

資料庫導論

朱克剛



何謂資料庫

■ 何謂「庫」？

- 車庫：放車子的地方
- 倉庫：放雜物的地方
- 彈藥庫：放彈藥的地方
- 資料庫：放資料的地方

■ 何謂「資料」？

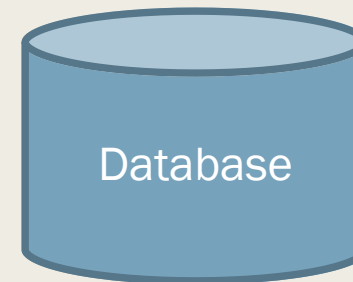
- 在電腦中所有可以被儲存起來的東西都是資料
 - 例如：一張圖片、一段聲音、一篇文章...等

■ 目的

- 大量資料下，加快資料存取速度

資料庫的「建築材料」

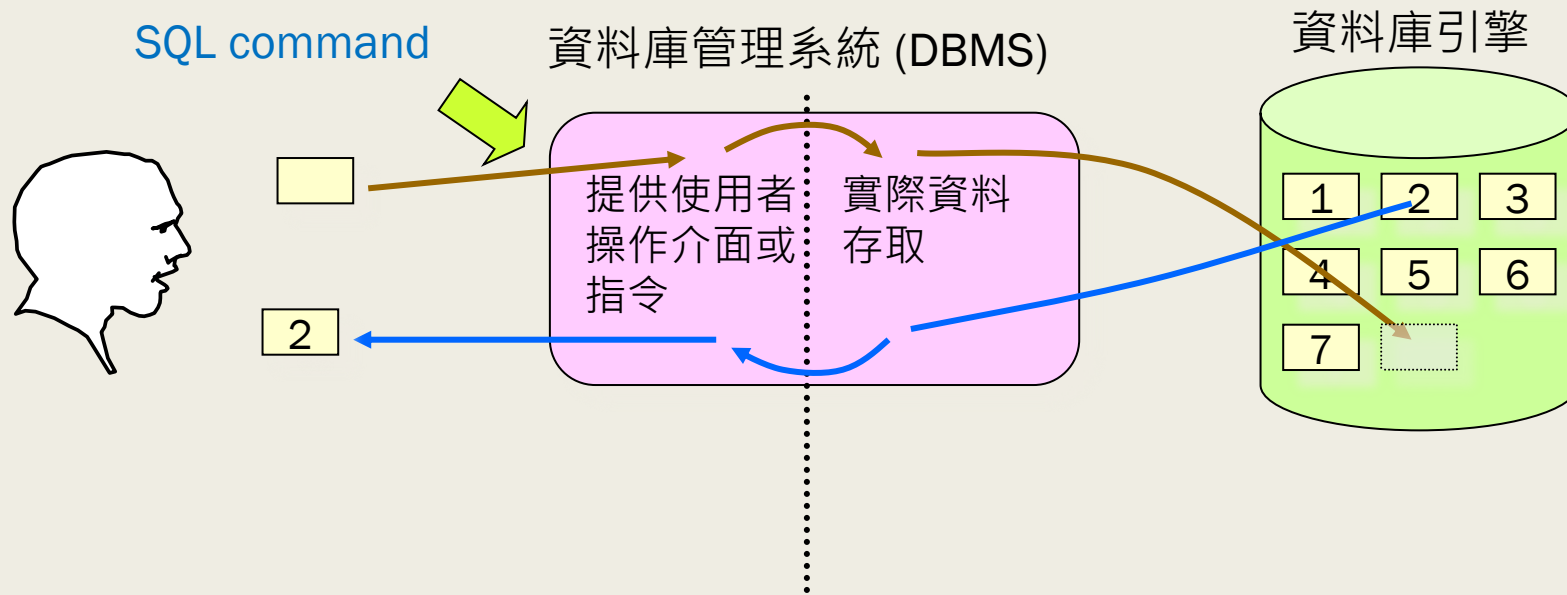
- 建築材料：
 - 車庫、倉庫用的是鋼筋水泥
 - 彈藥庫用的可能是鋼板
 - 資料庫用的是檔案
- 檔案：
 - SQL Server: .mdf
 - Access: .mdb
 - Oracle: .dbf
 - SQLite: .sqlite



資料庫類型

- 關連式 (relational database)
 - 著重在資料與資料分割後的關連性
 - 儲存空間小，規劃時間長
 - 例如：MySQL、SQL Server、Oracle、Access... 等
- NoSQL (not-only SQL)
 - 資料類型多元，不需調整成二維矩陣形式
 - 不在意儲存空間需求，規劃時間短
 - 例如：MongoDB、Apache Cassandra

資料庫管理系統



SQL command 種類

- 資料定義語言 (DDL, data definition language)
 - Create: 建立資料庫物件
 - Alter: 變更資料庫物件
 - Drop: 刪除資料庫物件
- 資料操作語言 (DML, data manipulation language)
 - Insert Into: 插入資料
 - Update: 修改資料
 - Delete: 刪除資料
- 資料查詢語言 (DQL, data query language)
 - Select: 查詢資料
- 資料控制語言 (DCL, data control language)
 - Grant: 設定權限
 - Revoke: 取消權限
 - Commit: 確認交易完成
 - Rollback: 交易失敗回復原始狀態

關連式資料庫的組成

- 實體 (entity) -> 資料表 (table)
- 屬性 (attribute) -> 欄位 (field)
- 關連性 (relationship) -> 心中

通訊錄

■ 姓名 / 住址 / 電話

- 請用 Excel 建
- Excel 的工作表相當於具有一個資料表的資料庫

cname	address	tel
王大明	台中市台灣大道一段 50 號	1111
李大媽	台中市台灣大道一段 50 號	1111
王小毛	台中市台灣大道一段 50 號	1111

問題

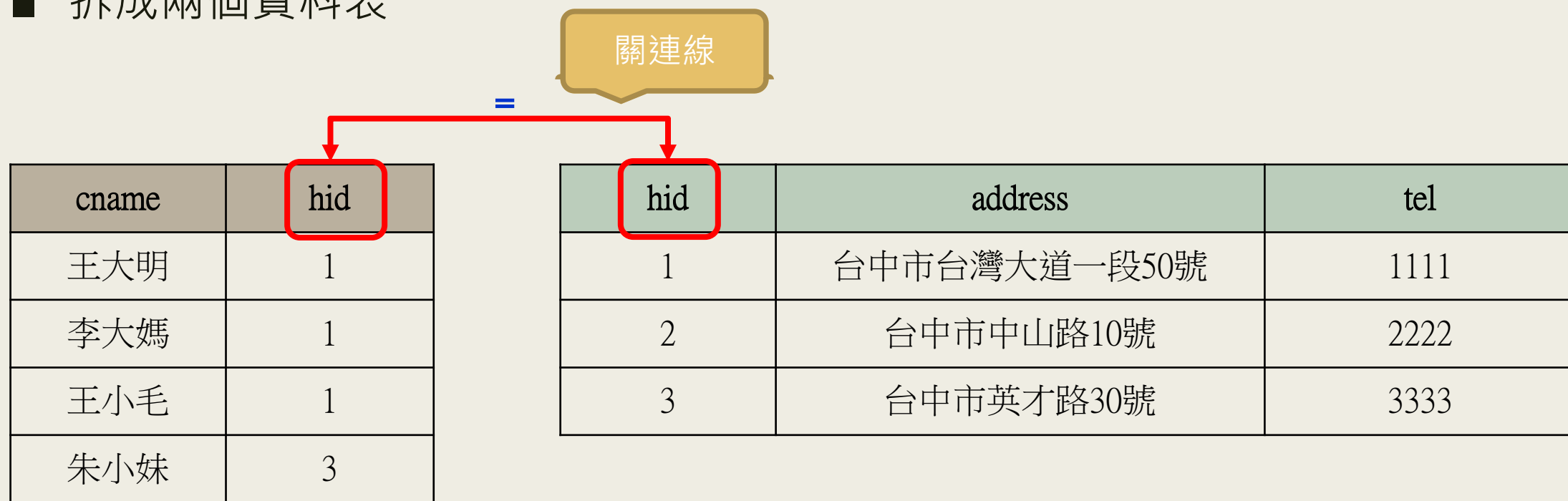
- 資料大量重複造成容易出錯、維護困難且浪費儲存空間

cname	address	tel
王大明	台中市台灣大道1段 50 號	1111
李大媽	台中市台灣大道一段 50 號	1111
王小毛	台中市台灣大道一段50號	1111

這是三個不同的地址，
看的出來嗎？

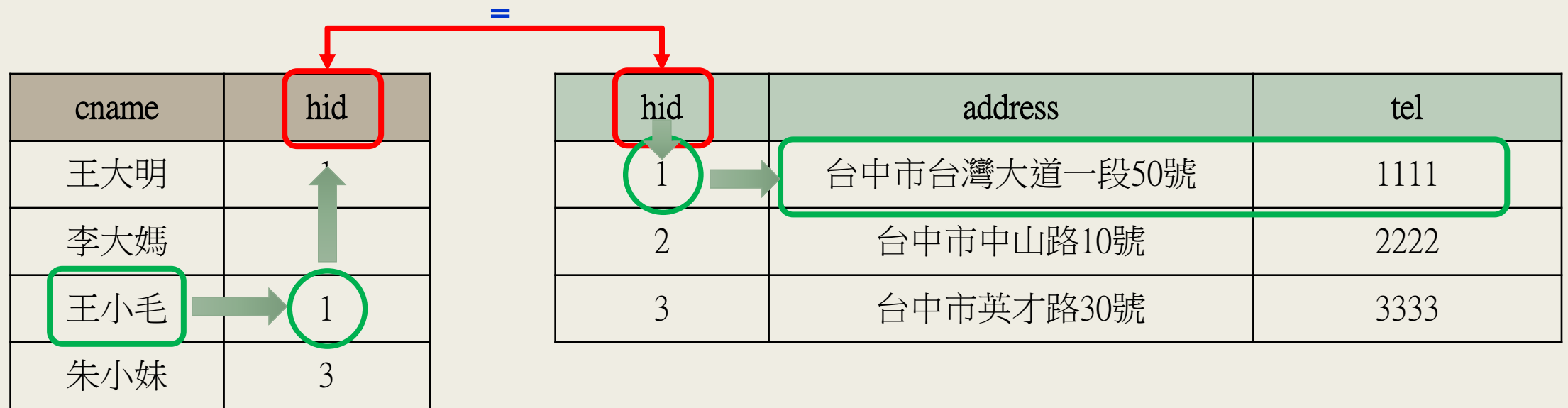
解決

■ 拆成兩個資料表



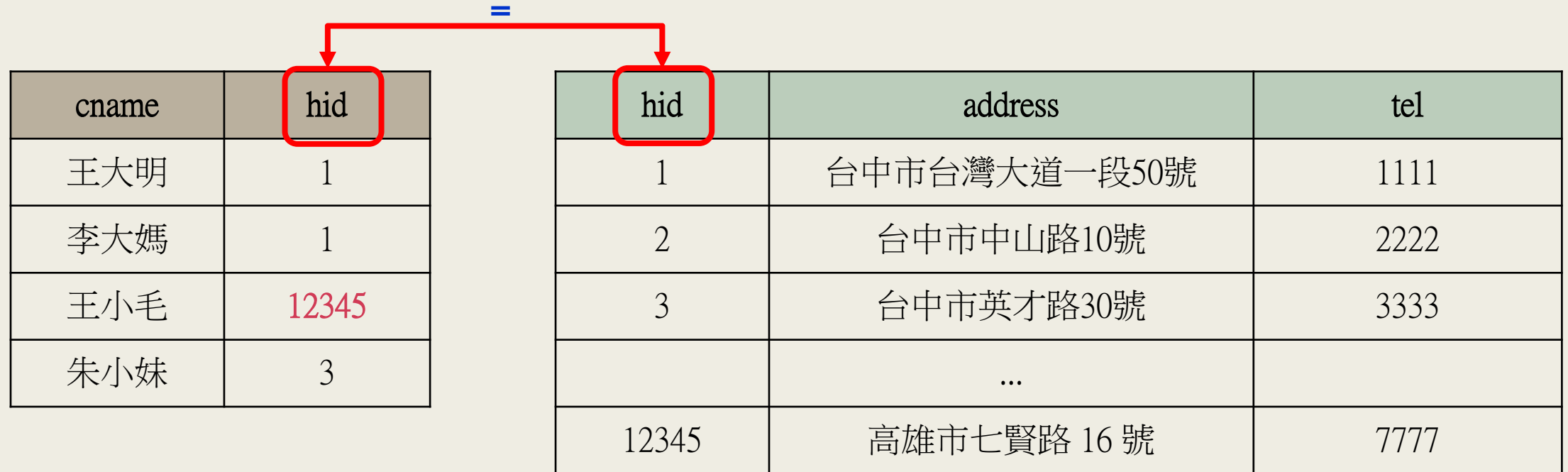
查詢軌跡

- 王小毛住哪以及電話為何
- 搜尋起點為何？搜尋終點為何？需要哪些資料表？



資料維護容易

- 王小毛搬到高雄工作，只要找到王小毛的資料，更新hid內容即可



索引 (Index)

- 唯一目的：加快資料搜尋速度
 - 以空間換取時間
 - 想想原文書最後的索引就會知道運作原理了
- 特性：
 - 一個資料表可設定的索引數目不限
 - 索引可由一個或一個以上的欄位所組成
 - 例如：可設「cname」或是「cname + hid」
- 建議設索引的欄位：
 - 作為查詢「條件」的欄位
 - 關連線兩端的欄位
 - 作為排序的欄位

主索引、主鍵、Primary Key、PK

■ 目的：

- 設成主索引的欄位其值：

- 不可重複
- 不可空白不填

■ 特性

- 主索引通常預設是索引
- 可由一個或一個以上的欄位構成

■ 每個資料表都應該有主索引

加上底線
代表PK

<u>cname</u>	hid
王大明	1
李大媽	1
王小毛	1
朱小妹	3

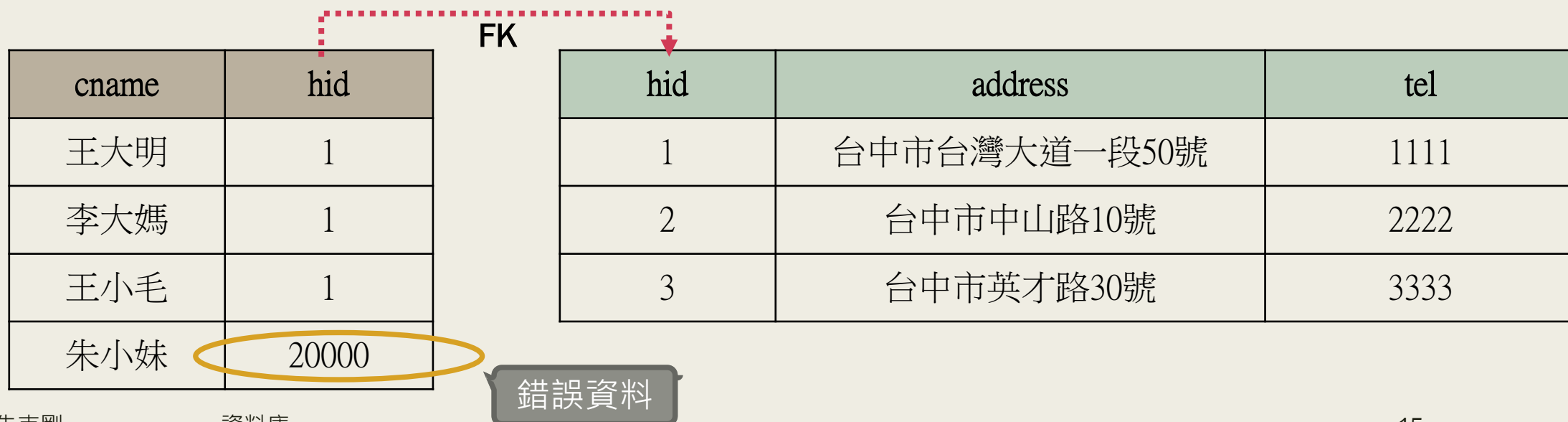
參考索引、外來鍵、Foreign Key、FK

■ 目的：

- 維持資料正確性

■ 特性：

- 參考欄位的值只能在被參考欄位值的範圍內
- 被參考欄位必須為主索引或不可重複索引
- 如未設定"串接刪除" (cascade delete) ，則無法直接刪除被參考資料



正規化

■ 資料庫設計時的遵循準則

- 遵守越多資料越破碎，越少資料重複性越高
- 實務上通常遵守三個正規化：1NF、2NF與3NF

	UNF (1970)	1NF (1971)	2NF (1971)	3NF (1971)	EKNF (1982)	BCNF (1974)	4NF (1977)	ETNF (2012)	5NF (1979)	DKNF (1981)	6NF (2003)
主鍵（無重複元組）	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
沒有重複組	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
欄位原子性（元組只有一個值）	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
沒有部分函數相依（值相依於每個主鍵這一整體）	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
沒有傳遞函數相依（值僅相依於候選鍵）	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓
每個非平凡的函數相依涉及一個超鍵或者主鍵的子鍵	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	不適用
沒有函數相依造成的冗餘	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	不適用
每個非平凡的多值相依都有一個超鍵	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	不適用
超鍵是每個顯式連接相依的一部分 ^[1]	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	不適用
候選鍵隱含了每個非平凡的連接相依關係	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	不適用
每個約束都是域約束和鍵約束的結果	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	不適用
每個連接相依都是平凡的	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓

第一正規化

- 欄位內只能放一筆資料
- 例如王大明有兩棟屋子...

cname	hid
王大明	1、2

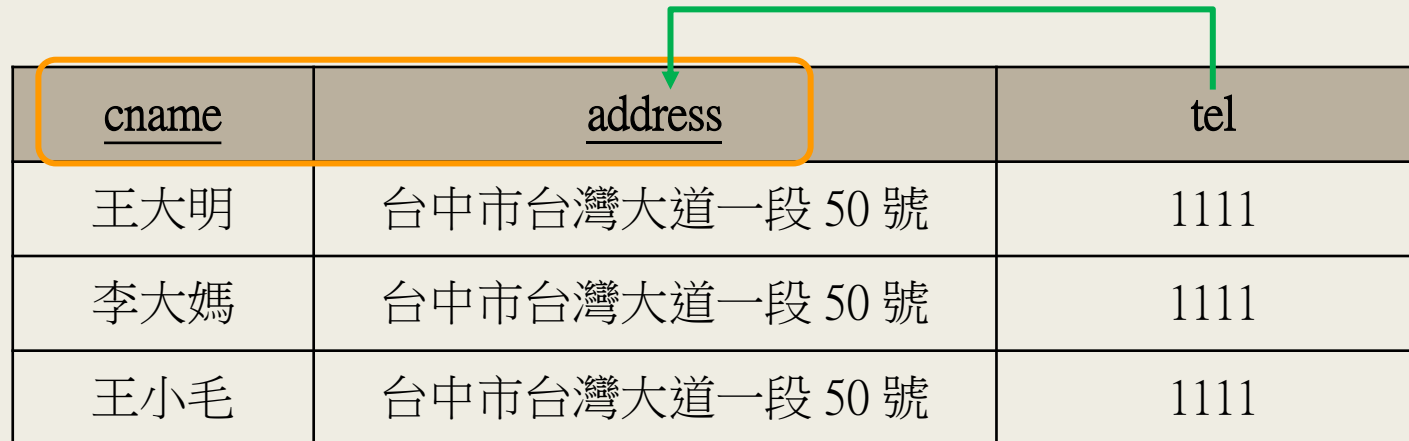
✗違反第一正規化

cname	hid
王大明	1
王大明	2

遵守第一正規化

第二正規化

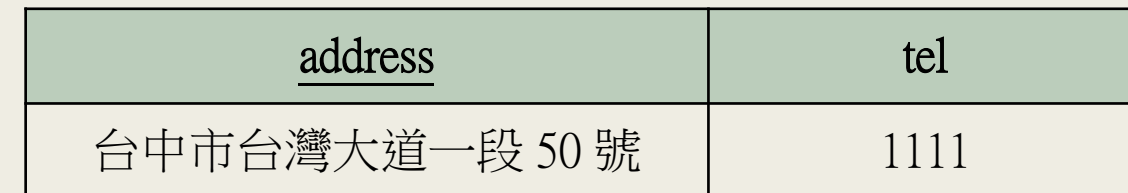
- 每個欄位都必須與主索引完全有關
- 例如此資料表 cname + address 為主索引時，tel 只與 address 有關



<u>cname</u>	<u>address</u>	tel
王大明	台中市台灣大道一段 50 號	1111
李大媽	台中市台灣大道一段 50 號	1111
王小毛	台中市台灣大道一段 50 號	1111



<u>cname</u>	<u>address</u>
王大明	台中市台灣大道一段 50 號
李大媽	台中市台灣大道一段 50 號
王小毛	台中市台灣大道一段 50 號



<u>address</u>	tel
台中市台灣大道一段 50 號	1111

第三正規化

此為電話帳單地址，與tel有關，而tel非PK或部分PK

- 非主鍵欄位不可彼此有關
- hid為PK

<u>hid</u>	address	tel	bill_hid
1	台中市台灣大道一段50號	1111	1
2	台中市中山路10號	2222	2
3	台中市英才路30號	3333	1

<u>hid</u>	address	tel
1	台中市台灣大道一段50號	1111
2	台中市中山路10號	2222
3	台中市英才路30號	3333

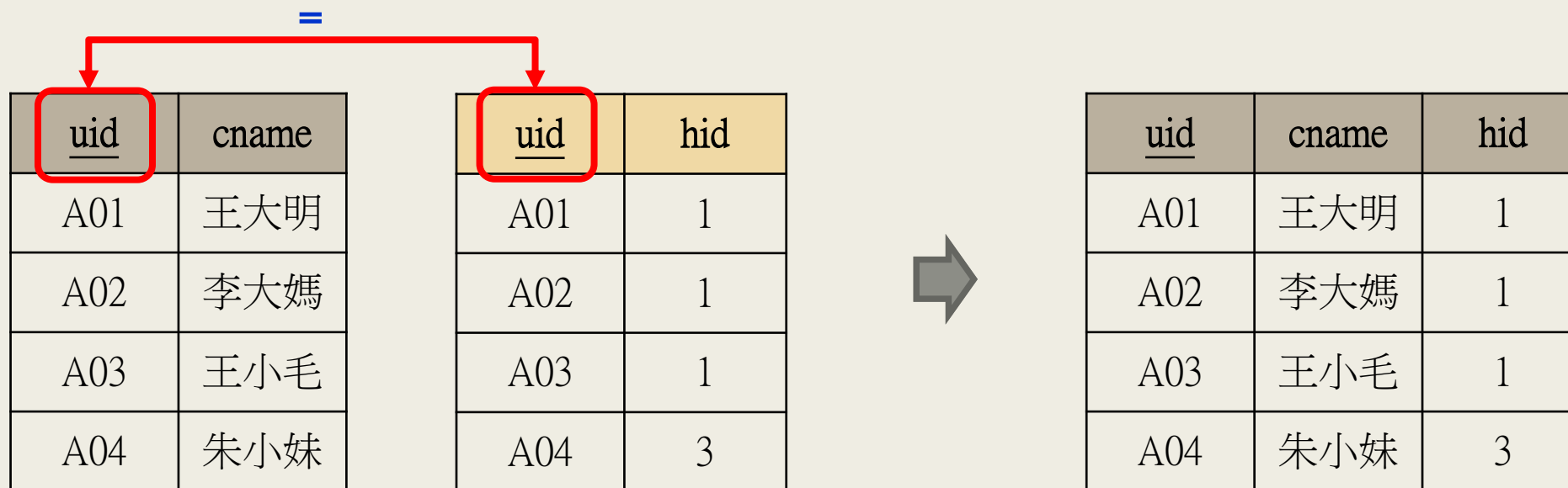
<u>tel</u>	hid
1111	1
2222	2
3333	1

關連種類

- 一對一
 - 通常可合併成一個資料表
- 一對多、多對一
 - 兩個資料表
- 多對多
 - 三個資料表

一對一

- 關連線兩端都是主索引



一對多、多對一

- 一個人只能有一棟屋子，一棟屋子中可以住很多人

<u>uid</u>	cname	hid
A01	王大明	1
A02	李大媽	1
A03	王小毛	1
A04	朱小妹	3

<u>hid</u>	address	tel
1	台中市台灣大道一段50號	1111
2	台中市中山路10號	2222
3	台中市英才路30號	3333

多對多

- 一個人可以有很多棟屋子，一棟屋子中也可以住很多人

<u>uid</u>	cname
A01	王大明
A02	李大媽
A03	王小毛
A04	朱小妹

<u>uid</u>	<u>hid</u>
A01	1
A01	2
A02	1
A03	1
A04	3

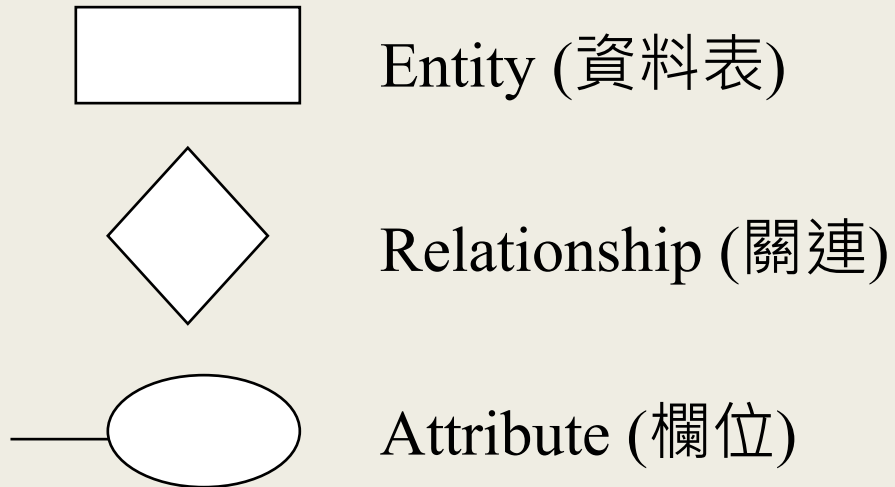
<u>hid</u>	address	tel
1	台中市台灣大道一段50號	1111
2	台中市中山路10號	2222
3	台中市英才路30號	3333

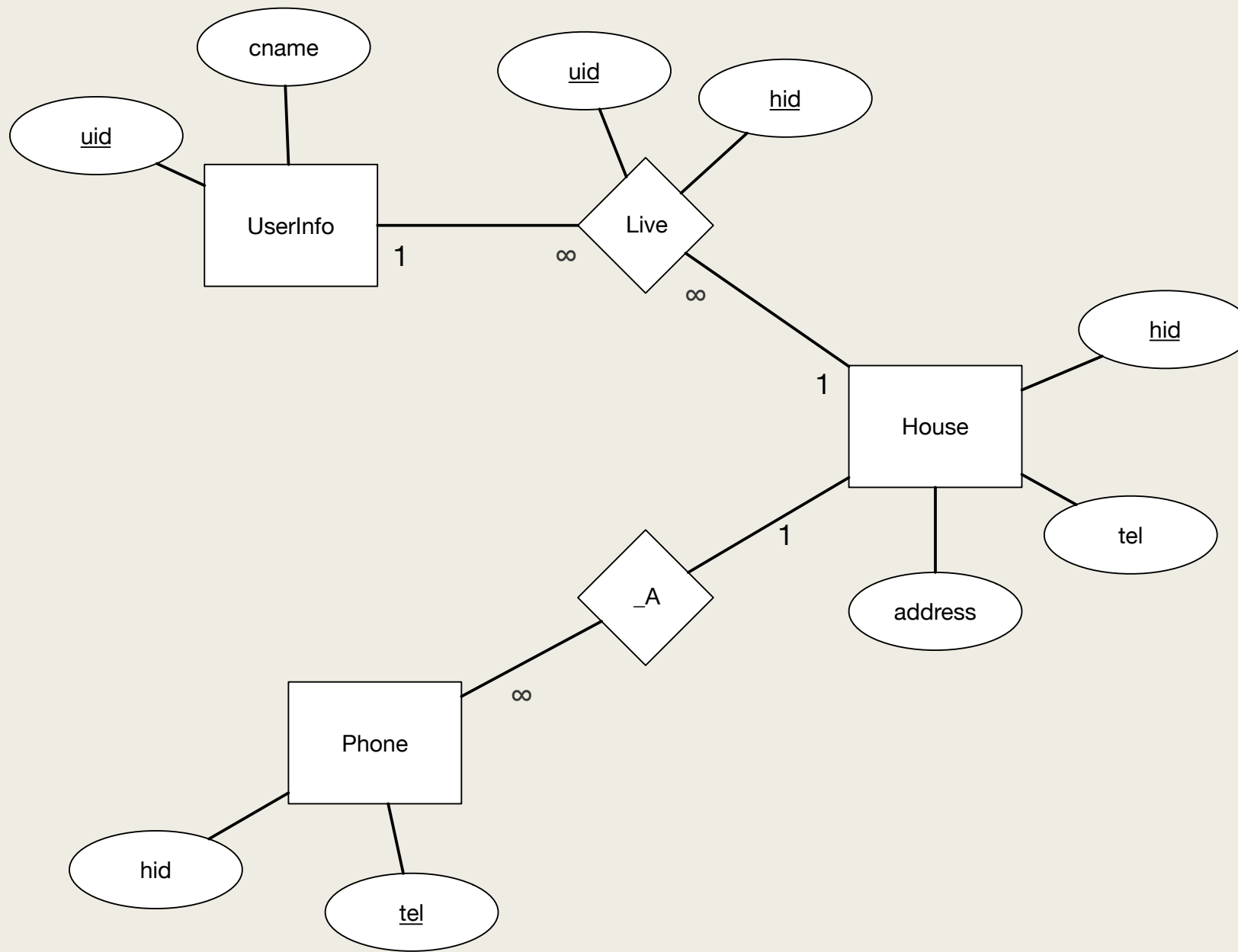
關連圖

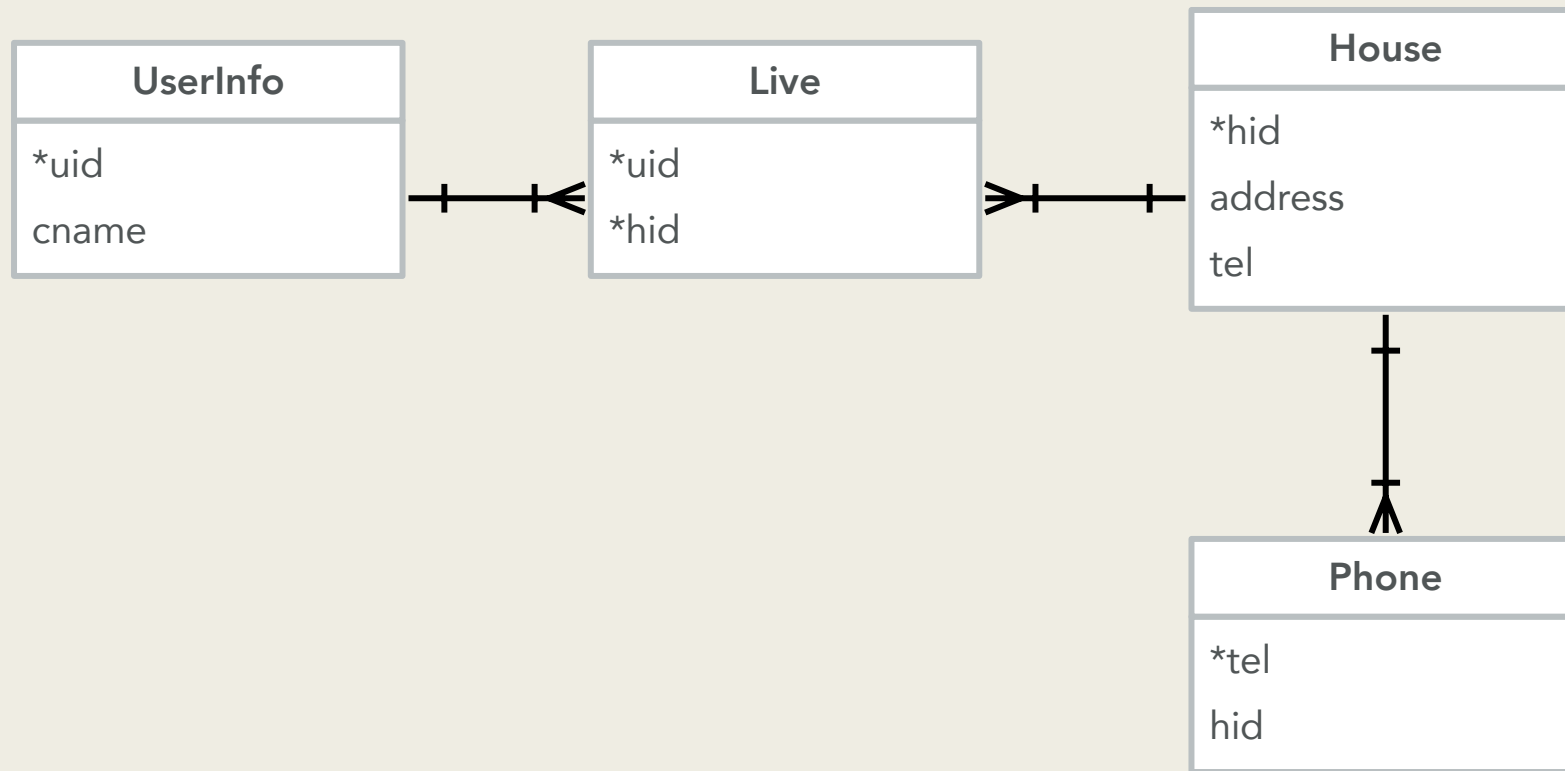
Entity-Relationship (ER) Diagrams

■ 目的：

- 此為資料庫的「架構圖」，通常在系統分析時製作，確定無誤後才開始實作資料庫







資料字典

UserInfo

序號	說明	欄位名稱	資料型態	NULL?
*1	身份證字號	uid	CHAR(10)	N
2	中文姓名	cname	NVARCHAR(50)	Y