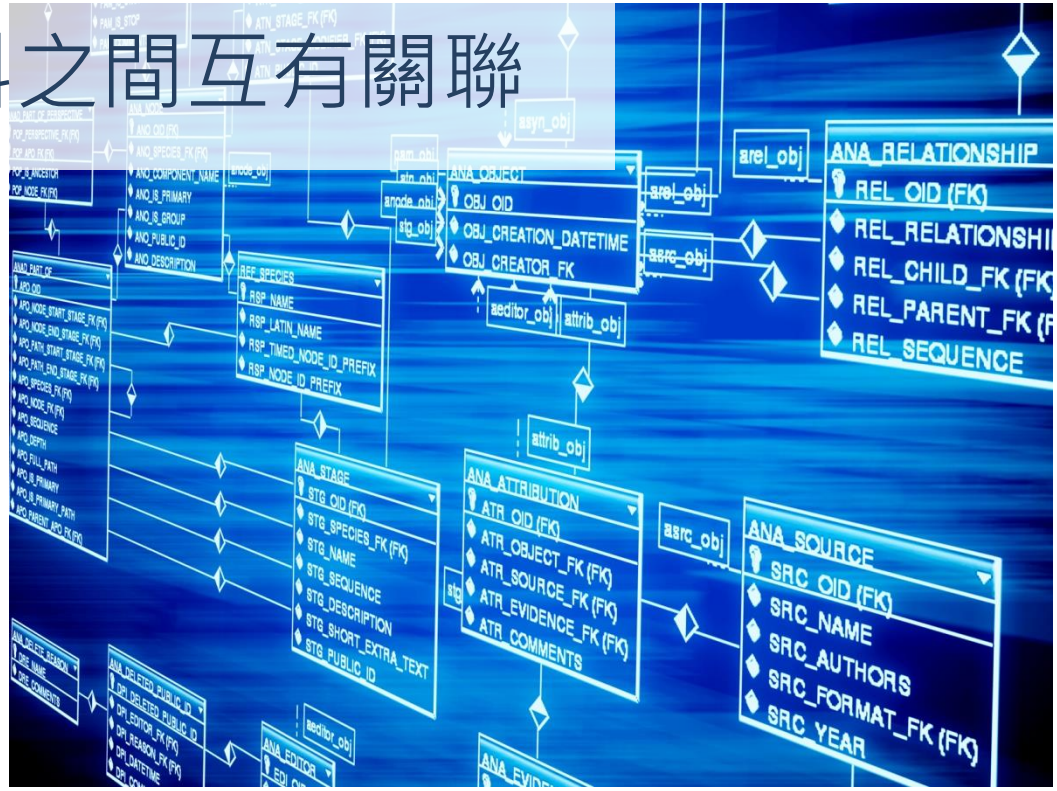
逢甲資工 許懷中

開始之前，問一下

- 知道什麼是資料庫嗎？(Database)
- 知道什麼是資料庫管理系統嗎？(DBMS)
- 知道交易 (Transaction) 有什麼特性嗎？
- 知道用來存取資料庫語言有哪幾種嗎？

什麼是資料庫 (Database)

- 資料的集合
- 一般而言，資料之間互有關聯



什麼是資料庫管理系統 (DBMS)

- 管理一個或多個資料庫 (Database)
- 存取資料庫的工具
- 令存取資料庫變得**方便**而**有效率**

交易 (Transaction) 的特性

- Transaction 是存取資料庫的基本單位
- 四大特性
  - ACID
  - Atomicity, Consistency, Isolation, Durability
  - 原子性(單元性)、一致性、獨立性、持續性

交易的特性 (cont.)

- 原子性 (Atomicity) a.k.a 單元性
  - 交易不可分割
  - 全部做完或完全沒做
  - 銀貨兩訖

交易的特性 (cont.)

- 一致性 (Consistency)
  - 一致的狀態 (Consistent State) 沒有違反任何約束 (Constraints) 的狀態
  - 交易得令資料庫從一個 Consistent State 轉移到另一個 Consistent State
  - 例如，行內轉帳前後，總金額必須相同

交易的特性 (cont.)

- 獨立性 (Isolation)
  - 同時進行的交易，其結果與順序執行相同

交易的特性 (cont.)

- 持續性 (Durability)
  - 交易一旦完成，其結果永久 (Permanent) 存在

用來操作資料庫的語言

- 資料定義語言 (DDL)
  - Data Definition Language
  - 產生資料表、定義 schema
- 資料操作語言 (DML)
  - Data Manipulation Language
  - 查找、結合

資料庫系統

關聯式資料模型

關聯式資料庫

- 資料表 (Table) 的集合

dept_name	*building*	*budget*
Biology	Watson	90000
Comp. Sci.	Taylor	100000
Elec. Eng.	Taylor	85000
Finance	Painter	120000
History	Painter	50000
Music	Packard	80000
Physics	Watson	70000

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

ID	*course_id*	*sec_id*	*semester*	*year*
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

何為關聯 (Relation)

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

Relation 的構成

欄 (column) => 屬性 (attribute)

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

列 (row) => 元組 (tuple)

tuple: 一連串的值
(3, 4, 5, 6, 7)
一個 5-tuple
(“John”, 31, “Taichung”)
一個 3-tuple

數據域 (Data Domain)

- 每個關聯 (relation) 中，一個屬性 (attribute) 可能出現的值
- 姓名：許懷中, “9527”, 9[redacted]
- 如果某屬性是不可分割的，那麼其相對應的域 (domain) 具備原子性(單元性, atomic)
- 許懷中 可不可以分割？

數據域 (Data Domain)

- 將『姓名』分割成『姓』與『名』兩個子屬性
  - E.g. 許懷中 => 許 懷中
- **在這種狀況**下『姓名』是一個可分割的屬性，那麼姓名的域即不具備原子性

架構 (Schema) vs. 實體 (Instance)

- instructor 關聯 (relation) 的 架構 (schema)
  - Instructor(ID, name, dept_name, salary)
- instructor 關聯 的 實體 (instance)

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	70000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	80000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	100000
98345	Kim	Elec. Eng.	80000

鍵 (Keys)

- 用來分辨一個關聯(relation) 中的不同元組(tuple)
- 超鍵 (superkey)
  - 可用來分辨一個關聯中不同元組的屬性(們)

超鍵 (Superkey)

ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	**Brandt**	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

ID

~~Name~~

Name, dept_name

ID, Name, salary

候選鍵 (Candidate Key)

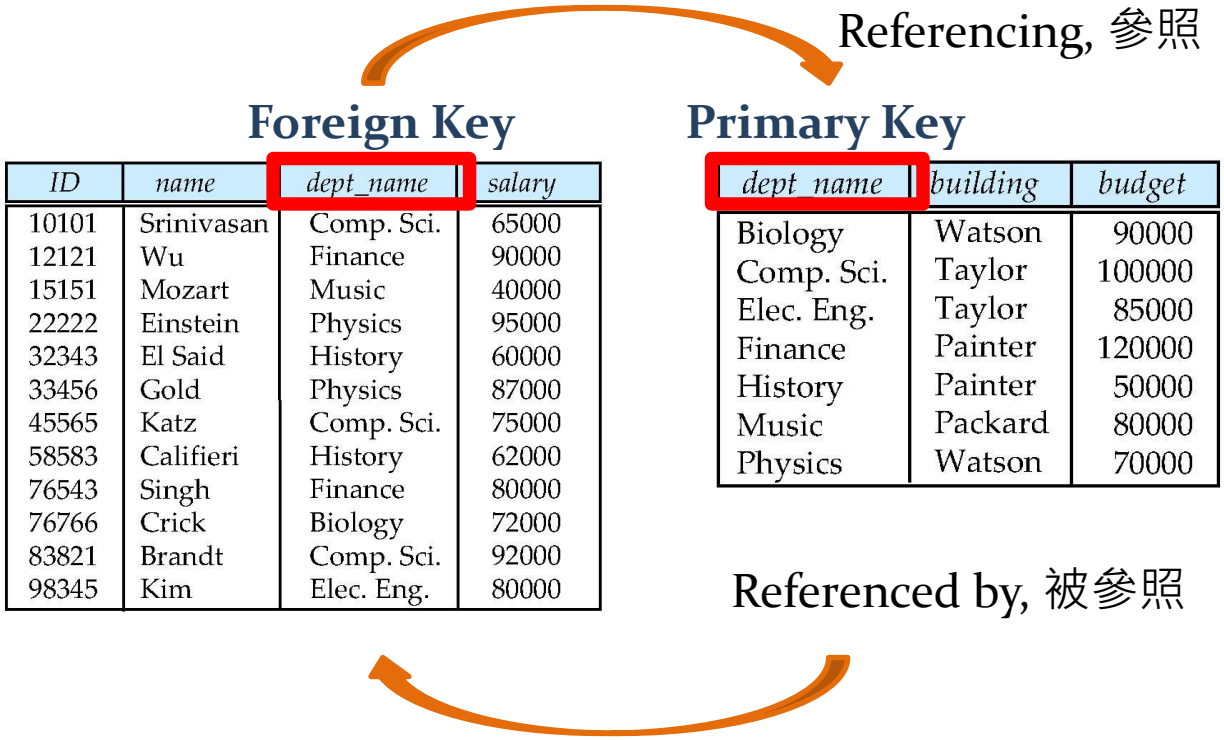
- 最少成員的超鍵 (superkey)
  - 候選鍵 (candidate key) 是超鍵
  - 候選鍵的子集合不可以是候選鍵
- 對 instructor relation 而言
  - ID 是 超鍵**也是** 候選鍵
  - name, dept_name 是超鍵**也是**候選鍵
  - ID, name, dept_name是超鍵**不是** 候選鍵
    - 因為 只需要 ID 就可以當作候選鍵

主鍵 (Primary Key)

- 被選擇來代表一個關聯 (relation) 中各個元組 (tuple) 的候選鍵 (candidate Key)
- 在一個關聯中，任意元組的主鍵不可重複
  - 這是一個約束 (constraint)
  - 表示任何對於該資料庫的操作，都不可違背
- 避免選擇可能會變更的候選鍵為主鍵

外來鍵 (Foreign Key)

- 用來參照其他關聯 (relation) 的鍵 (key)
- 必須是被參照關聯的主鍵 (Primary Key)



ID	*name*	*dept_name*	*salary*
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

dept_name	*building*	*budget*
Biology	Watson	90000
Comp. Sci.	Taylor	100000
Elec. Eng.	Taylor	85000
Finance	Painter	120000
History	Painter	50000
Music	Packard	80000
Physics	Watson	70000

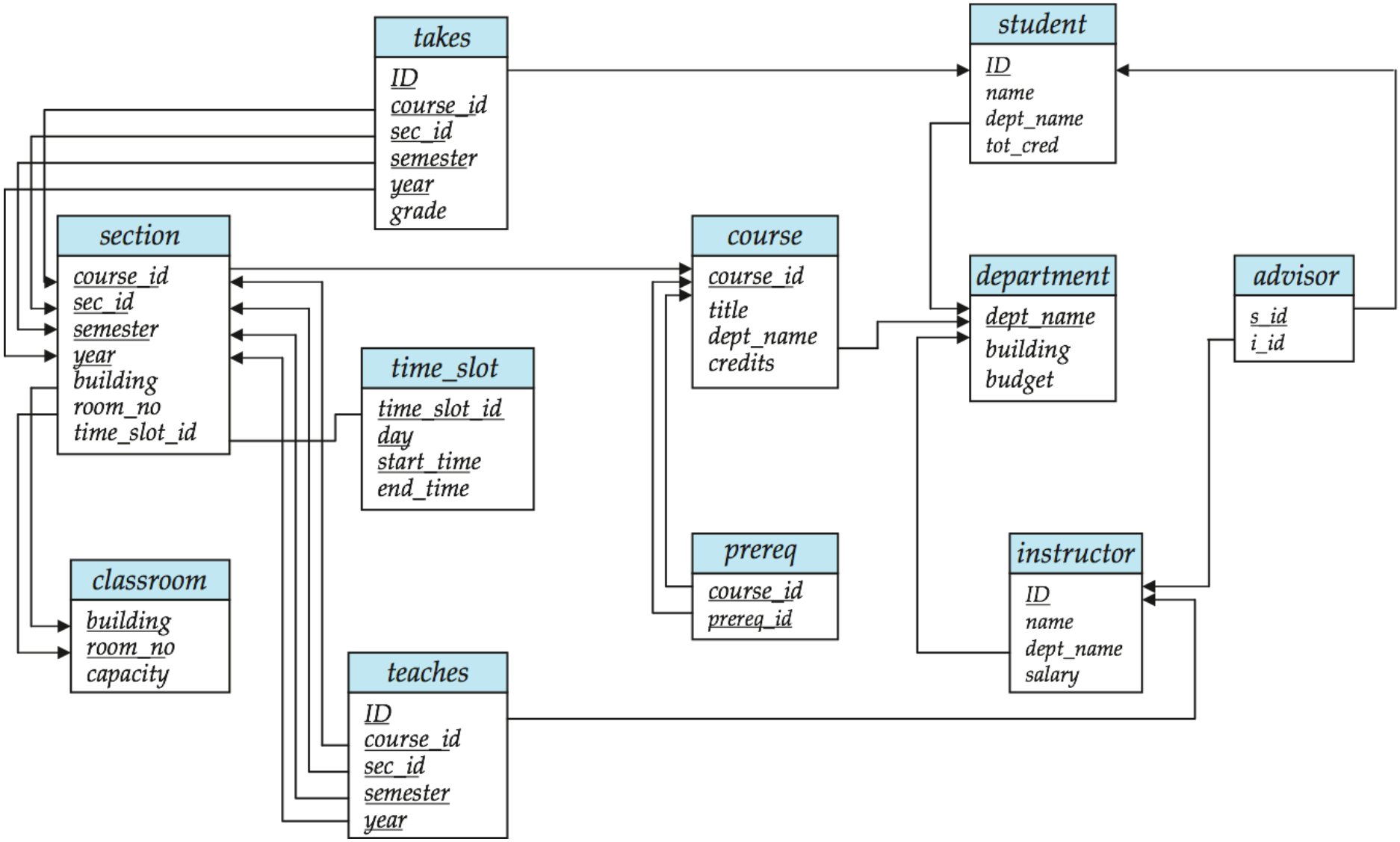
外來鍵 (cont.)

ID	*course_id*	*sec_id*	*semester*	*year*
10101	CS-101	1	Fall	2009
10101	CS-315	1	Spring	2010
10101	CS-347	1	Fall	2009
12121	FIN-201	1	Spring	2010
15151	MU-199	1	Spring	2010
22222	PHY-101	1	Fall	2009
32343	HIS-351	1	Spring	2010
45565	CS-101	1	Spring	2010
45565	CS-319	1	Spring	2010
76766	BIO-101	1	Summer	2009
76766	BIO-301	1	Summer	2010
83821	CS-190	1	Spring	2009
83821	CS-190	2	Spring	2009
83821	CS-319	2	Spring	2010
98345	EE-181	1	Spring	2009

teaches

section

course_id	*sec_id*	*semester*	*year*	*building*	*room_number*	*time_slot_id*
CS-101	1	Summer	2009	Painter	514	B
CS-101	1	Summer	2010	Painter	514	A
CS-190	1	Fall	2009	Packard	101	H
CS-190	1	Spring	2010	Packard	101	F
CS-190	2	Spring	2009	Taylor	3128	E
CS-315	2	Spring	2009	Taylor	3128	A
CS-315	1	Spring	2010	Watson	120	D
CS-319	1	Spring	2010	Watson	100	B
CS-319	2	Spring	2010	Taylor	3128	C
CS-347	1	Fall	2009	Taylor	3128	A
EE-181	1	Spring	2009	Taylor	3128	C
FIN-201	1	Spring	2010	Packard	101	B
HIS-351	1	Spring	2010	Painter	514	C
MU-199	1	Spring	2010	Packard	101	D
PHY-101	1	Fall	2009	Watson	100	A


takes
ID
course_id
sec_id
semester
year
grade
student
ID
name
dept_name
tot_cred
section
course_id
sec_id
semester
year
building
room_no
time_slot_id
course
course_id
title
dept_name
credits
department
dept_name
building
budget
advisor
s_id
i_id
time_slot
time_slot_id
day
start_time
end_time
prereq
course_id
prereq_id
instructor
ID
name
dept_name
salary
classroom
building
room_no
capacity
teaches
ID
course_id
sec_id
semester
year

Summary Quiz

- 在資料表中加入一筆資料，以關聯性資料模型的說法，應該如何描述
- 在使用者資料這個關聯中，有電話號碼這項欄位，請說明，該欄位在何種狀況下具有原子性，何種狀況下沒有
- 試說明超鍵、候選鍵、主鍵、外來鍵之間的關係

Summary

- 在關聯 (relation) 中加入一個元組 (tuple)
- 具有原子性
  - 若一個使用者只允許有一組電話號碼
- 不具有原子性
  - 一個使用者允許有多組電話號碼
  - 電話號碼可以被分割成國碼、區碼、本地號碼
- 候選鍵、主鍵都是所屬關聯的超鍵
- 主鍵是由候選鍵中選出來的
- 外來鍵是用來參照另一個關聯用的，為被參照關聯的主鍵

關聯式代數 (Relational Algebra)

- Selection (σ)
- Projection (Π)
- Cartesian Product (×)
- Join
  - Natural Join (⋈)
  - Semijoin (⋉)(⋊)
  - Antijoin (▷)

Selection (σ)

Relation r

A	B	C	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

$$\sigma_{A=B \text{ and } D>5}(r)$$

A	B	C	D
α	α	1	7
β	β	23	10

Projection (σ)

Relation r

A	B	C	D
α	α	1	7
α	β	5	7
β	β	12	3
β	β	23	10

$\Pi_{A,D}(r)$

A	D
α	7
α	7
β	3
β	10

A	D
α	7
β	3
β	10

Cartesian Product (×)

Relation r

A	B
α	1
β	2

Relation s

C	D	E
α	10	a
β	20	b
γ	30	c

r × s

A	B	C	D	E
α	1	α	10	a
α	1	β	20	b
α	1	γ	30	c
β	2	α	10	a
β	2	β	20	b
β	2	γ	30	c

Natural Join (⋈)

superheroes

name	alignment	gender	publisher
Magneto	bad	male	Marvel
Storm	good	female	Marvel
Mystique	bad	female	Marvel
Batman	good	male	DC
Joker	bad	male	DC
Catwoman	bad	female	DC
Hellboy	good	male	Dark Horse Comics

publishers

publisher	yr_founded
DC	1934
Marvel	1939
Image	1992

superheroes ⋈ publishers

name	alignment	gender	publisher	yr_founded
Magneto	bad	male	Marvel	1939
Storm	good	female	Marvel	1939
Mystique	bad	female	Marvel	1939
Batman	good	male	DC	1934
Joker	bad	male	DC	1934
Catwoman	bad	female	DC	1934

*from https://stat545-ubc.github.io/bit001_dplyr-cheatsheet.html

Semijoin (⋉, left-join)

superheroes

name	alignment	gender	publisher
Magneto	bad	male	Marvel
Storm	good	female	Marvel
Mystique	bad	female	Marvel
Batman	good	male	DC
Joker	bad	male	DC
Catwoman	bad	female	DC
Hellboy	good	male	Dark Horse Comics

publishers

publisher	yr_founded
DC	1934
Marvel	1939
Image	1992

superheroes ⋉ publishers

name	alignment	gender	publisher	yr_founded
Magneto	bad	male	Marvel	1939
Storm	good	female	Marvel	1939
Mystique	bad	female	Marvel	1939
Batman	good	male	DC	1934
Joker	bad	male	DC	1934
Catwoman	bad	female	DC	1934
Hellboy	good	male	Dark Horse Comics	NA

*from https://stat545-ubc.github.io/bit001_dplyr-cheatsheet.html

Semijoin (⋉, left-join)

publishers

publisher	yr_founded
DC	1934
Marvel	1939
Image	1992

superheroes

name	alignment	gender	publisher
Magneto	bad	male	Marvel
Storm	good	female	Marvel
Mystique	bad	female	Marvel
Batman	good	male	DC
Joker	bad	male	DC
Catwoman	bad	female	DC
Hellboy	good	male	Dark Horse Comics

publishers ⋉ superheroes

publisher	yr_founded	name	alignment	gender
DC	1934	Batman	good	male
DC	1934	Joker	bad	male
DC	1934	Catwoman	bad	female
Marvel	1939	Magneto	bad	male
Marvel	1939	Storm	good	female
Marvel	1939	Mystique	bad	female
Image	1992	NA	NA	NA

*from https://stat545-ubc.github.io/bit001_dplyr-cheatsheet.html

Antijoin (▷)

superheroes

name	alignment	gender	publisher
Magneto	bad	male	Marvel
Storm	good	female	Marvel
Mystique	bad	female	Marvel
Batman	good	male	DC
Joker	bad	male	DC
Catwoman	bad	female	DC
Hellboy	good	male	Dark Horse Comics

publishers

publisher	yr_founded
DC	1934
Marvel	1939
Image	1992

superheroes ▷ publishers

name	alignment	gender	publisher
Hellboy	good	male	Dark Horse Comics

*from https://stat545-ubc.github.io/bit001_dplyr-cheatsheet.html