

6. 正規劃

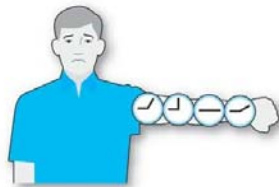
Normalization

為何「正規劃」很重要？

差在哪裡？



好的設計



不良的設計

當沒有一致更新冗餘資料的所有副本時，
便存在數據完整性的問題。

冗餘資料

- 重複的數據
- 可以從其它的數據推導出來的資料

如何判斷「冗餘資料」？

- 冗餘的真正意義不在於有多少副本
 - 即使是唯一的拷貝
- 而在於是否移除此一欄位之後，從此就遺失該資料
- 因此，如果您刪除某一欄位之後，該刪除內容還是可以通過其它欄位資料獲得的話
 - 則該欄位便是「多餘的」
 - 稱為導出屬性 (Derived Attribute)

存有冗餘資料的單一資料表

Students

std_no	name	sex	dept_sname	dept_fname	dept_phone_no
B0341001	周潔倫	男	醫管	醫務管理	5815
B0341002	蔡依林	女	醫管	醫務管理	5815
B0344103	張惠妹	女	資管	資訊管理	5812
B0344204	李宗盛	男	資管	資訊管理	5812
B0244201	史彥文	男	資管	資訊管理	5812
B0244103	張三豐	男	資管	資訊管理	5812
B0345001	高斯	男	工設	工業設計	5826

冗餘資料

重複的資料

- 可以從其它的資料推導出來的資料

為何冗餘資料是一個問題？

- 小問題
 - 浪費輸入資料的人力
 - 浪費資料庫的儲存空間
- 大問題
 - 重大的性能問題 (Major Performance Issue)
 - 如果冗餘資料必須被更新，它需要花更多的時間
 - 產生資料維護上的問題
 - 在新增、修改，或刪除資料時，極有可能會造成無法執行的錯誤或資料不一致的情況

異常 (Anomalies) 現象

異動資料的異常現象 (Anomalies)

- 指在結構不佳的資料表，進行資料新增、修改或刪除時，可能會造成不良副作用的情況
 - 產生資料不一致的情況
 - 無法執行
- 三種異常情況
 - 新增異常 (Insertion Anomaly)
 - 修改異常 (Update Anomaly)
 - 刪除異常 (Deletion Anomaly)

新增異常 (Insertion Anomaly)

- 新成立「大數據分析」科系
 - 副作用：不能單純的新增一個科系的資料，除非該科系已經有至少一個學生。

Students

std_no	name	sex	dept_sname	dept_fname	dept_phone_no
B0341001	周潔倫	男	醫管	醫務管理	5815
B0341002	蔡依林	女	醫管	醫務管理	5815
B0344103	張惠妹	女	資管	資訊管理	5812
B0344204	李宗盛	男	資管	資訊管理	5812
B0244201	史彥文	男	資管	資訊管理	5812
B0244103	張三豐	男	資管	資訊管理	5812
B0345001	高斯	男	工設	工業設計	5826

修改異常 (Update Anomaly)

- 更改資管的系辦電話分機號碼
 - 副作用：必須修改多個地方
 - 可能造成資料不一致

Students

std_no	name	sex	dept_sname	dept_fname	dept_phone_no
B0341001	周潔倫	男	醫管	醫務管理	5815
B0341002	蔡依林	女	醫管	醫務管理	5815
B0344103	張惠妹	女	資管	資訊管理	5812
B0344204	李宗盛	男	資管	資訊管理	5812
B0244201	史彥文	男	資管	資訊管理	5812
B0244103	張三豐	男	資管	資訊管理	5812
B0345001	高斯	男	工設	工業設計	5826

刪除異常 (Deletion Anomaly)

- 如果你刪除某個科系唯一一個學生的紀錄時，則資料庫再也沒有該科系的資料
 - 例如：如果高斯同學被退學，必須移除該筆紀錄...
 - 副作用：工設科系的資料也一併從資料庫中移除，資料庫中再也沒有工設系的相關資料

Students

std_no	name	sex	dept_sname	dept_fname	dept_phone_no
B0341001	周潔倫	男	醫管	醫務管理	5815
B0341002	蔡依林	女	醫管	醫務管理	5815
B0344103	張惠妹	女	資管	資訊管理	5812
B0344204	李宗盛	男	資管	資訊管理	5812
B0244201	史彥文	男	資管	資訊管理	5812
B0244103	張三豐	男	資管	資訊管理	5812
B0345001	高斯	男	工設	工業設計	5826

什麼是欄位之間的「函數依賴」？ (Functional Dependency)

函數依賴 (Functional Dependency)

- 當我們確定某個欄位 (X) 的值時，便能夠確定另一個欄位 (Y) 的值
 - Y 函數依賴 X
 - X 決定 Y: $X \rightarrow Y$
 - X: 決定者 (Determinant)
 - 和候選鍵概念相似
 - 如果2筆記錄的X欄位的值相同，則Y欄位的值也一定相同

15

函數依賴 (Functional Dependency)

- 描述資料表中欄位間的依賴關係
- 給定一個 std_no 的值，就能夠確定 name 的值
 - name 欄位 “函數依賴” 在 std_no 欄位
 - std_no 欄位 決定 name 欄位
 - $\text{std_no} \rightarrow \text{name}$

Students

std_no	name	sex	dept_no	sname	fname	phone_no
B0341001	周潔倫	男	3100	醫管	醫務管理	5815
B0341002	蔡依林	女	3100	醫管	醫務管理	5815
B0344103	張惠妹	女	3500	資管	資訊管理	5812
B0344204	李宗盛	男	3500	資管	資訊管理	5812
B0244201	史彥文	男	3500	資管	資訊管理	5812
B0244103	張三豐	男	3500	資管	資訊管理	5812
B0345001	高斯	男	3600	工設	工業設計	5826

16

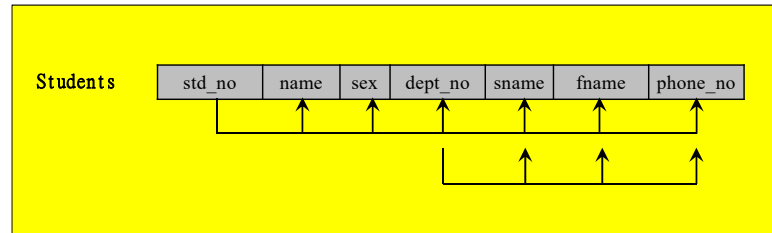
函數依賴 (Functional Dependency)

- 描述資料表中欄位間的依賴關係
- 給定一個 std_no 的值，就能夠確定 name 的值
 - name 欄位 “函數依賴” 在 std_no 欄位
 - std_no 欄位 決定 name 欄位
 - $\text{std_no} \rightarrow \text{name}$

Students

std_no	name	sex	dept_no	sname	fname	phone_no
B0341001	周潔倫	男	3100	醫管	醫務管理	5815
B0341002	蔡依林	女	3100	醫管	醫務管理	5815
B0344103	張惠妹	女	3500	資管	資訊管理	5812
B0344204	李宗盛	男	3500	資管	資訊管理	5812
B0244201	史彥文	男	3500	資管	資訊管理	5812
B0244103	張三豐	男	3500	資管	資訊管理	5812
B0345001	高斯	男	3600	工設	工業設計	5826

函數依賴 (Functional Dependency)



使用現成資料來驗證

Students

std_no	name	sex	dept_no	sname	fname	phone_no
B0341001	周潔倫	男	3100	醫管	醫務管理	5815
B0341002	蔡依林	女	3100	醫管	醫務管理	5815
B0344103	張惠妹	女	3500	資管	資訊管理	5812
B0344204	李宗盛	男	3500	資管	資訊管理	5812
B0244201	史彥文	男	3500	資管	資訊管理	5812
B0244103	張三豐	男	3500	資管	資訊管理	5812
B0345001	高斯	男	3600	工設	工業設計	5826

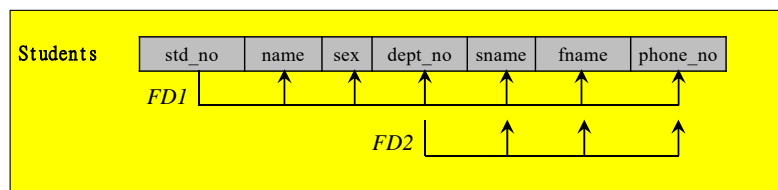
$std_no \rightarrow dept_no$ (?)
 $dept_no \rightarrow std_no$ (?)

如何呈現欄位間的函數依賴關係

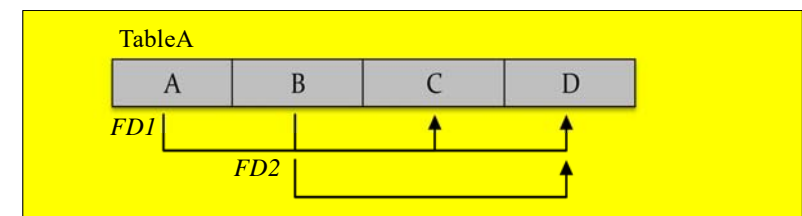
1. 函數依賴列表

(FD1) $std_no \rightarrow name, sex, dept_no, sname, fname, phone_no$
 (FD2) $dept_no \rightarrow sname, fname, phone_no$

2. 函數依賴關係圖 (Functional Dependency Diagram)

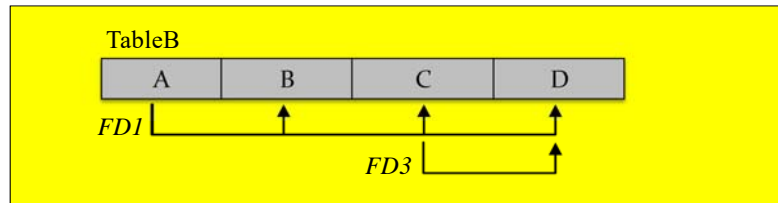


部分依賴 (Partial Dependency)



D 部分依賴在(A, B)上

遞移依賴(Transitive Dependency)



D 遞移依賴 在 A 上

什麼是「正規化」？

什麼是「正規化」？

- 「正規化」(Normalization)
 - 一套方法來組織直接相關的欄位成為資料表
 - 在組織的過程中，資料的冗餘情況被一一消除
 - 使的異常現象不再發生
- 正規化的目的：
 - 去除資料冗餘的現象
 - 力求一個事實的資料只出現在一個地方

如何進行「正規化」？
(Normalization)

正規化形式 (Normal Form)

- 一個規則
 - 規定一個資料表欄位之間不該存在的函數依賴種類
- 每個正規化的形式會移除某類函數依賴
 - 會移除某種的冗餘情況

正規化形式種類

正規化形式	特徵
第一正規化形式 (1NF)	不會有重複群組或多值欄位
第二正規化形式 (2NF)	1NF 且移除 <u>部分依賴</u>
第三正規化形式 (3NF)	2NF 且移除 <u>遞移依賴</u>

第一正規化形式 (1NF)

第一正規化形式 (1NF)

- 符合 1NF 的資料表
 - 每個儲存格最多只可以包含一個值，稱為單元值
 - 或沒有重複的組 (No Repeating Groups)
- 一個關聯資料庫的資料表，至少要符合第一正規化形式的規定

違反第一正規化形式:有重複群組

Enrollment

StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	CourseNo	CrsDesc	EnrGrade
S1	Seattle	JUN	O1	Fall	2006	C1	DB	3.5
			O2	Spring	2006	C2	VB	3.3
S2	Berkeley	JUN	O4	Fall	2007	C1	DB	3.1
			O2	Spring	2006	C2	VB	3.4
S3	New York	SEN	O3	Fall	2006	C1	DB	3.5
			O5	Spring	2007	C3	Java	3.6

如何轉成第一正規化形式

Enrollment

StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	CourseNo	CrsDesc	EnrGrade
S1	Seattle	JUN	O1	Fall	2006	C1	DB	3.5
			O2	Spring	2006	C2	VB	3.3
S2	Berkeley	JUN	O4	Fall	2007	C1	DB	3.1
			O2	Spring	2006	C2	VB	3.4
S3	New York	SEN	O3	Fall	2006	C1	DB	3.5
			O5	Spring	2007	C3	Java	3.6



Enrollment

StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	CourseNo	CrsDesc	EnrGrade
S1	Seattle	JUN	O1	Fall	2006	C1	DB	3.5
S1	Seattle	JUN	O2	Spring	2006	C2	VB	3.3
S2	Berkeley	JUN	O4	Fall	2007	C1	DB	3.1
S2	Berkeley	JUN	O2	Spring	2006	C2	VB	3.4
S3	New York	SEN	O3	Fall	2006	C1	DB	3.5
S3	New York	SEN	O5	Spring	2007	C3	Java	3.6

符合第一正規化形式資料表結構

```
Enrollment(StdSSN, StdCity, StdClass, OfferNo, OffTerm, OffYear,
           CourseNo, CrsDesc, EnrGrade);
```

找出主鍵欄位

```
Enrollment(StdSSN, StdCity, StdClass, OfferNo, OffTerm, OffYear,
           CourseNo, CrsDesc, EnrGrade);
```

找出欄位之間的函數依賴關係

Enrollment

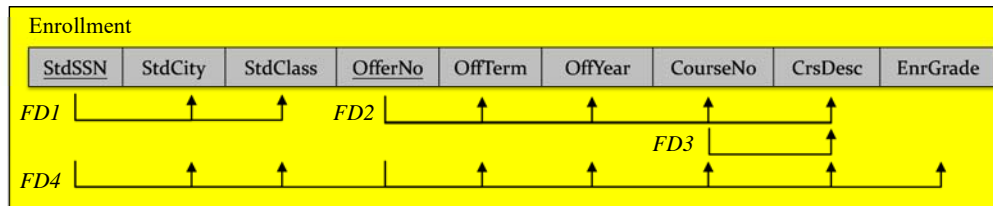
StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	CourseNo	CrsDesc	EnrGrade
S1	Seattle	JUN	O1	Fall	2006	C1	DB	3.5
S1	Seattle	JUN	O2	Spring	2006	C2	VB	3.3
S2	Berkeley	JUN	O4	Fall	2007	C1	DB	3.1
S2	Berkeley	JUN	O2	Spring	2006	C2	VB	3.4
S3	New York	SEN	O3	Fall	2006	C1	DB	3.5
S3	New York	SEN	O5	Spring	2007	C3	Java	3.6

欄位之間的函數依賴列表

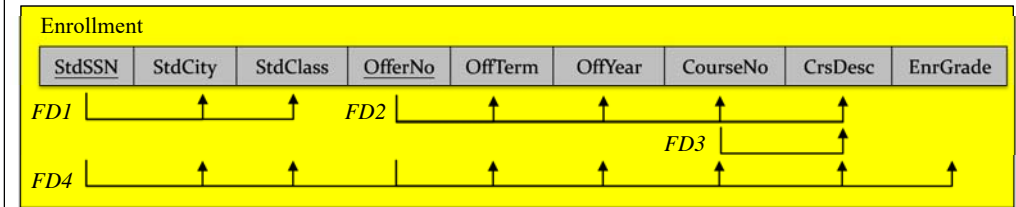
```
(FD1) StdSSN → StdCity, StdClass
(FD2) OfferNo → OffTerm, OffYear, CourseNo, CrsDesc
(FD3) CourseNo → CrsDesc
(FD4) StdSSN, OfferNo → StdCity, StdClass, OffTerm, OffYear,
                        CourseNo, CrsDesc, EnrGrade
```


將函數依賴列表畫成函數依賴關係圖

(FD1) StdSSN \rightarrow StdCity, StdClass
 (FD2) OfferNo \rightarrow OffTerm, OffYear, CourseNo, CrsDesc
 (FD3) CourseNo \rightarrow CrsDesc
 (FD4) StdSSN, OfferNo \rightarrow StdCity, StdClass, OffTerm, OffYear, CourseNo, CrsDesc, EnrGrade



分析函數依賴關係



主鍵欄位：

部分依賴之欄位：

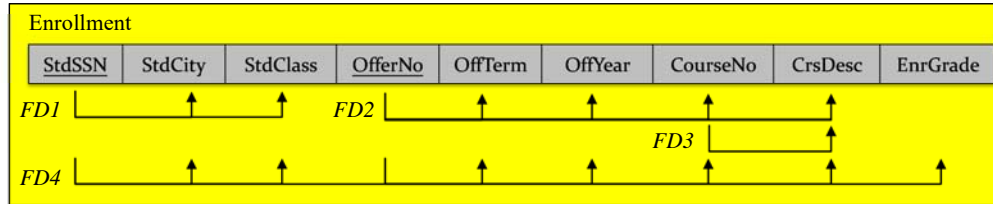
遞移依賴欄位：

第二正規化形式 (2NF)

2NF

- 符合 2NF 的資料表
 - 符合第一正規化形式規定，而且
 - 沒有部分依賴的存在
- 違反第二正規化形式
 - 只會發生在候選鍵是組合鍵的資料表中
- 任何符合第一正規化形式之資料表，如果它的候選鍵只包含單一欄位，此資料表一定符合第二正規化形式

有哪些部分依賴？



主鍵欄位：

部分依賴之欄位：

「部分依賴」造成的資料冗餘

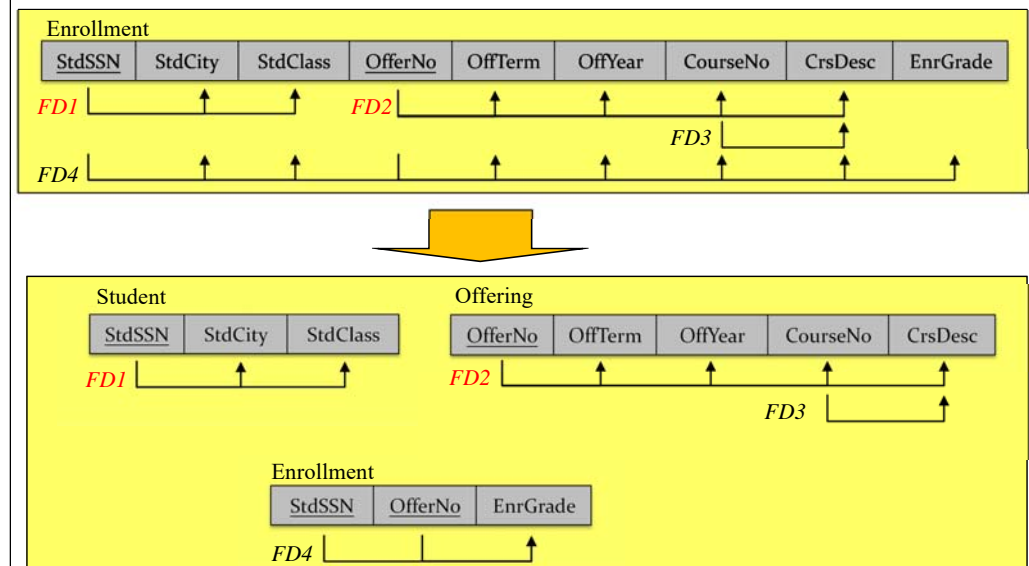
Enrollment

StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	CourseNo	CrsDesc	EnrGrade
S1	Seattle	JUN	O1	Fall	2006	C1	DB	3.5
S1	Seattle	JUN	O2	Spring	2006	C2	VB	3.3
S2	Berkeley	JUN	O4	Fall	2007	C1	DB	3.1
S2	Berkeley	JUN	O2	Spring	2006	C2	VB	3.4
S3	New York	SEN	O3	Fall	2006	C1	DB	3.5
S3	New York	SEN	O5	Spring	2007	C3	Java	3.6

如何轉換成 2NF?

- 將原資料表分裂成滿足2NF定義的數個小資料表
- 拆分後，你應該添加參照完整性(外鍵)來連接這些資料表。

如何轉換成 2NF?



轉換成 2NF 之後的資料表結構

1NF

```
Enrollment (StdSSN, StdCity, StdClass,
            OfferNo, OffTerm, OffYear, CourseNo, CrsDesc, EnrGrade)
```

2NF

```
Student (StdSSN, StdCity, StdClass)
```

```
Offering (OfferNo, OffTerm, OffYear, CourseNo, CrsDesc)
```

```
Enrollment (StdSSN, OfferNo, EnrGrade)
    FOREIGN KEY (StdSSN) REFERENCES Student (StdSSN)
    FOREIGN KEY (OfferNo) REFERENCES Offering (OfferNo)
```

是否有改善「資料冗餘」的情況？

Enrollment

StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	CourseNo	CrsDesc	EnrGrade
S1	Seattle	JUN	O1	Fall	2006	C1	DB	3.5
S1	Seattle	JUN	O2	Spring	2006	C2	VB	3.3
S2	Berkeley	JUN	O4	Fall	2007	C1	DB	3.1
S2	Berkeley	JUN	O2	Spring	2006	C2	VB	3.4
S3	New York	SEN	O3	Fall	2006	C1	DB	3.5
S3	New York	SEN	O5	Spring	2007	C3	Java	3.6

1NF

2NF

Student

StdSSN	StdCity	StdClass
S1	Seattle	JUN
S2	Berkeley	JUN
S3	New York	SEN

Enrollment

StdSSN	OfferNo	EnrGrade
S1	O1	3.5
S1	O2	3.3
S2	O4	3.1
S2	O2	3.4
S3	O3	3.5
S3	O5	3.6

Offering

OfferNo	OffTerm	OffYear	CourseNo	CrsDesc
O1	Fall	2006	C1	DB
O2	Fall	2006	C2	VB
O3	Spring	2007	C1	DB
O4	Fall	2007	C1	DB
O5	Spring	2007	C3	Java

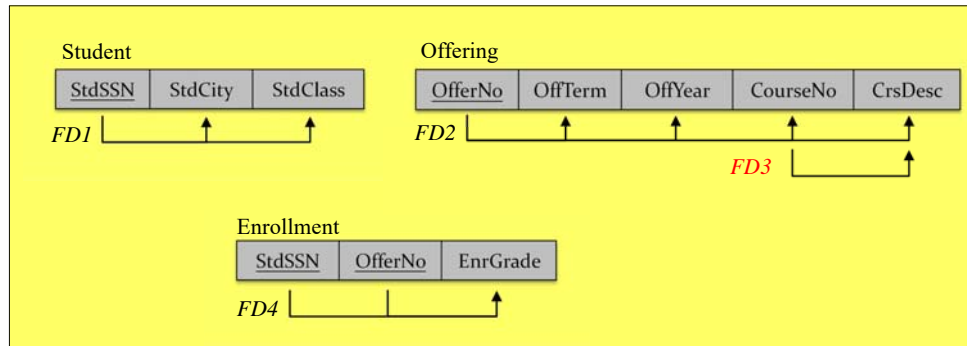
第三正規化形式 (3NF)

3NF

■ 符合 3NF 的資料表

- 符合第二正規化形式規定（無部分依賴），而且
- 沒有遞移依賴的存在

哪些資料表有遞移依賴?



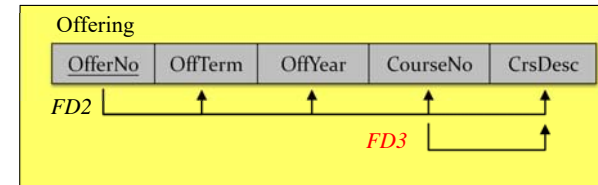
哪些是遞移依賴?

遞移依賴欄位:

遞移依賴造成的冗餘資料

Offering

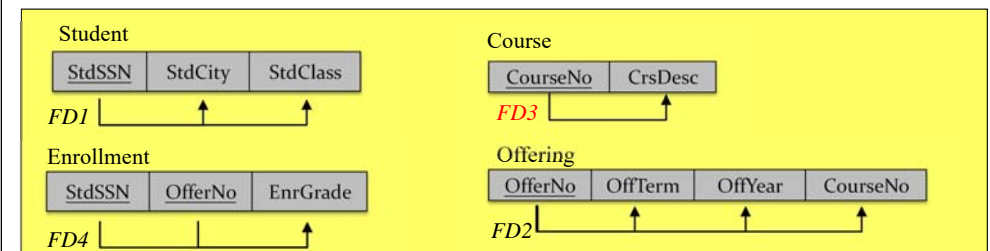
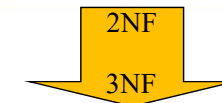
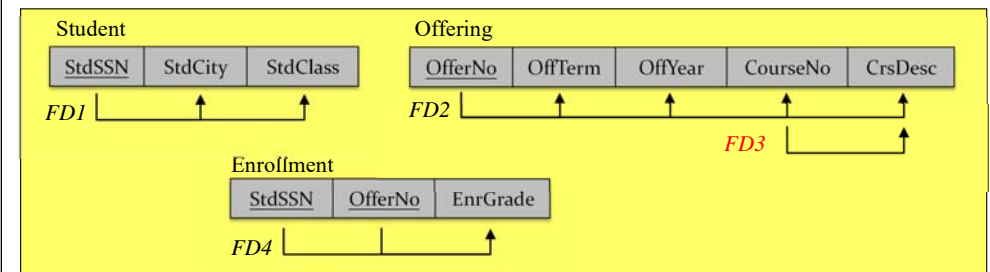
OfferNo	OffTerm	OffYear	CourseNo	CrsDesc
O ₁	Fall	2006	C ₁	DB
O ₂	Fall	2006	C ₂	VB
O ₃	Spring	2007	C ₁	DB
O ₄	Fall	2007	C ₁	DB
O ₅	Spring	2007	C ₃	Java



如何轉換成 3NF?

- 將原資料表分裂成滿足3NF定義的數個小資料表
- 在每一個較小的資料表，主鍵直接決定所有非關鍵欄位
- 拆分後，你應該添加參照完整性 (外鍵)來連接這些資料表。

如何轉換成 3NF?



轉換成 3NF

2NF

```

Student (StdSSN, StdCity, StdClass)
Offering (OfferNo, OffTerm, OffYear, CourseNo, CrsDesc)
Enrollment (StdSSN, OfferNo, EnrGrade)
    FOREIGN KEY (StdSSN) REFERENCES Student (StdSSN)
    FOREIGN KEY (OfferNo) REFERENCES Offering (OfferNo)
  
```

3NF

```

Student (StdSSN, StdCity, StdClass)
Course (CourseNo, CrsDesc)
Offering (OfferNo, OffTerm, OffYear, CourseNo)
    FOREIGN KEY (CourseNo) REFERENCES Course (CourseNo)
Enrollment (StdSSN, OfferNo, EnrGrade)
    FOREIGN KEY (StdSSN) REFERENCES Student (StdSSN)
    FOREIGN KEY (OfferNo) REFERENCES Offering (OfferNo)
  
```

是否有改善「資料冗餘」的情況？

Offering

OfferNo	OffTerm	OffYear	CourseNo	CrsDesc
O1	Fall	2006	C1	DB
O2	Fall	2006	C2	VB
O3	Spring	2007	C1	DB
O4	Fall	2007	C1	DB
O5	Spring	2007	C3	Java

2NF

3NF

Course

CourseNo	CrsDesc
C1	DB
C2	VB
C3	Java

Offering

OfferNo	OffTerm	OffYear	CourseNo
O1	Fall	2006	C1
O2	Fall	2006	C2
O3	Spring	2007	C1
O4	Fall	2007	C1
O5	Spring	2007	C3

正規化

- 第一階正規化型式 (1NF)
 - 在資料表刪除多重值屬性，讓資料表只擁有單元值屬性
- 第二階正規化型式 (2NF)
 - 滿足1NF，且資料表沒有部分依賴
- 第三階正規化型式 (3NF)
 - 滿足2NF，且資料表沒有遞移依賴

Enrollment

StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	CourseNo	CrsDesc	EnrGrade
FD1								
			FD2					
						FD3		
FD4								

Student

StdSSN	StdCity	StdClass
--------	---------	----------

FD1		
-----	--	--

StdSSN	OfferNo	EnrGrade
--------	---------	----------

FD4		
-----	--	--

Offering

OfferNo	OffTerm	OffYear	CourseNo	CrsDesc
---------	---------	---------	----------	---------

FD2				
-----	--	--	--	--

			FD3	
--	--	--	-----	--

Student

StdSSN	StdCity	StdClass
--------	---------	----------

FD1		
-----	--	--

StdSSN	OfferNo	EnrGrade
--------	---------	----------

FD4		
-----	--	--

Offering

OfferNo	OffTerm	OffYear	CourseNo
---------	---------	---------	----------

FD2			
-----	--	--	--

CourseNo	CrsDesc
----------	---------

	FD3
--	-----

每個正規化步驟的結果

1NF

Enrollment(StdSSN, StdCity, StdClass,
OfferNo, OffTerm, OffYear, CourseNo, CrsDesc, EnrGrade)



2NF

Student (StdSSN, StdCity, StdClass)
Offering(OfferNo, OffTerm, OffYear, CourseNo, CrsDesc)
Enrollment(StdSSN, OfferNo, EnrGrade)
FOREIGN KEY (StdSSN) REFERENCES Student(StdSSN)
FOREIGN KEY (OfferNo) REFERENCES Offering(OfferNo)



3NF

Student (StdSSN, StdCity, StdClass)
Course (CourseNo, CrsDesc)
Offering(OfferNo, OffTerm, OffYear, CourseNo)
Enrollment(StdSSN, OfferNo, EnrGrade)
FOREIGN KEY (CourseNo) REFERENCES Course(CourseNo)
FOREIGN KEY (StdSSN) REFERENCES Student(StdSSN)
FOREIGN KEY (OfferNo) REFERENCES Offering(OfferNo)

「資料冗餘」因正規化而獲得改善

Enrollment

StdSSN	StdCity	StdClass	OfferNo	OffTerm	OffYear	CourseNo	CrsDesc	EnrGrade
S1	Seattle	JUN	O1	Fall	2006	C1	DB	3.5
S1	Seattle	JUN	O2	Spring	2006	C2	VB	3.3
S2	Berkeley	JUN	O4	Fall	2007	C1	DB	3.1
S2	Berkeley	JUN	O2	Spring	2006	C2	VB	3.4
S3	New York	SEN	O3	Fall	2006	C1	DB	3.5
S3	New York	SEN	O5	Spring	2007	C3	Java	3.6



Student

StdSSN	StdCity	StdClass
S1	Seattle	JUN
S2	Berkeley	JUN
S3	New York	SEN

Course

CourseNo	CrsDesc
C1	DB
C2	VB
C3	Java

Offering

OfferNo	OffTerm	OffYear	CourseNo
O1	Fall	2006	C1
O2	Fall	2006	C2
O3	Spring	2007	C1
O4	Fall	2007	C1
O5	Spring	2007	C3

Enrollment

StdSSN	OfferNo	EnrGrade
S1	O1	3.5
S1	O2	3.3
S2	O4	3.1
S2	O2	3.4
S3	O3	3.5
S3	O5	3.6

作業

HW 第七章 239頁 Questions: 1, 2, 3, 14, 15, 24

HW 第七章 242頁 13, 14, 15