



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

200
1822 - 2022

Normalisation

Man Geni Addysg Uwch yng Nghymru
Where Higher Education Began in Wales

www.uwtsd.ac.uk

Week:

30-Sep-2019

Today's Lecture

- Recap - Relational Model Terminology
- Functional dependency
- Normalisation
- Indexing

<https://forms.office.com/e/VMLsr3pH4E>



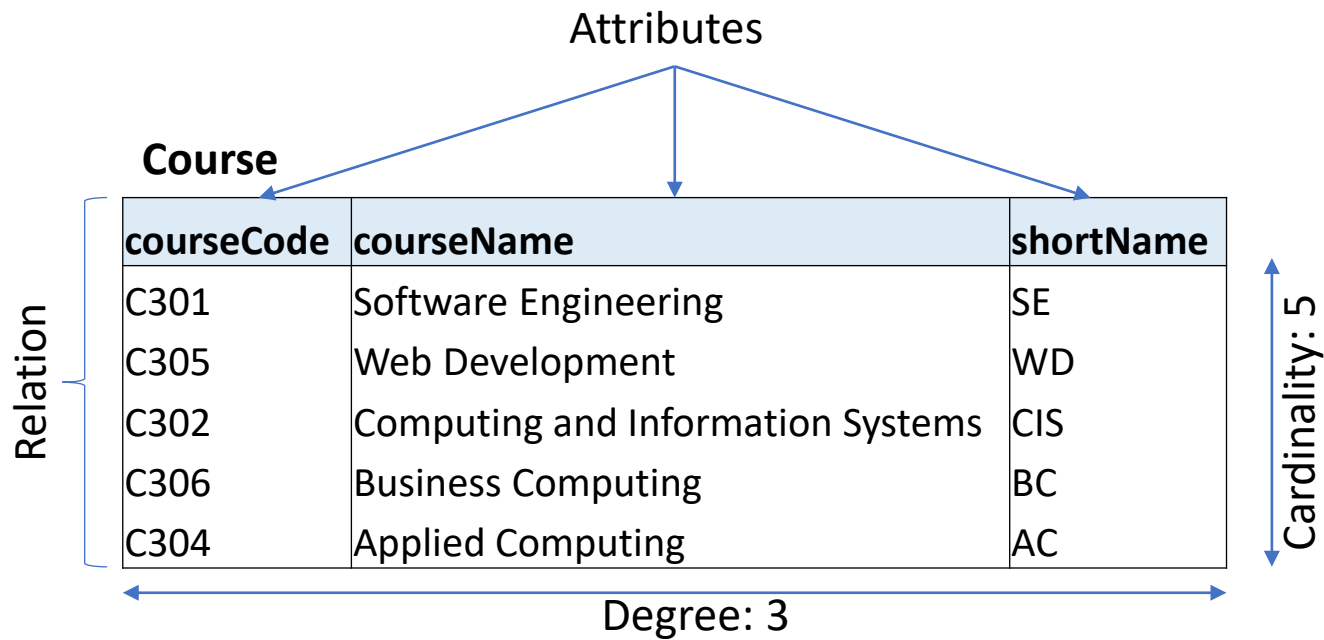
Recap - Relational Model Terminology

- **Relation:** a table with columns and rows
 - This concept only applies to the logical structure of the database: at the external and conceptual levels
- **Attribute:** a named column of a relation
- **Domain:** the set of allowed values for one or more attributes.
- **Tuple:** a row of a relation.

Relational Model Terminology

- **Degree:** the number of attributes in a relation.
- **Cardinality:** the number of tuples in a relation.
- **Relational Database:** collection of *normalised* relations with distinct relation names.

Instances of Course and Student Relations



Normalisation

What is Normalisation?

- **Normalisation** is a technique for producing a set of *suitable relations that support the data requirements* of an enterprise.

Employee

Man#	Name	Birthdate	JobHistory				Children	
123	Michael Swart	Nov. 22	JobDate	Title	SalaryHistory		Childname	Birthyear
			2000	Lackey	SalaryDate	Salary	Mini-me 1	2000
					2000	\$1,000,000	Mini-me 2	2002
					2001	\$2,000,000		
			2002	Senior Lackey	SalaryDate	Salary		
					2002	\$3,000,000		
					2003	\$4,000,000		

What is Normalisation?

- **Normalisation** is a technique for producing a set of *suitable relations that support the data requirements* of an enterprise.

Employee'

Man#	Name	Birthdate
123	Michael Swart	Nov. 22

JobHistory'

<u>Man#</u>	<u>JobDate</u>	Title
123	2000	Lackey
123	2002	Senior Lackey

Children'

Man#	ChildName	BirthYear
123	Mini-me 1	2000
123	Mini-me 2	2002

SalaryHistory'

<u>Man#</u>	<u>JobDate</u>	<u>SalaryDate</u>	Salary
123	2000	2000	\$1,000,000
123	2000	2001	\$2,000,000
123	2002	2002	\$3,000,000
123	2003	2003	\$4,000,000

The Purpose of Normalisation

- The main aim of a relational database design is to group attributes into relations to ***minimise data redundancy***.
- Normalisation, helps to achieve this.

Problems of not properly normalised tables

- There are two main problems associated with a relation that is not appropriately normalised:
 - Data Redundancy
 - Update Anomalies for:
 - Insertions (adding new),
 - Deletions, and
 - Updates (modifications).

Data Redundancy and Update Anomalies

- What will be the primary key for this table?
- Can you spot potential problems with this table design?

CampusStaff Table

staffNo	staffName	position	salary	campusID	campusAddress
SW101	John Smith	Lecturer	30000	C002	Swansea IQ, SA1 8PH
SW102	Diane Lee	PVC	70000	C001	Carmarthen Campus, Carmarthen, SA31 3EP
SW103	Ann Beech	Lecturer	18000	C002	Swansea IQ, SA1 8PH
SW105	David Ford	Snr Lecturer	41000	C001	Carmarthen Campus, Carmarthen, SA31 3EP
SW106	Mary Howe	Dean, Humanities	55000	C003	Lampeter Campus Ceredigion SA48 7ED
SW107	Susan Brand	Receptionist	18000	C001	Swansea IQ, SA1 8PH

- Swansea IQ address is repeated for 3 staff members
- Carmarthen is repeated for 2 staff members

CampusStaff Table

staffNo	staffName	position	salary	campusID	campusAddress
SW101	John Smith	Lecturer	30000	C002	Swansea IQ, SA1 8PH
SW102	Diane Lee	PVC	70000	C001	Carmarthen Campus, Carmarthen, SA31 3EP
SW103	Ann Beech	Lecturer	18000	C002	Swansea IQ, SA1 8PH
SW105	David Ford	Snr Lecturer	41000	C001	Carmarthen Campus, Carmarthen, SA31 3EP
SW106	Mary Howe	Dean, Humanities	55000	C003	Lampeter Campus Ceredigion SA48 7ED
SW107	Susan Brand	Receptionist	18000	C001	Swansea IQ, SA1 8PH

Deletion Anomalies

- If we delete a row from the CampusStaff table representing the last member of staff at the campus, the details about that campus are lost.
- For example, deleting the record for Mary Howe means the details for the campus C003 are also lost.

CampusStaff Table

staffNo	staffName	position	salary	campusID	campusAddress
SW101	John Smith	Lecturer	30000	C002	Swansea IQ, SA1 8PH
SW102	Diane Lee	PVC	70000	C001	Carmarthen Campus, Carmarthen, SA31 3EP
SW103	Ann Beech	Lecturer	18000	C002	Swansea IQ, SA1 8PH
SW105	David Ford	Snr Lecturer	41000	C001	Carmarthen Campus, Carmarthen, SA31 3EP
SW106	Mary Howe	Dean, Humanities	55000	C003	Lampeter Campus Ceredigion SA48 7ED
SW107	Susan Brand	Receptionist	18000	C001	Swansea IQ, SA1 8PH

The Better Design

- The anomaly problems of the Staff table can be eliminated by ***decomposing this relation into smaller relations***.
 - **Campus** (campusID, campusAddress)
 - **Staff** (staffNo, staffName, position, salary, campusID)
- The primary keys (PKs) for each table are underlined.
- CampusID is a foreign key (FK) in the Staff table linking a member of staff to the correct campus.

Introduction to Functional Dependencies

- What are Functional Dependencies (FD)?

Example Dataset 1: Student Information

Student_ID	Name	Department	HOD
101	Alice	Computer Science	Dr. John
102	Bob	Mechanical Eng.	Dr. Smith
103	Charlie	Computer Science	Dr. John

- Identify the attribute(s) that uniquely determine student's name and department?
- Identify the attribute(s) that uniquely determine the head of department?

Introduction to Functional Dependencies

- What are Functional Dependencies (FD)?

Example Dataset 1: Student Information

Student_ID	Name	Department	HOD
101	Alice	Computer Science	Dr. John
102	Bob	Mechanical Eng.	Dr. Smith
103	Charlie	Computer Science	Dr. John

- Student ID uniquely determines the student's name and department
- Department determines the Head of Department

Introduction to Functional Dependencies

- What are Functional Dependencies (FD)?

Example Dataset 1: Student Information

Student_ID	Name	Department	HOD
101	Alice	Computer Science	Dr. John
102	Bob	Mechanical Eng.	Dr. Smith
103	Charlie	Computer Science	Dr. John

- Student ID uniquely determines the student's name and department

Student_ID → Name, Department

Determinant

Dependents

- Department determines the Head of Department

Department → HOD

Determinant

Dependent

The arrowline is
read as
'determines'

Activity: Identifying Functional Dependencies

With the following dataset identify the functional dependencies.

Employee_ID	Employee_Name	Department	Manager
201	John	HR	Mr. Parker
202	Sarah	IT	Mr. Stevens
203	Alice	HR	Mr. Parker

Activity: Identifying Functional Dependencies

With the following dataset identify the functional dependencies.

Employee_ID	Employee_Name	Department	Manager
201	John	HR	Mr. Parker
202	Sarah	IT	Mr. Stevens
203	Alice	HR	Mr. Parker

Employee_ID → Employee_Name, Department

Department → Manager

Normalization - First Normal Form (1NF)

Example Dataset 2: Non-1NF

Student_ID	Name	Courses
101	Alice	CSE101, MTH102
102	Bob	PHY101
103	Charlie	CSE101, ENG101, PHY101

Normalization - First Normal Form (1NF)

Example Dataset 2: Non-1NF

Student_ID	Name	Courses
101	Alice	CSE101, MTH102
102	Bob	PHY101
103	Charlie	CSE101, ENG101, PHY101

1NF requires that all entries in a column are atomic (indivisible), and each row must be unique.

Normalization - First Normal Form (1NF)

Example Dataset 2: Non-1NF

Student_ID	Name	Courses
101	Alice	CSE101, MTH102
102	Bob	PHY101
103	Charlie	CSE101, ENG101, PHY101

contains
atomic values

1NF requires that all entries in a column are atomic (indivisible), and each row must be unique.

Convert to 1NF:

Student_ID	Name	Course
101	Alice	CSE101
101	Alice	MTH102
102	Bob	PHY101
103	Charlie	CSE101
103	Charlie	ENG101
103	Charlie	PHY101

Activity - First Normal Form (1NF)

Normalize the table below to 1NF

stu_id	name	department	hod	address
101	Alice	CS	Dr. Smith	12 Park Street, Swansea SA1 3EF
102	Bob	IT	Dr. Jones	45 Queen Road, Cardiff CF10 1AA

Activity - First Normal Form (1NF)

1NF requires that all entries in a column are atomic (indivisible), and each row must be unique.

stu_id	name	department	hod	door_no	street_name	city	postcode
101	Alice	CS	Dr. Smith	12	Park Street	Swansea	SA1 3EF
102	Bob	IT	Dr. Jones	45	Queen Road	Cardiff	CF10 1AA

Normalization - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

StudentID	StudentName	CourseID	CourseName	InstructorName	Grade
101	Alice Johnson	C101	Database Systems	Dr. Smith	A
101	Alice Johnson	C102	Data Structures	Dr. Johnson	B
102	Bob Brown	C101	Database Systems	Dr. Smith	B

Identify the composite key from the above table?

Normalization - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

StudentID	StudentName	CourseID	CourseName	InstructorName	Grade
101	Alice Johnson	C101	Database Systems	Dr. Smith	A
101	Alice Johnson	C102	Data Structures	Dr. Johnson	B
102	Bob Brown	C101	Database Systems	Dr. Smith	B

Identify the composite key from the above table?

composite key is **{StudentID, CourseID}**

Which FD is enough to uniquely determine studentname?

a) {StudentID, CourseID} -> {StudentName}

b) {StudentID} -> {StudentName}

Normalization - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

StudentID	StudentName	CourseID	CourseName	InstructorName	Grade
101	Alice Johnson	C101	Database Systems	Dr. Smith	A
101	Alice Johnson	C102	Data Structures	Dr. Johnson	B
102	Bob Brown	C101	Database Systems	Dr. Smith	B

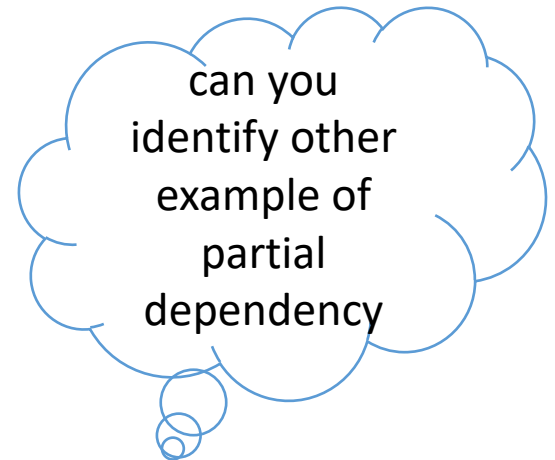
Identify the composite key from the above table?

composite key is **{StudentID, CourseID}**

Which FD is enough to uniquely determine studentname?

a) {StudentID, CourseID} -> {StudentName}

b) {StudentID} -> {StudentName}



Normalization - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

StudentID	StudentName	CourseID	CourseName	InstructorName	Grade
101	Alice Johnson	C101	Database Systems	Dr. Smith	A
101	Alice Johnson	C102	Data Structures	Dr. Johnson	B
102	Bob Brown	C101	Database Systems	Dr. Smith	B

Identify the composite key from the above table?

composite key is {**StudentID**, **CourseID**}

Which FD is enough to uniquely determine CourseName?

a) {StudentID, CourseID} -> {CourseName}

b) {CourseID} -> {CourseName}

c) {CourseID} -> {InstructorName}

Normalization - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

StudentID	StudentName	CourseID	CourseName	InstructorName	Grade
101	Alice Johnson	C101	Database Systems	Dr. Smith	A
101	Alice Johnson	C102	Data Structures	Dr. Johnson	B
102	Bob Brown	C101	Database Systems	Dr. Smith	B

Can you create two tables to eliminate partial dependencies?

Normalization - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

StudentID	StudentName	CourseID	CourseName	InstructorName	Grade
101	Alice Johnson	C101	Database Systems	Dr. Smith	A
101	Alice Johnson	C102	Data Structures	Dr. Johnson	B
102	Bob Brown	C101	Database Systems	Dr. Smith	B

Students Table (after removing partial dependencies):

StudentID	StudentName
101	Alice Johnson
102	Bob Brown

Courses Table (after removing partial dependencies):

CourseID	CourseName	InstructorName
C101	Database Systems	Dr. Smith
C102	Data Structures	Dr. Johnson

Where to
place
'Grade'?

Normalization - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

Students Table (after removing partial dependencies):

StudentID	StudentName
101	Alice Johnson
102	Bob Brown

Courses Table (after removing partial dependencies):

CourseID	CourseName	InstructorName
C101	Database Systems	Dr. Smith
C102	Data Structures	Dr. Johnson

Enrollment Table (no partial dependencies, stores relationships and grades):

StudentID	CourseID	Grade
101	C101	A
101	C102	B
102	C101	B

Activity - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

EmployeeID	EmployeeName	DeptID	DeptName	ProjectID	ProjectName	HoursWorked
E001	Alice Johnson	D01	HR	P101	Payroll System	30
E001	Alice Johnson	D01	HR	P102	Recruitment App	25
E002	Bob Brown	D02	IT	P101	Payroll System	40
E003	Charlie Davis	D01	HR	P103	Employee Portal	20
E004	Daisy White	D03	Finance	P104	Budgeting App	35

Step 1: Identify the Composite Key

Step 2: Identify Partial Dependencies

Step 3: Break the Table into 2NF Tables

Activity - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

Employee Table (2NF)

EmployeeID	EmployeeName	DeptID
E001	Alice Johnson	D01
E002	Bob Brown	D02
E003	Charlie Davis	D01
E004	Daisy White	D03

Department Table (2NF)

DeptID	DeptName
D01	HR
D02	IT
D03	Finance

Activity - Second Normal Form (2NF)

- The table must be in 1NF.
- There should be no **partial dependency**

Project Table (2NF)

ProjectID	ProjectName
P101	Payroll System
P102	Recruitment App
P103	Employee Portal
P104	Budgeting App

Assignment Table (2NF)

This table contains the relationships between employees and the projects they are working on, including the HoursWorked.

EmployeeID	ProjectID	HoursWorked
E001	P101	30
E001	P102	25
E002	P101	40
E003	P103	20
E004	P104	35

Normalization - Third Normal Form (3NF)

- The table must be in 2NF.
- There should be no **transitive dependency**

In **mathematics** (especially set theory):

A relation **R** is *transitive* if
whenever aRb and bRc , then aRc .

Example with numbers:

If $a < b$ and $b < c \rightarrow$ then $a < c$.

$2 < 3$ and $3 < 4 \rightarrow$ then $2 < 4$

(Here “<” is a transitive relation.)

Normalization - Third Normal Form (3NF)

StudentID	StudentName	CourseID	CourseName	InstructorName	InstructorPhone
101	Alice Johnson	C101	Math	Dr. Smith	123-456-7890
102	Bob Brown	C102	History	Dr. Adams	987-654-3210
101	Alice Johnson	C102	History	Dr. Adams	987-654-3210

StudentID → **StudentName**: The student's name depends on the student's ID.

CourseID → **CourseName**: The course name depends on the course's ID.

CourseID → **InstructorName**: The instructor's name depends on the course.

InstructorName → **InstructorPhone**: The instructor's phone number depends on the instructor.

Normalization - Third Normal Form (3NF)

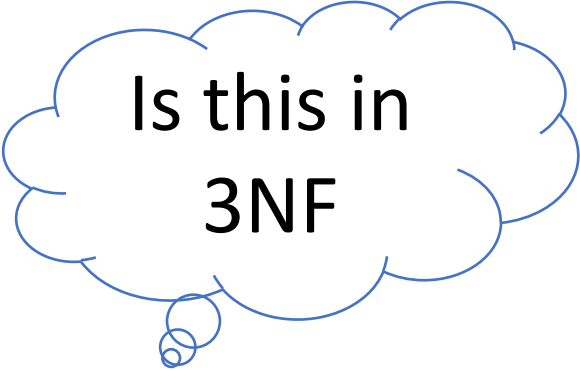
- The table must be in 2NF.
- There should be no **transitive dependency**

StudentID → **StudentName**

CourseID → **CourseName**

CourseID → **InstructorName**

InstructorName → **InstructorPhone**



Is this in
3NF

1. Student Table (No changes):

StudentID	StudentName
101	Alice Johnson
102	Bob Brown

2. Course Table (No transitive dependency):

CourseID	CourseName	InstructorName
C101	Math	Dr. Smith
C102	History	Dr. Adams

3. Instructor Table (Instructor phone moved to a separate table):

InstructorName	InstructorPhone
Dr. Smith	123-456-7890
Dr. Adams	987-654-3210

Recommended 3NF Design

Student Table

stu_id	name
101	Alice
102	Bob

Instructor Table

instr_id	instr_name	instr_phone
1	Dr. Smith	12345678
2	Dr. Jones	8754321

Course Table

course_id	course_name	instr_id
C1	Math	1
C2	Physics	2

Enrolment Table

stu_id	course_id
101	C1
101	C2
102	C2

Lab Exercise – Normalize to 3NF

Below is a table of library borrowings and associated student and book information.

BorrowID	StudentID	StudentName	StudentPhone	BookID	BookTitle	Author	BorrowDate	ReturnDate	LibraryBranch
B001	S101	Alice Johnson	123-456-7890	BK101	Database Systems	Dr. Tan	2024-09-01	2024-09-10	Main Branch
B002	S102	Bob Brown	987-654-3210	BK102	Operating Systems	Dr. Wong	2024-09-02	2024-09-12	Main Branch
B003	S101	Alice Johnson	123-456-7890	BK103	Data Structures	Dr. Gupta	2024-09-03	2024-09-13	West Branch
B004	S103	Charlie Davis	555-987-1234	BK101	Database Systems	Dr. Tan	2024-09-05	2024-09-15	Main Branch
B005	S104	Daisy White	666-789-3456	BK104	Network Security	Dr. Smith	2024-09-06	2024-09-16	East Branch

Database Normalization!

1NF made you clean up the mess,

2NF asked you to stop being so *partially dependent*, and

3NF taught you that transitive dependencies are more confusing than they sound!

Further Reading :

Chapter 11 Database Analysis and the DreamHome Case Study

Chapter 14: Normalisation

Chapter 12: Entity–Relationship Modelling

Chapter 13 Enhanced Entity–Relationship Modelling

Database Systems: A Practical Approach to Design, Implementation,
and Management Sixth edition by • Thomas Connolly • Carolyn Begg