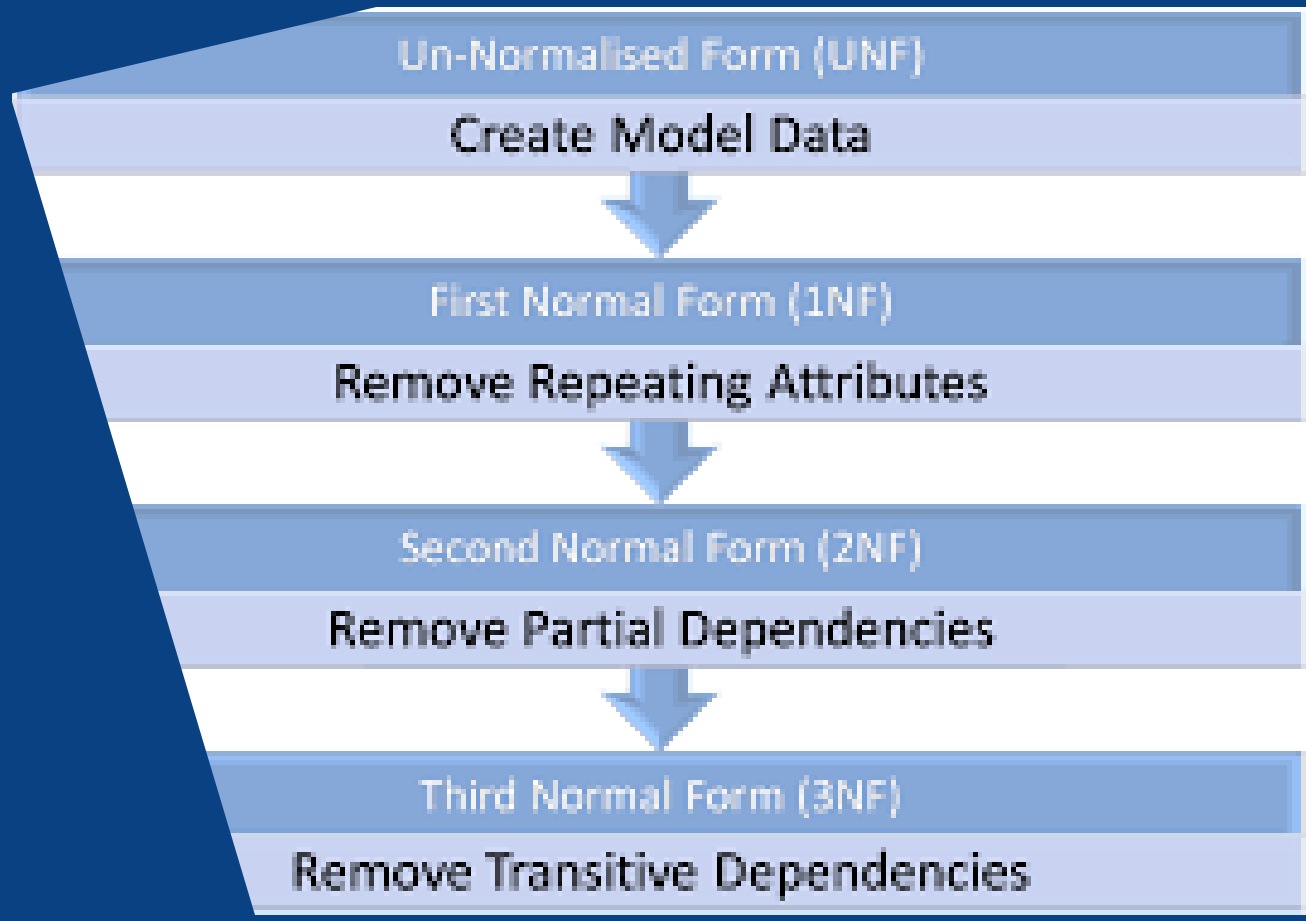
THE UNIVERSITY OF MELBOURNE

Normalisation

Database Systems & Information Modelling
INFO90002



Week 3 – Three normal forms

Dr Tanya Linden
Dr Renata Borovica-Gajic
David Eccles

Learning Objectives

By the end of this lecture, you should be able to:

- Define normalisation
- Explain and identify database anomalies
- Define and identify functional dependencies
- Normalise relations to:
  - 1st Normal Form (1NF)
  - 2nd Normal Form (2NF)
  - 3rd Normal Form (3NF)
  - Boyce-Codd Normal Form (BCNF)

Motivation for normalisation

What happens if we don’t normalise?

What's wrong with the *organisation* of data in this table?

Student ID#	Student Name	Campus Address	Degree	Phone	Subject ID	Subject Title	Lecturer Name	Lecturer Office	Lecturer Phone	Sem.	Grade
A121	Joy Egbert	166 Grattan Street	B.Com.	555-7771	ACC101	Accounting	Davern	T240C	8344-1846	1-11	H1
A121	Joy Egbert	166 Grattan Street	B.Com.	555-7771	ECO101	Economics	Smyth	T240F	8344-1868	1-11	H2B
A121	Joy Egbert	166 Grattan Street	B.Com.	555-7771	ECO104	Quant. M.	Collier	T240D	8344-5716	1-11	H2B
A121	Joy Egbert	166 Grattan Street	B.Com.	555-7771	FIN101	Finance.	James	T240D	8344-5275	1-11	H2A
A121	Joy Egbert	166 Grattan Street	B.Com.	555-7771	ACC103	Processes	Wise	T240E	8344-5309	1-11	H3
A123	Larry Mueller	302 Royal Parade	B.Com.	555-1235	ACC101	Accounting	Davern	T240C	8344-1846	1-11	H1
A123	Larry Mueller	302 Royal Parade	B.Com.	555-1235	ECO101	Economics	Smyth	T240F	8344-1868	1-11	H2B
A123	Larry Mueller	302 Royal Parade	B.Com.	555-1235	ECO104	Quant. M.	Collier	T240D	8344-5716	1-11	H2A
A123	Larry Mueller	302 Royal Parade	B.Com.	555-1235	FIN101	Finance.	James	T240D	8344-5275	1-11	H3
A124	Mike Guon	224 Swanston St.	B.Eco.	555-2214	ACC101	Accounting	Davern	T240C	8344-1846	1-11	H2A
A124	Mike Guon	224 Swanston St.	B.Eco.	555-2214	ECO101	Economics	Smyth	T240F	8344-1868	1-11	H2A
A124	Mike Guon	224 Swanston St.	B.Eco.	555-2214	ECO104	Quant. M.	Collier	T240D	8344-5716	1-11	H2B
A124	Mike Guon	224 Swanston St.	B.Eco.	555-2214	ACC103	Processes	Wise	T240E	8344-5309	1-11	H2B
A126	Jackie Judson	85 Barry Street	B.Eco.	555-1245	ACC101	Accounting	Davern	T240C	8344-1846	1-11	H1
A126	Jackie Judson	85 Barry Street	B.Eco.	555-1245	ECO101	Economics	Smyth	T240F	8344-1868	1-11	H2B
A126	Jackie Judson	85 Barry Street	B.Eco.	555-1245	ECO104	Quant. M.	Collier	T240D	8344-5716	1-11	H2B
A126	Jackie Judson	85 Barry Street	B.Eco.	555-1245	ACC103	Processes	Wise	T240E	8344-5309	1-11	H2A
...	...	...	...	...	...	...	...	...	...	...	...

Anomalies in Denormalised Data

Consider the following denormalised table (relation) :

Student-ID	Course-ID	Fee
130	C200	75
200	C300	100
250	C200	75
425	C400	150
500	C300	100
575	C500	50
…	…	…

Insertion Anomaly: A new course cannot be added until at least one student has enrolled (which comes first student or course?)

Deletion Anomaly: If student 425 withdraws, we lose all record of course C400 and its fee!

Update Anomaly: If the fee for course C200 changes, we have to change it in multiple records (rows), else the data will be inconsistent.

Normalisation

A technique used to remove undesired redundancy from databases (Break one large table into several smaller tables).

A relation is normalised if all determinants are candidate keys

A determinant in a database table is **any attribute that you can use to determine the values assigned to other attribute(s) in the same row**.

How do we normalise?

Invoice example

Bill To

John

Synex Inc
128 AA Juanita Ave
Glendora
CA 91740 US

Ship To

John

Synex Inc
128 AA Juanita Ave
Glendora
CA 91740 US

Date	14-Aug-2009	Order No		Sales Person	Charles Wooten
Shipping Date	13-Aug-2009	Shipping Terms		Terms	COD

ID	SKU / Description	Unit Price (USD)	Qty	Amount (USD)
PS.V860.005	AMD Athlon X2DC-7450, 2.4GHz/1GB/160GB/SMP-DVD/VB	580.00	6.00	3,480.00
PS.V880.037	PDC-E5300 - 2.6GHz/1GB/320GB/SMP-DVD/FDD/VB	645.00	4.00	2,580.00
LC.V890.002	LG 18.5" WLCD	230.00	10.00	2,300.00
HP.Q754.071	HP LaserJet 5200	1,103.00	1.00	1,103.00

Invoice example – Spreadsheet

This is not a relational model

Invoice Number	Date	Customer Name	Customer Address	Sales Person	Terms	Product ID	Product Name	Unit Price	Quantity	Amount	Sub Total
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	Charles Wooten	COD	PSV880.006	AMD Athlon X2DC	580	6	3480	9463
						PSV880.037	PDC E5300	645	4	2580	
						LC.V890.002	LG 8.5" LCD	230	10	2300	
						HPQ754.071	HP LaserJet 5200	1103	1	1103	
INV0013	15-Aug-09	Ann / ThisCo	123 Smith Street...	Mary Smith	COD	HP Q754.071	HP LaserJet 5200	1103	2	2206	3356
						LCV890.002	LG 8.5" LCD	230	5	1150	

Invoice example – Spreadsheet Format

Invoice Number	Date	Customer Name	Customer Address	Sales Person	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	Charles Wooten	COD	9463	0	780.70	0
INV0013	15-Aug-09	Ann / ThisCo	123 Smith Street...	Mary Smith	COD	3356	0	100	0

Product ID	Product Name	Unit Price	Quantity	Amount
PSV880.006	AMD Athlon X2DC	580	6	3480
PSV880.037	PDC E5300	645	4	2580
LC.V890.002	LG 8.5” LCD	230	10	2300
HPQ754.071	HP LaserJet 5200	1103	1	1103
HPQ754.071	HP LaserJet 5200	1103	2	2206
LCV890.002	LG 8.5” LCD	230	5	1150

Break into two
But…
How do we connect?

Invoice example – Spreadsheet Format

Invoice Number	Date	Customer Name	Customer Address	Sales Person	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	Charles Wooten	COD	9463	0	780.70	0
INV0013	15-Aug-09	Ann/This Co	123 Smith Street...	Mary Smith	COD	3356	0	100	0

Product ID	Product Name	Unit Price	Quantity	Amount	*Invoice Number*
PSV880.006	AMD Athlon X2DC	580	6	3480	INV0012
PSV880.037	PDC E5300	645	4	2580	INV0012
LC.V890.002	LG 8.5” LCD	230	10	2300	INV0012
HPQ754.071	HP LaserJet 5200	1103	1	1103	INV0012
HPQ754.071	HP LaserJet 5200	1103	2	2206	INV0013
LCV890.002	LG 8.5” LCD	230	5	1150	INV0013

Add FK

Invoice example – keep fixing

Invoice Number	Date	Customer Name	Customer Address	Sales Person	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	Charles Wooten	COD	9463	0	780.70	0
INV0013	15-Aug-09	Ann / ThisCo	123 Smith Street...	Mary Smith	COD	3356	0	100	0

This is about product

Product ID	Product Name	Unit Price	Quantity	Amount	*Invoice Number*
PSV880.006	AMD Athlon X2DC	580	6	3480	INV0012
PSV880.037	PDC E5300	645	4	2580	INV0012
LC.V890.002	LG 8.5” LCD	230	10	2300	INV0012
HPQ754.071	HP LaserJet 5200	1103	1	1103	INV0012
HPQ754.071	HP LaserJet 5200	1103	2	2206	INV0013
LCV890.002	LG 8.5” LCD	230	5	1150	INV0013

This is about Order (invoice)

Invoice example – keep fixing

Invoice Number	Date	Customer Name	Customer Address	Sales Person	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	Charles Wooten	COD	9463	0	780.70	0
INV0013	15-Aug-09	Ann / ThisCo	123 Smith Street...	Mary Smith	COD	3356	0	100	0

Product ID	Quantity	Amount	*Invoice Number*
PSV880.006	6	3480	INV0012
PSV880.037	4	2580	INV0012
LC.V890.002	10	2300	INV0012
HPQ754.071	1	1103	INV0012
HPQ754.071	2	2206	INV0013
LCV890.00	5	1150	INV0013

Break into two

Product ID	Product Name	Unit Price
PSV880.006	AMD Athlon X2DC	580
PSV880.037	PDC E5300	645
LC.V890.002	LG 8.5” LCD	230
HPQ754.071	HP LaserJet 5200	1103

Invoice example – keep fixing

Invoice Number	Date	Customer Name	Customer Address	Sales Person	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	Charles Wooten	COD	9463	0	780.70	0
INV0013	15-Aug-09	Ann / ThisCo	123 Smith Street...	Mary Smith	COD	3356	0	100	0

Product ID	Quantity	Amount	*Invoice Number*
PSV880.006	6	3480	INV0012
PSV880.037	4	2580	INV0012
LC.V890.002	10	2300	INV0012
HPQ754.071	1	1103	INV0012
HPQ754.071	2	2206	INV0013
LCV890.00	5	1150	INV0013

What about amount?

Product ID	Product Name	Unit Price
PSV880.006	AMD Athlon X2DC	580
PSV880.037	PDC E5300	645
LC.V890.002	LG 8.5” LCD	230
HPQ754.071	HP LaserJet 5200	1103

Invoice example – keep fixing

Invoice Number	Date	Customer Name	Customer Address	Sales Person	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	Charles Wooten	COD	9463	0	780.70	0
INV0013	15-Aug-09	Ann / ThisCo	123 Smith Street...	Mary Smith	COD	3356	0	100	0

Product ID	Quantity	*Invoice Number*
PSV880.006	6	INV0012
PSV880.037	4	INV0012
LC.V890.002	10	INV0012
HPQ754.071	1	INV0012
HPQ754.071	2	INV0013
LCV890.00	5	INV0013

Could be derived

Product ID	Product Name	Unit Price
PSV880.006	AMD Athlon X2DC	580
PSV880.037	PDC E5300	645
LC.V890.002	LG 8.5” LCD	230
HPQ754.071	HP LaserJet 5200	1103

Invoice example – keep fixing

Invoice Number	Date	Customer Name	Customer Address	Sales Person	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	Charles Wooten	COD	9463	0	780.70	0
INV0013	15-Aug-09	Ann / ThisCo	123 Smith Street...	Mary Smith	COD	3356	0	100	0

Product ID	Quantity	*Invoice Number*
PSV880.006	6	INV0012
PSV880.037	4	INV0012
LC.V890.002	10	INV0012
HPQ754.071	1	INV0012
HPQ754.071	2	INV0013
LCV890.00	5	INV0013

What about sales person?

Product ID	Product Name	Unit Price
PSV880.006	AMD Athlon X2DC	580
PSV880.037	PDC E5300	645
LC.V890.002	LG 8.5” LCD	230
HPQ754.071	HP LaserJet 5200	1103

Invoice example – keep fixing

Invoice Number	Date	Customer Name	Customer Address	*Sales Person ID*	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	John / Synex	128 Juanita Ave...	1	COD	9463	0	780.70	0
INV0013	15-Aug-09	Ann / This co	123 Smith Street...	2	COD	3356	0	100	0

What about customer?

Product ID	Quantity	*Invoice Number*
PSV880.006	6	INV0012
PSV880.037	4	INV0012
LC.V890.002	10	INV0012
HPQ754.071	1	INV0012
HPQ754.071	2	INV0013
LCV890.00	5	INV0013

Sales Person ID	Sales Person
1	Charles Wooten
2	Mary Smith

Product ID	Product Name	Unit Price
PSV880.006	AMD Athlon X2DC	580
PSV880.037	PDC E5300	645
LC.V890.002	LG 8.5” LCD	230
HPQ754.071	HP LaserJet 5200	1103

Invoice example – Done!

Invoice Number	Date	Customer ID	*Sales Person ID*	Terms	Sub Total	Discount	Sales Tax	Shipping
INV0012	14-Aug-09	1	1	COD	9463	0	780.70	0
INV0013	15-Aug-09	2	2	COD	3356	0	100	0

Product ID	Quantity	*Invoice Number*
PSV880.006	6	INV0012
PSV880.037	4	INV0012
LC.V890.002	10	INV0012
HPQ754.071	1	INV0012
HPQ754.071	2	INV0013
LCV890.00	5	INV0013

Sales Person ID	Sales Person
1	Charles Wooten
2	Mary James

Customer ID	Customer Name	Customer Address
1	John / Synex	128 Juanita Ave...
2	Ann / ThisCo	123 Smith Street...

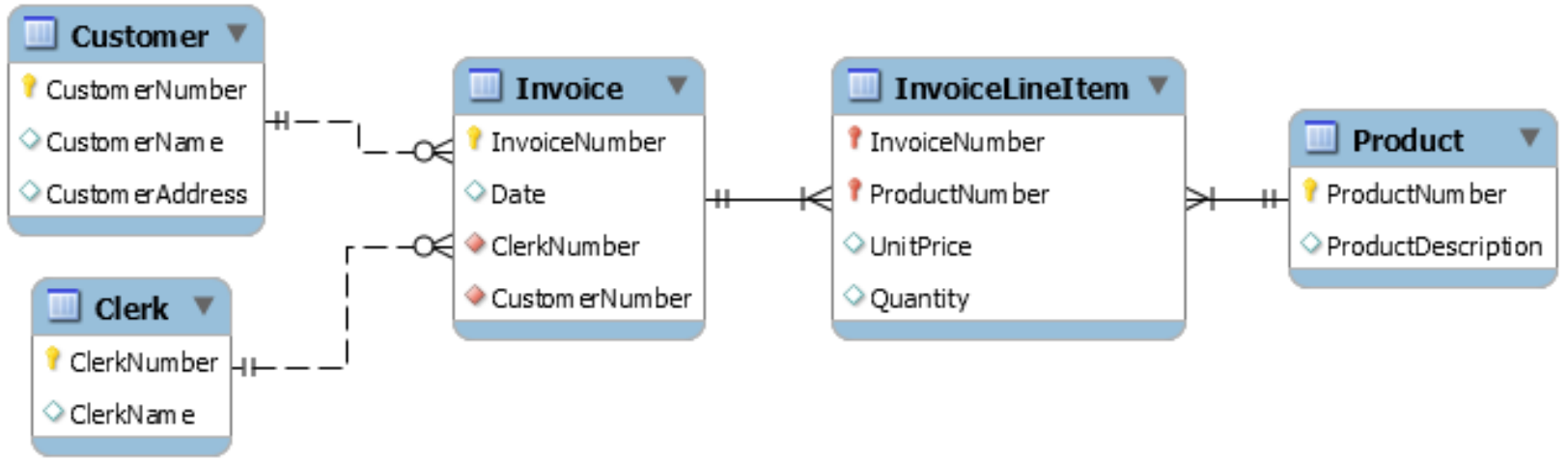
Product ID	Product Name	Unit Price
PSV880.006	AMD Athlon X2DC	580
PSV880.037	PDC E5300	645
LC.V890.002	LG 8.5” LCD	230
HPQ754.071	HP LaserJet 5200	1103

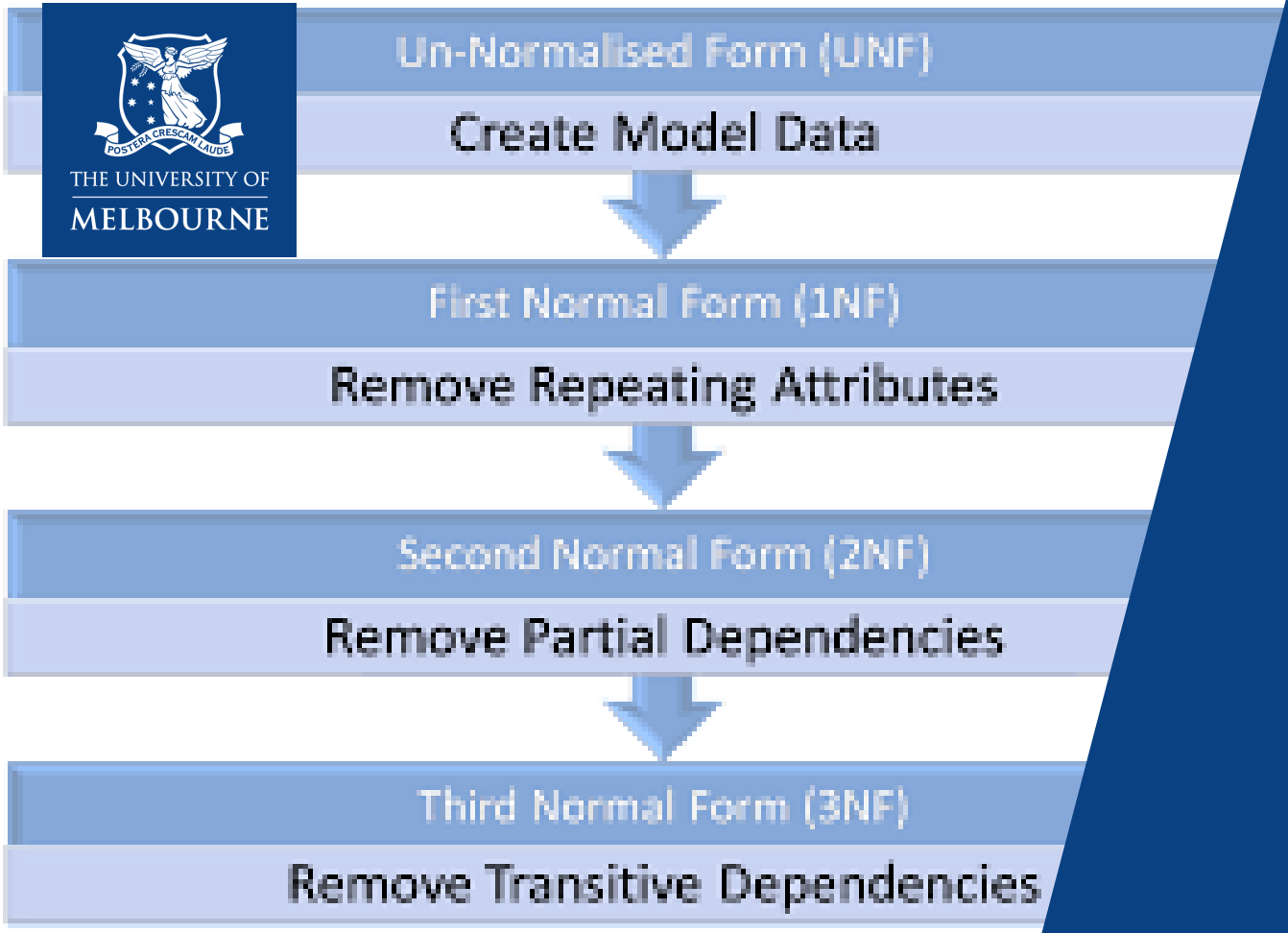
Normalised Relations and ER Diagram

We can name the relations now

Customer (CustomerNumber, CustomerName, CustomerAddress)
Clerk (ClerkNumber, ClerkName)
Product (ProductNumber, ProductDescription)
Invoice (InvoiceNumber, Date, *CustomerNumber*, *ClerkNumber*)
InvoiceLineItem (*InvoiceNumber*, *ProductNumber*, UniPrice, Quantity)

KEY:
PK = Underline
FK = Italic
PFK Underline + Italic





Normalisation and Associated Concepts

Functional Dependency

- A functional dependency concerns values of attributes in a relation
- The attribute Y is **fully functionally dependent** on the attribute X if each value of X determines one and only value of Y
  Notation: X → Y
- X determines Y (If I know X then I also know Y)

StudentID → Student-name

StudentID → Student-email

Functional Dependency:

- Determinants (*X,Y → Z*)
  - the attribute(s) on the left hand side of the arrow

A(X, Y, Z, D)

- Key and Non-Key attributes
  - each attribute is either part of the primary key or it is not

- Partial functional dependency *(Y → Z)*
  - a functional dependency of one or more non-key attributes upon part (but not all) of the primary key, e.g. (StudentID, SubjectCode) -> StudentName

- Transitive dependency *(Z → D)*
  - a functional dependency between 2 (or more) non-key attributes, e.g. SubjCode|SubjectName|LecturerID|LecturerName results in SubjCode->LecturerName

Armstrong's Axioms

Functional dependencies can be identified using Armstrong's Axioms

$$A = (X1, X2, \dots, Xn) \; and \; B = (Y1, Y2, \dots, Yn)$$

1. Reflexivity: $B \subseteq A \Longrightarrow A \rightarrow B$

   B is a subset of A, means
   A functionally determines B

   Example: Student_ID, name -> name

2. Augmentation: $A \rightarrow B \Longrightarrow \mathrm{AC} \rightarrow BC$

   Example: Student_ID -> name => Student_ID, surname ->name, surname

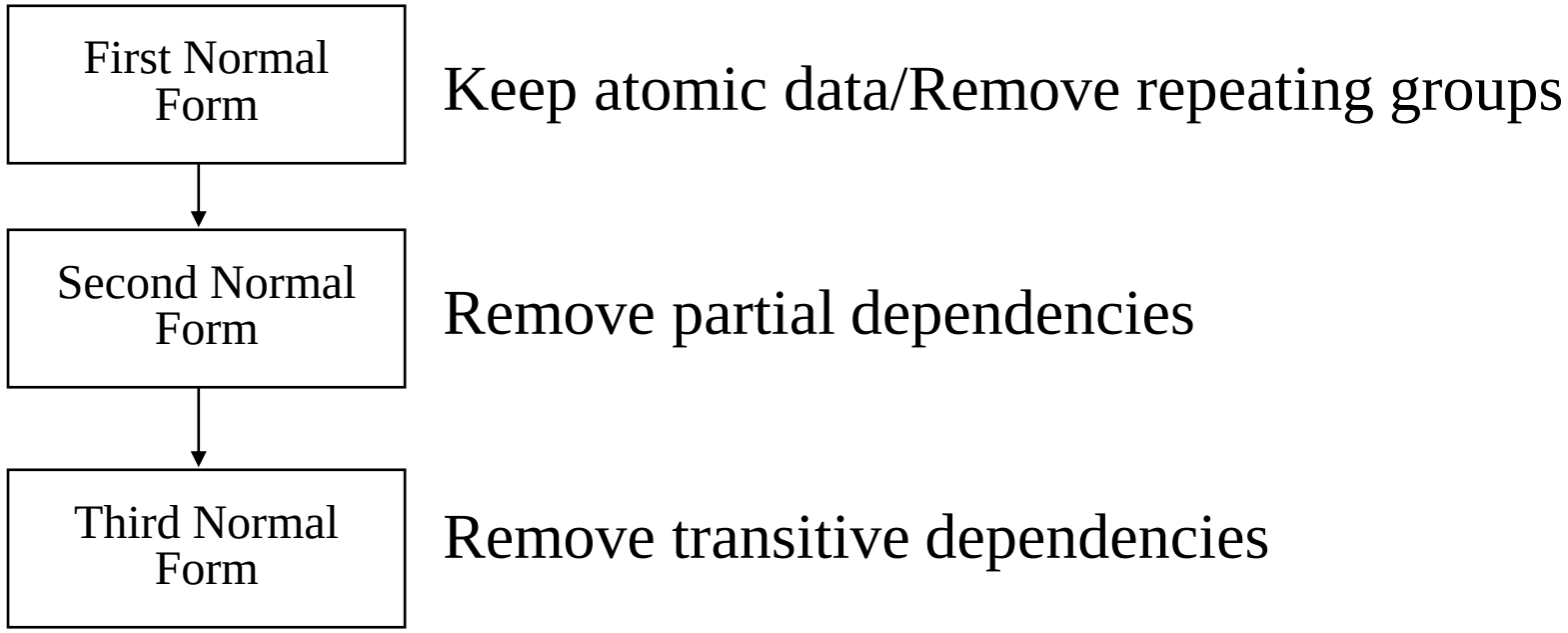
3. Transitivity: $A \rightarrow B \text{ and } \mathrm{B} \rightarrow C \Longrightarrow A \rightarrow C$

   If A functionally determines B and B functionally determines C, then A functionally determines C

   Example: ID -> birthdate, and birthdate -> age then ID ->age

Steps in Normalisation

Normal form - the state of a relation resulting from applying rules about the functional dependency of some attributes upon others



Example: UNF (unnormalised form)

Order#	Customer#	Item#	Description	Qty
1234-5	789912	L236	Lenovo Q27q 27" QHD Monitor with Built-in Speaker	1
		D722	Dell KM5221W Pro Wireless Keyboard and Mouse	1
2577-1	321154	B614	Belkin High Speed HDMI Cable with Ethernet (5M)	2
		A343	ALOGIC Ultra USB-C to HDMI Cable (2m)	1
		M766	Microsoft Surface Dock 2	1

First Normal Form (1NF)

Remove Repeating Groups/Keep Atomic Data

- repeating groups of attributes cannot be represented in a flat, two-dimensional table
  - removing cells with multiple values (keep atomic data)
    - eliminate nulls

OrderItem

Order#	Customer#	Item#	Description	Qty
1234-5	789912	L236	Lenovo Q27q 27" QHD Monitor with Built-in Speaker	1
1234-5	789912	D722	Dell KM5221W Pro Wireless Keyboard and Mouse	1
2577-1	321154	B614	Belkin High Speed HDMI Cable with Ethernet (5M)	2
2577-1	321154	A343	ALOGIC Ultra USB-C to HDMI Cable (2m)	1
2577-1	321154	M766	Microsoft Surface Dock 2	1

First Normal Form (1NF) - cont

Identify PK

Example: Order-Item (**Order#**, Customer#, (**Item#**, Desc, Qty))

Identify all dependencies

The obvious dependency:

Order#, Item# →Customer#, Desc, Qty

Additional dependencies:

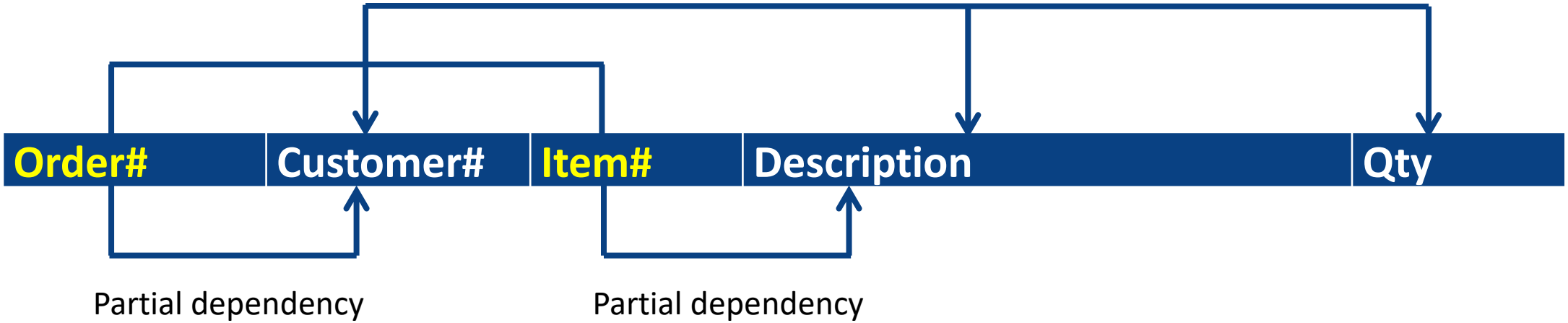
Order# → Customer#

Order#, Item# → Desc, Qty

Partial dependency:

Item# → Desc

First Normal Form Dependency Diagram



- The arrows above the attributes indicate all desirable dependencies, i.e. dependencies based on PK Order# and Item#
- The arrows below the attributes indicate partial dependencies

Partial Dependency Anomalies

Order-Item (Order#, Item#, Desc, Qty)

OrderItem

Order#	Item#	Desc	Qty
1234-5	L236	Lenovo Q27q 27" QHD Monitor with Built-in Speaker	1
1234-5	D722	Dell KM5221W Pro Wireless Keyboard and Mouse	1
2577-1	B614	Belkin High Speed HDMI Cable with Ethernet (5M)	2
2577-1	A343	ALOGIC Ultra USB-C to HDMI Cable (2m)	1
2577-1	M766	Microsoft Surface Dock 2	1

- *UPDATE* change item desc in many places
- *DELETE* data for last item lost when last order for that item is deleted
- *INSERT* cannot add new item until it is ordered

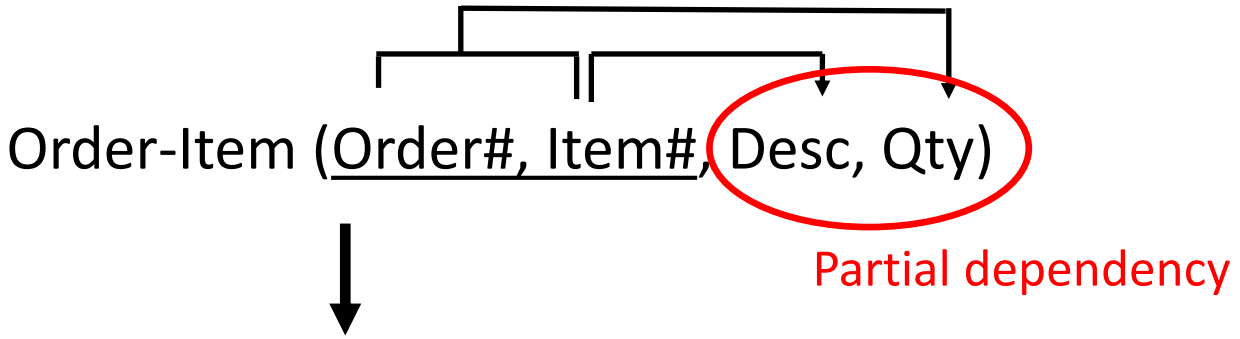
Second Normal Form (2NF)

2NF occurs only when the 1NF has a composite PK

If a table has a single-attribute PK, it's already in 2NF

Eliminate Partial Dependencies - create new tables

Example: Order-Item (Order#, Item#, Desc, Qty)



Item (Item#, Desc)

Order-Item (Order#, Item#, Qty)

Solution to these Anomalies

OrderItem

Order#	Item#	Qty
1234-5	L236	1
1234-5	D722	1
2577-1	B614	2
2577-1	A343	1
2577-1	M766	1

delete last order for item, but item remains

Item

Item#	Description
L236	Lenovo Q27q 27" QHD Monitor with Built-in Speaker
D722	Dell KM5221W Pro Wireless Keyboard and Mouse
B614	Belkin High Speed HDMI Cable with Ethernet (5M)
A343	ALOGIC Ultra USB-C to HDMI Cable (2m)
M766	Microsoft Surface Dock 2

add new item at any time

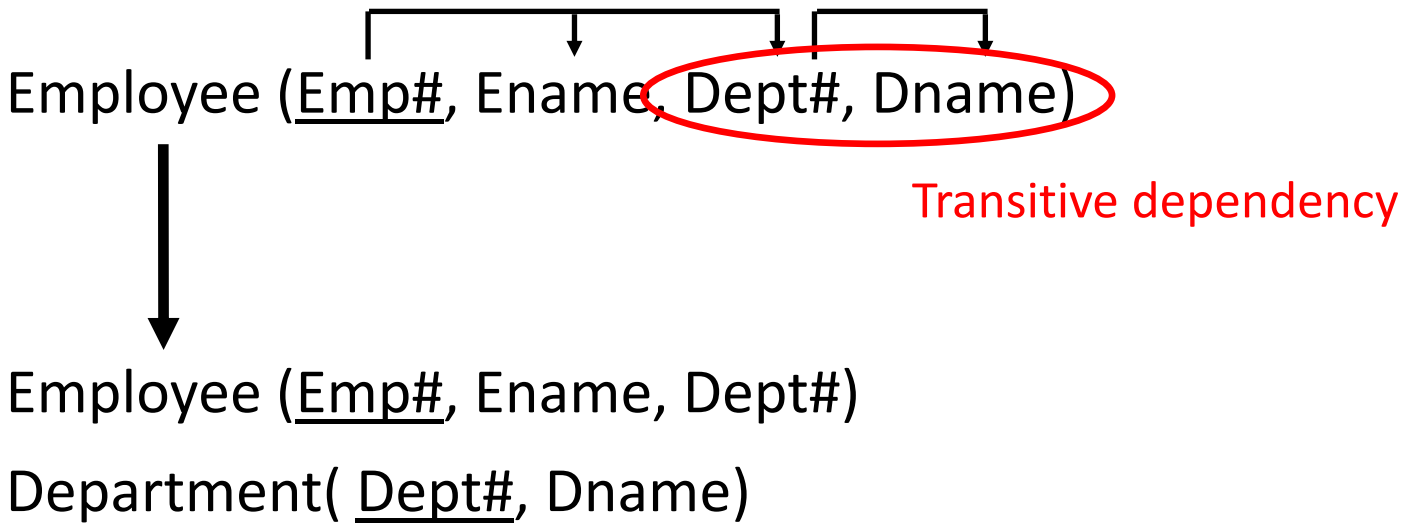
change item description in one place

Third Normal Form

Eliminate Transitive Dependencies

a non-key attribute **cannot be identified** by another non-key attribute

Example: Employee (Emp#, Ename, Dept#, Dname)



Transitive Dependency Anomalies

Example: Employee (Emp#, Ename, Dept#, Dname)

Emp#	EName	Dept#	DName
101	Emma Smith	D3	IT
203	Aron Lee	D4	Sales
251	Sizan Black	D4	Sales
304	June Chen	D1	HR
341	Nick Russo	D3	IT

- *UPDATE* change dept name in many places
- *DELETE* data for dept lost when last employee for that dept is deleted
- *INSERT* cannot add new dept until an employee is allocated to it

Solution to these Anomalies

Employee

Emp#	EName	Dept#
101	Emma Smith	D3
203	Aron Lee	D4
251	Sizan Black	D4
304	June Chen	D1
341	Nick Russo	D3

delete last employee in dept, but dept remains

add new department at any time

Dept#	DName
D3	IT
D4	Sales
D4	Sales
D1	HR
D3	IT

Department

change department name in one place

Boyce-Codd Normal Form

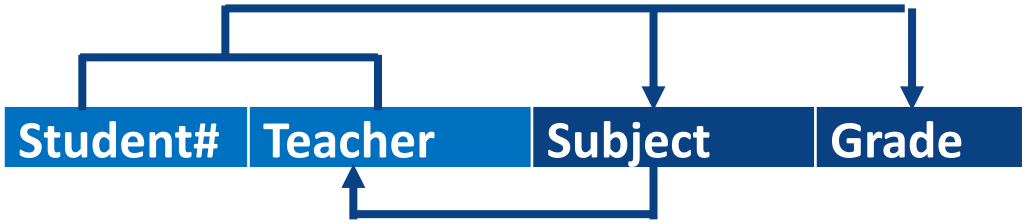
Every determinant must be a candidate key

"Every non-key attribute must provide a fact about the key, the whole key, and nothing but the key. (So help me Codd)"

Example: Allocation (Student#, Teacher, Subject, Grade)

Student#	Teacher	Subject	Grade
123344	Darwin	Biology	D
123344	Curie	Chemistry	C
455667	Curie	Chemistry	HD
788994	Curie	Physics	C
877995	Darwin	Biology	HD

Student#+Teacher → Subject, Grade
Subject → Teacher



Boyce-Codd Anomalies

Allocation

Student#	Teacher	Subject	Grade
123344	Darwin	Biology	D
123344	Curie	Chemistry	C
455667	Curie	Chemistry	HD
788994	Curie	Physics	C
877995	Darwin	Biology	HD

UPDATE student changing subject may lose teacher for that subject

DELETE deleting student may lose subject allocated to that teacher

CREATE cannot assign a subject to a teacher until a student takes it

Solution to these Anomalies

Student#	Subject	Grade
123344	Biology	D
123344	Chemistry	C
455667	Chemistry	HD
788994	Physics	C
877995	Biology	HD

delete student but subject remains

Change enrolment without losing teacher/subject

Subject	Teacher
Biology	Darwin
Chemistry	Curie
Chemistry	Curie
Physics	Curie
Biology	Darwin

add a subject and assign a teacher to it

Boyce-Codd is taught for completeness but is not examinable

What's examinable

- **Normalisation Process (1NF -> 2NF -> 3NF)**
- **Anomalies**
- **Functional dependencies**
- **Denormalisation**

THE UNIVERSITY OF MELBOURNE

Thank you

Subtitle

Identifier first line

Second line