Tamkang Universty Software Engineering Group
淡 江 軟 體 工 程 實 驗 室
http://www.tkse.tku.edu.tw/

Database

Presented by：Ying-Hong Wang
E-mail：inhon@mail.tku.edu.tw
2014/2/28

目次

- 基本規定
- 導論
- 資料模型
- 關聯式資料模型
- 運用E-R方法定義資料模型

基本規定

- 本課程分為講授、論文研讀與聆聽演講
- 每場聆聽演講需繳交聽講報告
  - 預計配合資工週演講活動進行
  - 每場演講結束後立即繳交隨堂報告，逾時不受理
  - 報告視同點名，擇優報告予以加分
  - 空白報告表請於授課網頁下載
- 論文研讀報告
  - 依座號依序輪流
  - 研讀論文指定請參閱授課網頁
- 學期評分
  - 出席15%，平時評量20%、期中考核15%、論文研讀報告50%
  - 每週統計出席，每人有三次免責，無其他請假理由，超過三次才開始扣出席分數(1%/per)，低於三次則加分(1%/per)

TSEG Tamkang Universty Software Engineering Group 淡江軟體工程實驗室 http://www.tkse.tku.edu.tw/

導論

- 相關名詞
  - 資料庫 (database, DB)
  - 集合的特性
  - 資料庫系統
  - 資料庫管理系統
    (database management systems, DBMS)

http://mail.tku.edu.tw/inhon

導論

- 資料庫 (DB)
  - 資料庫是相關資料的集合
- 集合的特性
  - 集合中的元素沒有順序性
  - 集合中的元素不會重複
  - 集合可以是空的

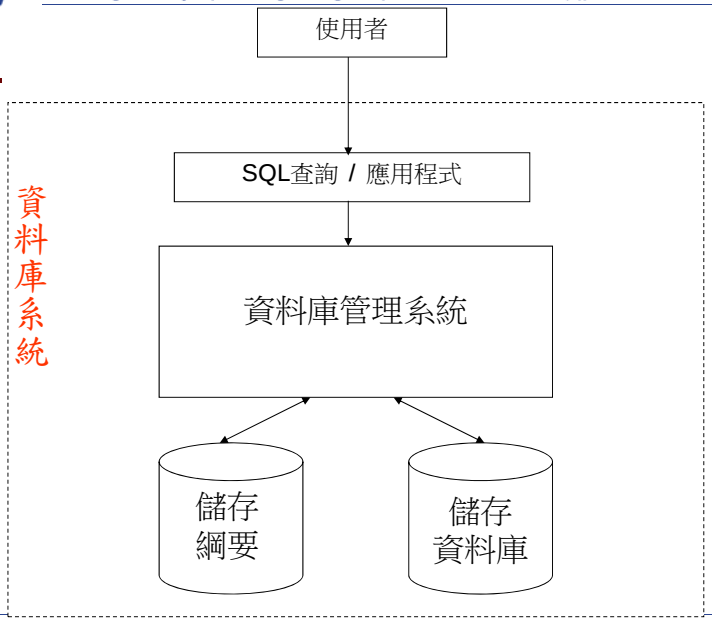
TSEG Tamkang Universty Software Engineering Group 淡江軟體工程實驗室 http://www.tkse.tku.edu.tw/

導論

- 資料庫管理系統
  - 資料庫管理系統 (DBMS)是許多程式的集合，讓使用者得以定義、建構，與處理資料庫
    - 定義:資料的型態、結構與限制
    - 建構:在DBMS的控制下,將資料儲存到媒體
    - 處理:查詢、更新

http://mail.tku.edu.tw/inhon

導論



Tamkang Universty Software Engineering Group
淡 江 軟 體 工 程 實 驗 室
http://www.tkse.tku.edu.tw/

Data Model

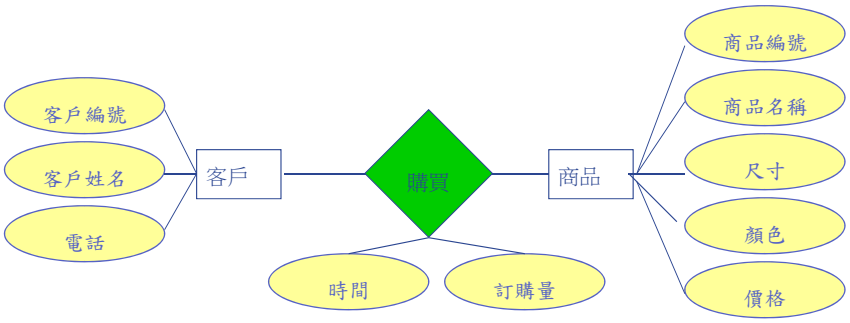
Presented by : Ying-Hong Wang
E-mail : inhon@mail.tku.edu.tw

資料模型

- 資料模型(Data Model) 是提供資料抽象化的主要工具
- Data Model 是用來描述資料庫結構的一種概念
- Data Model的內容包含
  - 結構的表示方式
  - 限制
  - 運算(操作)

資料模型 Data Model

- Data Model使用的概念
  - 實體 Entity
  - 屬性 Attribute
  - 關係 Relationship



Tamkang Universty Software Engineering Group
淡 江 軟 體 工 程 實 驗 室
http://www.tkse.tku.edu.tw/

Relational Data Model

Presented by : Ying-Hong Wang
E-mail : inhon@mail.tku.edu.tw

關聯式資料模型概念 - Relational Model Concepts

- 關聯式模型將資料庫表示成一個關聯的集合
- 相關名詞
  - 關聯(Relation)
  - 值組(Tuple)
  - 屬性(Attribute)
  - 定義域(Domain)

Relation Schemas – 名詞簡介

主鍵

關聯表名稱

PRODUCT

屬性集合

Pid	Pname	Size	Color	Price
11	襯衫	L	藍	100
22	帽子	M	藍	200
13	夾克	S	紅	300
34	牛仔褲	L	黑	400

值組

價格資料

品號屬性　品名屬性　尺寸屬性　顏色屬性　價格屬性

Relation Schemas – 特性與內容

- 關聯表的屬性數目稱為它的維度
- 一個關聯表中任兩個屬性名稱不可以重覆
- 「關聯表綱要」(Relation Schema) 包含:
  - 關聯表名稱
  - 屬性集宣告
  - 所有屬性的相對資料型態宣告
  - 主鍵的宣告
  - 外來鍵 (Foreign Key) 之宣告

DBMS的架構與資料獨立

- Three Schema (三綱要架構)
  - 外層(景觀層) External Level:外部綱要
  - 概念層 Conceptual Level:概念綱要
  - 內層 Internal Level:內部綱要
- 資料獨立
  - 邏輯資料獨立
  - 實體資料獨立

Three Schema

- 外層(又稱為景觀層) External Level:外部綱要
  - 描述特定使用者所感興趣的資料庫部分
  - 隱藏儲存結構等其他細節
- 概念層 Conceptual Level:概念綱要
  - 描述整個資料庫的結構
  - 隱藏儲存等細節
  - 著重在實體、資料型態、關係、使用者的操作與限制
- 內層Internal Level:內部綱要
  - 用來描述資料庫的實際儲存結構與路徑

Data Independence 資料獨立

- Logical data independence (邏輯資料獨立)
  - 改變 conceptual schema 時,不需改變 external schema (or application)
- Physical data independence (實體資料獨立)
  - 改變 internal schema 時,不需改變 conceptual schema

SQL

- DDL 資料定義語言
  - 用來定義 concept schema
  - 與資料庫綱要之定義有關的操作,如資料庫的建立、表格的建立等命令
- DML 資料處理語言
  - 對資料庫資料進行擷取、新增、刪除與修改等處理之命令
  - 僅對資料有所影響,與資料庫綱要定義無關

Tamkang Universty Software Engineering Group
淡 江 軟 體 工 程 實 驗 室
http://www.tkse.tku.edu.tw/

Modeling Using the Entity-Relationship Approach

Presented by：Ying-Hong Wang
E-mail：inhon@mail.tku.edu.tw

19

屬性型態

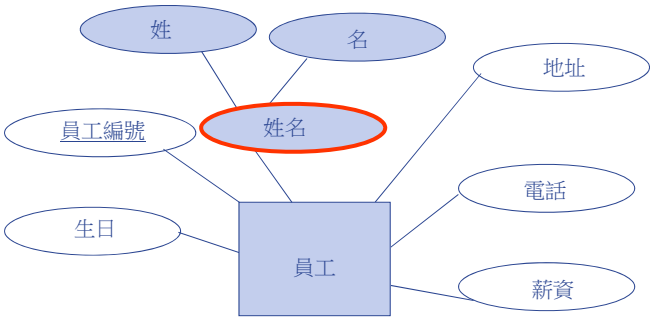
- 簡單(Simple)
- 複合(Composite)

- 單值(Single-valued)
- 多值(Multi-valued)

- 儲存(Stored)
- 導出(Derived)

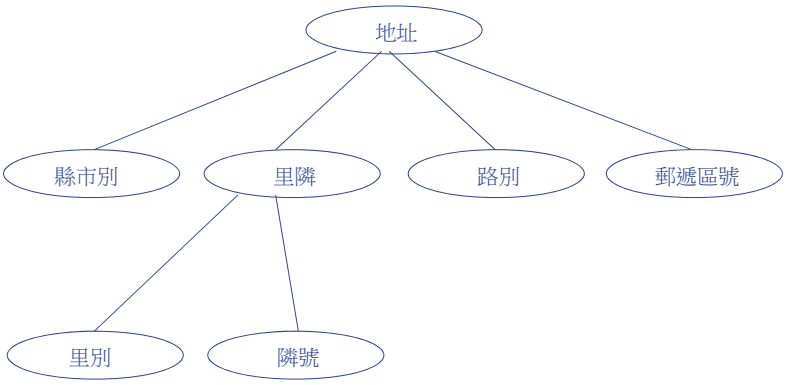
屬性型態

- 簡單(simple) 與 複合(composite)



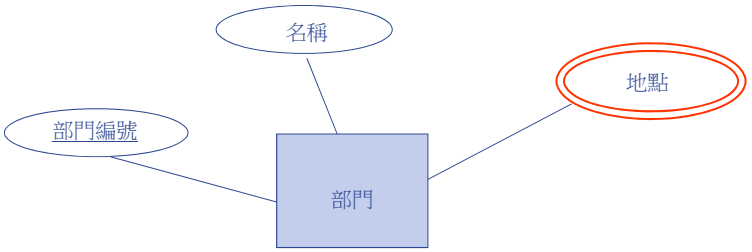
TSEG Tamkang Universty Software Engineering Group 淡江軟體工程實驗室 http://www.tkse.tku.edu.tw/

應該是簡單還是複合屬性?



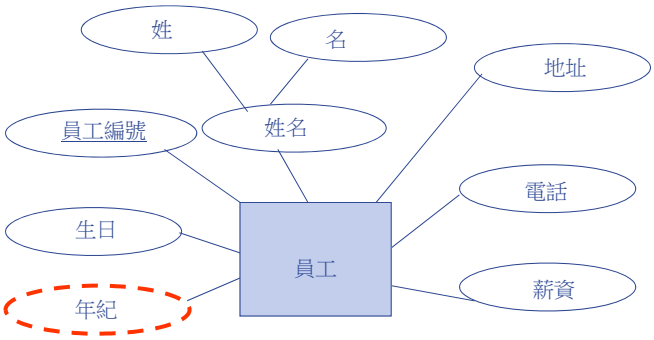
屬性型態

- 單值(single-valued) 與多值(multi-valued)
- Ex:一個部門只有一個部門編號與部門名稱,某些部門可能存同時在兩個地點(例如:一個部門分別在 *台中* 與 *彰化* 各有據點)



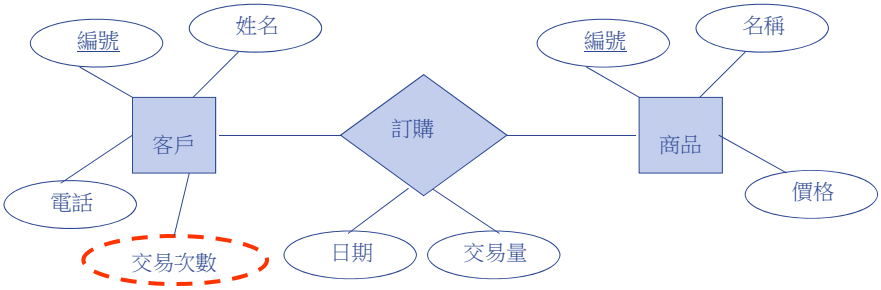
屬性型態

- 儲存(stored) 與 導出(derived)
- Ex:年紀可以經由**生日**換算得到



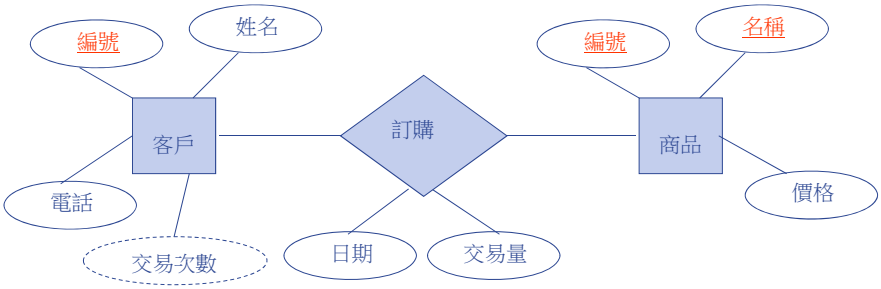
屬性型態 (範例二)

- Ex:一個客戶的交易次數可以經由訂購紀錄中計算出來



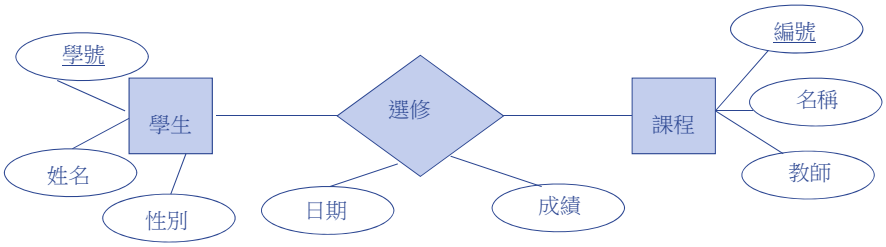
鍵值屬性

- 一個具有唯一性的屬性
- 假設:
  - 每一個客戶的編號會唯一
  - 每一個商品的編號會唯一,且每一商品的名稱也會唯一



實體,關係,屬性

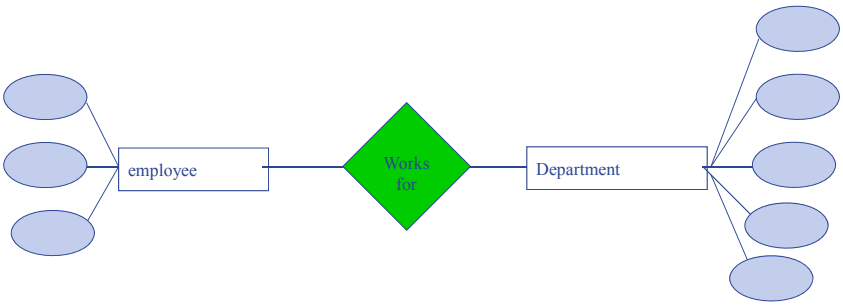
- 實體型態
  - 具有相同屬性集合的元素所形成的集合
- 關係型態
  - 用來連接某些相關的實體型態
- 屬性可以用來描述實體或關係



關係

Relationship

- 關係型態的Degree是參與此關係的實體型態數目
- Ex:Works_for的Degree為2,又稱二元關係

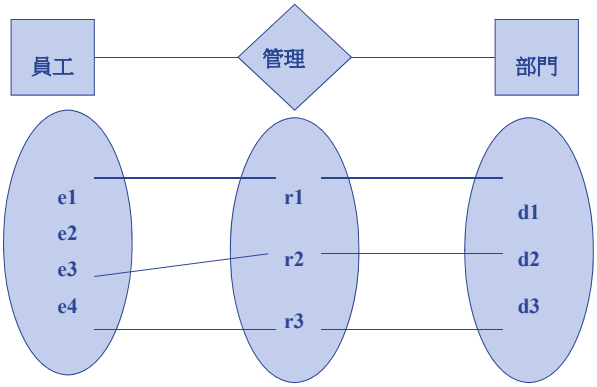


關係型態的限制

- Constraints on Relationship Types
  - 基數率與參與
- 基數率
  - 一個實體所能參與的關係案例數
  - Ex:一個部門僅會有一個主管
- 參與
  - Ex:任何一個部門一定有一個主管(去管理它)

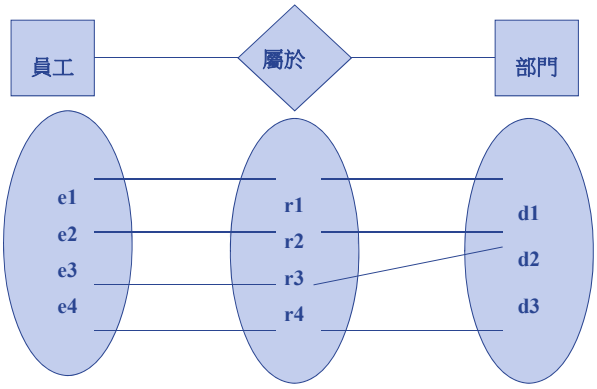
基數率(cardinality ratio)

- 1:1的範例



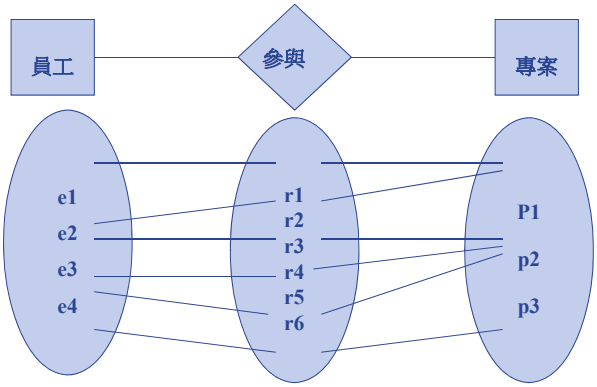
基數率

- 1:N的範例



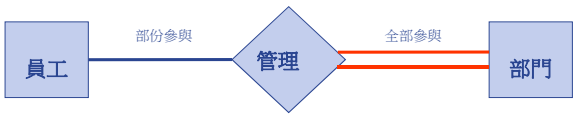
基數率

- N:M的範例



參與限制

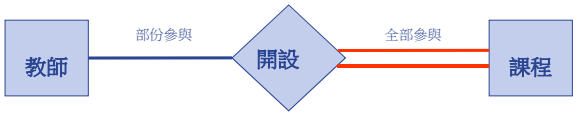
- 實體的存在是否依靠關係型態與另一實體產生關聯來決定
  - 部份參與 (以單線表示)
  - 全部參與 (以雙線表示)



TSEG Tamkang Universty Software Engineering Group 淡江軟體工程實驗室 http://www.tkse.tku.edu.tw/

參與限制 (範例二)

- 課程實體的存在是否依靠開設關係型態與另一教師實體產生關聯來決定
- 課程實體要存在必須所對應的教師實體與開設關係都存在,該課程才會存在
- 任何一個課程一定要有教師去開設它

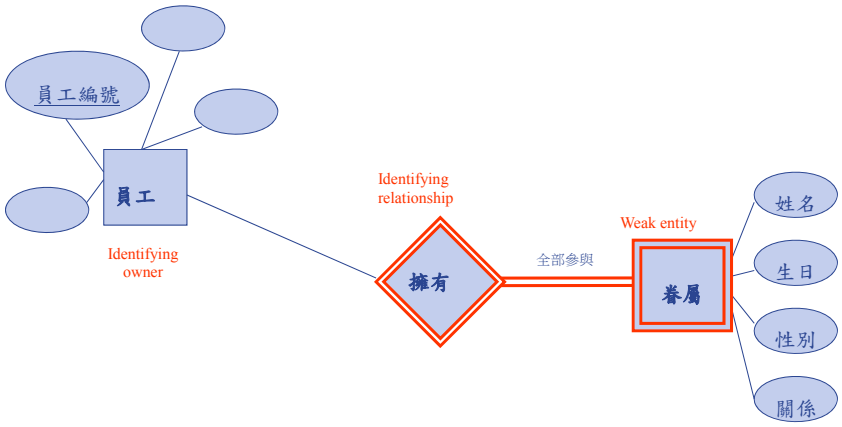


弱實體型態(Weak Entity Type)

- Weak Entity Types 本身可以沒有任何的鍵值屬性
- 透過另一實體型態的一些屬性所組成的相關特定實體(Identifying Owner)來辨認
- 辨認關係(Identifying Relationship)
- Weak Entity在沒有Identifying Owner 的狀況下是無法辨認的,所以Weak Entity 總是有一個完全參與的限制來與它的辨認關係產生關聯

TSEG Tamkang Universty Software Engineering Group 淡江軟體工程實驗室 http://www.tkse.tku.edu.tw/

弱實體型態(Weak Entity Type)



適當名稱

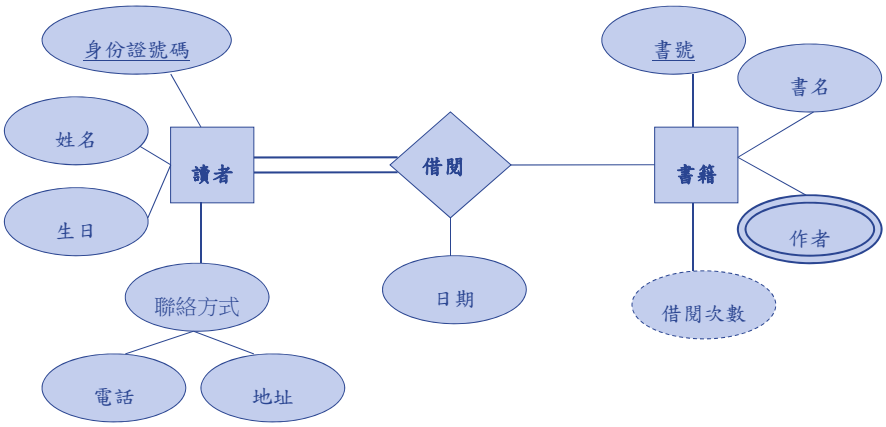
- 大寫字母表示實體與關係型態
- 屬性名稱字首大寫

- 實體多使用名詞
- 關係則採用動詞
- 方向：由左到右,由上到下

範例:讀者借閱書籍

- 系統必須紀錄讀者的身份證號碼、姓名、生日,與聯絡方式,其中聯絡方式包含了電話與地址資料
- 系統紀錄了館藏中每一本書籍的書名與作者,雖然有些書籍有兩個以上的作者,但一定都只有一個唯一書號. 為了掌握每本書的熱門程度,系統必須提供每本書籍的曾經被借閱的次數,以提供讀者參考
- 當一個讀者要借閱書籍時,系統必須紀錄借閱的日期,以便管理. 而每一位讀者至少都有借閱過一次以上書籍的紀錄

範例:讀者借閱書籍



TSEG

Tamkang University Software Engineering Group
淡 江 軟 體 工 程 實 驗 室
http://www.tkse.tku.edu.tw/

Relational Database Constraints

Presented by : Ying-Hong Wang
E-mail : inhon@mail.tku.edu.tw

關聯式資料模型的限制

- 定義域限制
- 鍵值限制
- 實體完整性限制
- 參考完整性限制

屬性與定義域(Domain)

- 欄位中所允許的數值(Value)資料稱為定義域(Domain)
- 定義域D是基元(Atomic)數值資料的集合
  - Domain中的數值是不可分割個所以稱為基元
  - Dom(Sex)={男生,女生} 則"男生" 與 " 女生" 皆為sex屬性的的合理基元值
- Ex:
  - age:大於0的整數
  - ID_numbers:一個英文字母加上九位數字的集合
  - 年級:介於1~4之間的數值

定義域限制

- 定義域限制指定了一屬性A的值必須在該屬性值的定義域Dom(A)裡的一個基元值
- 定義方式
  - 列舉
  - 範圍
  - 格式
  - 資料型態

鍵值限制

- 關聯可視為一個集合,所含值組必須相異
- $t1[SK] \neq t2[SK]$
  - SK: 超鍵 (superkey)
  - 每個關聯至少存在一個SK

包含SK的屬性集合亦為SK

Name	Sid	Tel_num	Sex	Age
張大森	s0023	047111111	男生	22
陳小東	s0052	047112235	女生	25
林中正	s0043	047223344	男生	23

- Ex: Sid是一SK,所以包含Sid的屬性集合皆為SK
  {Name,Sid} 、{Sid,Tel_num} 、{Sid,Name,Age}
- 鍵值屬性可以用來確認關聯中的唯一的值組
- 鍵值是由關聯綱要中屬性的意義來決定

主鍵(Primary Key)的產生

- 鍵值是一個最小的SuperKey
- 若有一個以上的鍵值,則每一個鍵值都稱為是一個候選鍵(Candidate Key)
- 必須從Candidate Key中選擇一個作為主鍵(Primary Key)
- 主鍵的特性
  - 唯一性
  - 最小性

關聯式資料模型的限制

- 定義域限制
- 鍵值限制
- 實體完整性限制
- 參考完整性限制

與關聯中的主鍵(PK)有關

鍵值限制與實體完整性

- 鍵值限制 (Key Constraint)
  - 主鍵值不能重複
- 實體完整性(Entity Integrity Constraint)
  - 主鍵值不能有空值

參考完整性

- 若一個關聯滿足定義域限制、鍵值限制、與實體完整性限制,我們只也能相信個別的關聯本身是合理的,但並無法保證關聯與關聯之間的資料狀態是一致的
- 參考完整性限制是指定在兩個關聯之間,用以維持兩個關聯值組的一致性

外鍵與參考

- R_1的外鍵FK (Foreign Key)
  - FK的屬性與另一關聯綱要R_2的Primary Key有相同的定義域 (FK 參考R_2)
  - R_1的值組t_1中的FK可以和R_2的某一值組t_2 的PK相符,或是空值
    - $t_1[FK] = t_2[PK]$

參考完整性

- FK的屬性值必須在所參考之關聯的PK屬性中被找到,或者為Null
- 關聯式資料庫之參考完整性限制,藉由FK與PK間的參考關係,以達到關聯表之間的一致性

EMPLOYEE.Dno 與 DEPARTMENT.Dnum 之間的一致性

EMPLOYEE

Eid	Name	Age	Sex	Dno
e1	張三	28	男生	1
e2	李四	33	女生	1
e3	王五	24	男生	3

DEPARTMENT

Dnum	Dname	Location
1	行政	台北
3	研發	新竹

若一員工不屬任何部門時，EMPLOYEE.Dno 可以是空值

EMPLOYEE

Eid	Name	Age	Sex	Dno
e1	張三	28	男生	1
e2	李四	33	女生	1
e3	王五	24	男生	3
e4	陳七	31	女生	

DEPARTMENT

Dnum	Dname	Location
1	行政	台北
3	研發	新竹

限制的實作

- 使用DDL指定限制

```
create table orders
   (no int not null,
    id int not null,
    quantity int,
    odate datetime default getdate(),  /*以系統日期及時間為odate的初值*/
    primary key (no,id,odate),         /*宣告複合欄位(no,id,odate)為主鍵*/
    foreign key (no) references bookstores,
    foreign key (id) references books,
    check(quantity>0 and quantity<5001)) /*設定quantity值的限制條件*/
```

Tamkang Universty Software Engineering Group
淡 江 軟 體 工 程 實 驗 室
http://www.tkse.tku.edu.tw/

Functional Dependencies and Normalization

Presented by : Ying-Hong Wang
E-mail : inhon@mail.tku.edu.tw

55

重要觀念

- 良好的語意
- 減少tuple的重複值
- 減少tuple的空值
- 不允許假值組(Spurious Tuples)的存在

語意設計上的優劣

- 設計易於解釋的Schema
  - 一個實體僅描述單一意義,勿結合多個於一關聯
- 不良的設計

EMP_DEPT

ENAME	EID	SEX	TEL	DNUM	DNAME	DMGR

TSEG Tamkang Universty Software Engineering Group 淡江軟體工程實驗室 http://www.tkse.tku.edu.tw/

tuple的重複值問題

EMP_DEPT

ENAME	EID	SEX	TEL	DNUM	DNAME	DMGR
陳花依	E1	女生	7111111	1	行政	E1
林大城	E2	男生	7111112	1	行政	E1
陳富強	E3	男生	7111113	2	研發	E3
張珠珠	E4	女生	7111114	2	研發	E3
莊勝為	E5	男生	7111115	3	生產	E5

減少tuple的空值

EID	ENAME	ADDRESS	TEL	PNO	PNAME
E1	陳花依	TAIPEI	123456	1	PROJ_1
E2	林大城			2	
E3	陳富強	HOUSTON	654321	3	PROJ_3
E2	林大城			3	PROJ_3

不允許假值組的存在

- 適當地分割關聯可以有效減少空間的浪費,並藉由正規化與關聯式運算(處理)來維持資料的完整性並滿足各項操作
- 假值組的問題,係因不良的關聯表格切割所導致
- 關聯表的分割必須考量FK與PK

EID	PNO	HOUR	ENAME	PNAME	PLOC
E1	1	32	陳花依	Proj1	TP
E1	2	7	陳花依	Proj2	CH
E2	3	40	林大城	Proj3	CH
E3	1	20	陳富強	Proj1	TP
E4	2	20	張珠珠	Proj3	KS

ENAME	PLOC
陳花依	TP
陳花依	CH
林大城	CH
陳富強	TP
張珠珠	KS

我們可以由
分割成,
但無法由
回覆成原來的

EID	PNO	HOUR	PNAME	PLOC
E1	1	32	Proj1	TP
E1	2	7	Proj2	CH
E2	3	40	Proj3	CH
E3	1	20	Proj1	TP
E4	2	20	Proj3	KS

Functional Dependencies

- 功能相依
  - 是資料庫的兩個屬性集合的限制
  - 關聯表分割的重要參考依據
- 表示法(X→Y)與意義
  - Relation R 中的任意兩個tuples t_1與t_2,若$t_1[X]=t_2[X]$,則使得$t_1[Y]=t_2[Y]$
  - Tuple中attribute Y的值是依靠X來決定
  - 稱為X到Y的功能相依,或Y功能相依於X
  - X可以決定Y

X→Y

- X→Y iff r(R)的X值相同,則Y值必定相同
- Key points
  - 若一關聯不存在超過一個以上的X值,則表示X是一個Candidate Key
  - 若 X→Y 並不表示 Y→X 成立
- 功能相依是一屬性語意的特性
- 功能相依無法從特定的關聯中推論得到

功能相依性 與 (非)主要屬性

EMP_PROJ

EID	PNO	HOUR	ENAME	PNAME	PLOC

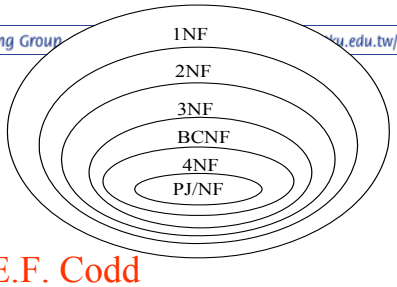
fd1

fd2

fd3

- 在Relation R 中,若一Attribute(s) 為任意R 的成員(每一Tuple都有該屬性),此Attribute(s)稱為主要屬性,否則則稱為非主要屬性
- EID與PNO為EMP_PROJ的主要屬性,其餘屬非主要屬性

正規化



- 第一正規化型式 (1NF) by E.F. Codd
- 第二正規化型式 (2NF) by E.F. Codd
- 第三正規化型式 (3NF) by E.F. Codd
- 廣義 3NF—BCNF (Boyce/Codd Normal Form)
- 第四正規化型式 (4NF) by R. Fagin
- 第五正規化型式 (5NF) by R. Fagin
  又稱為 Projection/Join Normal Form (PJ/NF)

1NF

- 1NF：所有屬性定義域僅含基元值(Atomic)
  (不能有複合屬性/多值的存在)

DEPARTMENT

DNUM	DNAME	DMGR	DLOCS
1	行政	E1	{TP, CH, TC}
2	研發	E3	{TP}
3	生產	E5	{KS}

不符合1NF

- DLOCS的定義域包含基元值,但某些值組包含這些值組的集合
- DLOCS屬性並非基元值

non-1NF分割 (Case1)

- 將違反1NF的屬性從原來的關聯中刪除，並與主鍵放到一個個別的關聯

DEPARTMENT

DNUM	DNAME	DMGR	~~DLOCS~~

DNUM	DLOC

DEPT_LOCATIONS

DNUM	DLOC
1	TP
1	CH
1	TC
2	TP
3	KS

1NF

- 1NF不允許有複合屬性的存在,這些屬性本身是屬於多值
- 巢狀關聯:一個值組有一個關聯存在
- PNO為巢狀關聯PROJS的部份主鍵

EMP_PORJ

EID	ENAME	PROJS	
		PNO	HOURS

EID	ENAME	PNO	HOURS
E1	陳花依	1	32
		2	7
E2	林大城	3	40
E3	陳富強	1	20
		2	12
E4	張珠珠	2	14
		3	14
		10	14
		20	14
E5	莊勝為	30	31
		20	10

non-1NF分割 (Case2)

EMP_PORJ

EID	ENAME	PNO	HOURS

1NF

EID	ENAME

EID	PNO	HOURS

主鍵與部份主鍵
形成新關聯的主鍵

TSEG Tamkang Universty Software Engineering Group 淡江軟體工程實驗室 http://www.tkse.tku.edu.tw/

2NF

- 2NF是以完全功能相依為基礎
- 完全功能相依
  - 若X→Y為完全功能相依,則從X中移除任一屬性A,將使得此功能相依性不再成立

EMP_PROJ

EID	PNO	HOUR	ENAME	PNAME	PLOC

fd1
fd2
fd3

- {EID,PNO} → HOUR (完全相依)
- EID → ENAME (去掉EID後不滿足相依性)
- PNO → {PNAME,PLOC} (去掉PNO後不滿足相依性)
- {EID,PNO} → ENAME (不滿足完全相依性,部份相依)

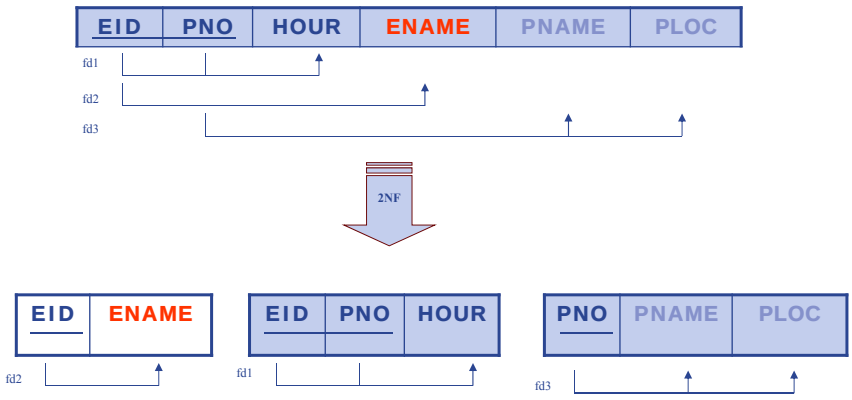
2NF

- 2NF
  - 若關聯R中的每一個非主要屬性是完全功能相依於R的Primary Key則稱Relation R屬於2NF
- non-2NF的處理
  - Non-2NF的Relation可以進一部被分解成多個2NF的Relations
  - 所產生的新關聯將只有非主要屬性與在主鍵內其完全功能相依的部份

TSEG Tamkang University Software Engineering Group 淡江軟體工程實驗室 http://www.tkse.tku.edu.tw/

2NF

EMP_PROJ



3NF 概念基礎

- 3NF是以遞移相依性為基礎
- 遞移相依性
  - Z不屬於任何鍵值的子集合，且X→Z,Z→Y成立

EMP_DEPT

ENAME	EID	SEX	TEL	DNUM	DNAME	DMGR

DNUM不屬於任何鍵值的子集合,且
EID→DNUM, DNUM→{DNAME,DMGR}成立
存在遞移相依 … (不符合3NF)

3NF 定義

- Relation R 是3NF:Relation R 是2NF且沒有非主要屬性是遞移相依於主鍵
- Non-3NF的Relation可以進一部被分解成多個3NF的Relations
- 所產生的新關聯將只有非主要屬性及與其完全功能相依的部分

3NF

EMP_DEPT

ENAME	EID	SEX	TEL	DNUM	DNAME	DMGR

3NF

員工

ENAME	EID	SEX	TEL	DNUM

部門

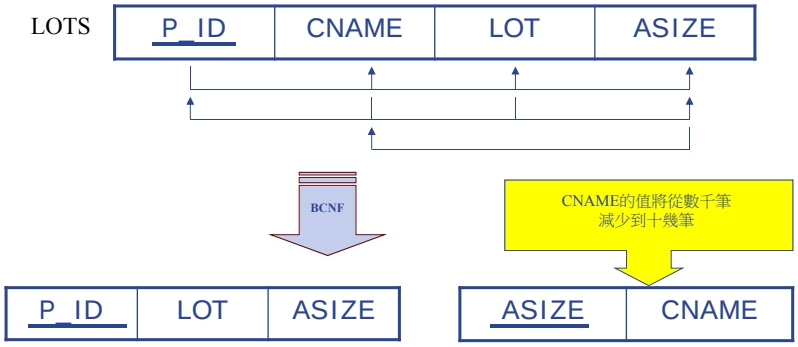
DNUM	DNAME	DMGR

Boyce-Codd正規式 (BCNF)

- BCNF比3NF更嚴格，BCNF成立3NF一定成立,但3NF不一定屬於BCNF
- BCNF的定義
  - 在Relation R中每一個功能相依X→A都滿足X是R的Super Key
- Example
  - CNAME={M, L}
  - ASIZE=0.5、0.6、0.7、0.8、0.9、1.0則CNAME=M
  - ASIZE=1.1、1.2、…、2.0則CNAME=L
  - 則存在一個FD: ASIZE → CNAME
  - 假設有數千筆Tuples

3NF → BCNF

- LOTS依舊符合3NF,但其實CNAME只有 M與L兩種值



Tamkang University Software Engineering Group
淡 江 軟 體 工 程 實 驗 室
http://www.tkse.tku.edu.tw/

SQL 語法

Presented by : Ying-Hong Wang
E-mail : inhon@mail.tku.edu.tw