8.7 Normalizes database schema to improve performance

Time: 6 periods

Learning Outcomes

- Describes the functional dependencies and categorizes them
- Describes abnormalities of an improperly designed table when modifying in terms of insert, update and delete
- Describes the zero normal form
- Explains the abnormalities which are reduced after the first normal form
- Lists the conditions for executing the second normal form
- Explains the abnormalities which are reduced after the second normal form
- Explains the abnormalities which are reduced after the third normal form

Normalization

Normalization is a process to **evaluate and correct the table structure**

Normalization is a another **DB design tool**

Why Normalization? A Case Study

- BCG company manages building group projects
- Purpose of BCG's data
  - Indicate the number of hours spent on each project
  - Indicate the hourly rate for each employee
  - To generate report about each project
  - Also auxiliary data reporting

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
		101	John G. News	Database Designer	105.00	19.4
		105	Alice K. Johnson *	Database Designer	105.00	35.7
		106	William Smithfield	Programmer	35.75	12.6
		102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
		118	James J. Frommer	General Support	18.36	45.3
		104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
		112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
		104	Anne K. Ramoras	Systems Analyst	96.75	48.4
		113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
		111	Geoff B. Wabash	Clerical Support	26.87	22.0
		106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
		115	Travis B. Bawangi	Systems Analyst	96.75	45.8
		101	John G. News *	Database Designer	105.00	56.3
		114	Annelise Jones	Applications Designer	48.10	33.1
		108	Ralph B. Washington	Systems Analyst	96.75	23.6
		118	James J. Frommer	General Support	18.36	30.5
		112	Darlene M. Smithson	DSS Analyst	45.95	41.4

Uniquely identify records?

Unnormalised Normal Form (UNF)

- Definition: A relation is unnormalised when it has not had any normalisation rules applied to it, and it suffers from various anomalies.

This only tends to occur where the relation has been designed using a 'bottom-up approach'. *i.e., the capturing of attributes to a 'Universal Relation' from a screen layout, manual report, manual document, etc...*

Normalization

- Normalization is a process to **evaluate and correct the table structure**
- Purpose
  - **To eliminate data redundancies**
- Normalization process
  - Goes through **a number of stages**
  - Each stage is a "**normal form(NF)**"

- First stage is called the first normal form(1NF)
- Generally, a higher normal form is better than a lower normal form, i.e. (n+1)-NF is better than n-NF
- **3NF** or **BCNF** is the furthest we go, for most situations
- Higher normal forms has less redundancies
  - But not always desirable. Why?

Normal forms

NORMAL FORM	CHARACTERISTIC
First normal form (1NF)	Table format, no repeating groups, and PK identified
Second normal form (2NF)	1NF and no partial dependencies
Third normal form (3NF)	2NF and no transitive dependencies
Boyce-Codd normal form (BCNF)	Every determinant is a candidate key (special case of 3NF)
Fourth normal form (4NF)	3NF and no independent multivalued dependencies

Normalization Objectives

- To ensure tables have the following characteristics:
  - Each table represents a **single subject**
  - **No duplication**, i.e., no data item will be *unnecessarily* stored in more than one table
  - All attributes in a table are uniquely identified by a **primary key**
  - Each table is **void of anomalies** for data insertion, updating, or deletion

First Normal Form (1NF)

- A table **must not contain repeating groups**
  - A **repeating group** is a group of entries/records that exist for any single key attribute occurrence
- Informally
  - A "key" **should not identify a group of records**
  - A "key" **should identify one record**
- 1NF is to **get rid of repeating groups**

A table not in 1NF

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	$84.50	23.8
		101	John G. News	Database Designer	$105.00	19.4
		105	Alice K. Johnson *	Database Designer	$105.00	35.7
		106	William Smithfield	Programmer	$35.75	12.6
		102	David H. Senior	Systems Analyst	$96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	$48.10	24.6
		118	James J. Frommer	General Support	$18.36	45.3
		104	Anne K. Ramoras *	Systems Analyst	$96.75	32.4
		112	Darlene M. Smithson	DSS Analyst	$45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	$105.00	64.7
		104	Anne K. Ramoras	Systems Analyst	$96.75	48.4
		113	Delbert K. Joenbrood *	Applications Designer	$48.10	23.6
		111	Geoff B. Wabash	Clerical Support	$26.87	22.0
		106	William Smithfield	Programmer	$35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	$35.75	24.6
		115	Travis B. Bawangi	Systems Analyst	$96.75	45.8
		101	John G. News *	Database Designer	$105.00	56.3
		114	Annelise Jones	Applications Designer	$48.10	33.1
		108	Ralph B. Washington	Systems Analyst	$96.75	23.6
		118	James J. Frommer	General Support	$18.36	30.5
		112	Darlene M. Smithson	DSS Analyst	$45.95	41.4

Conversion to1NF

- To achieve 1NF, you should
  - Eliminate the repeating groups by making sure **each attribute contains an appropriate data value**
  - Identify the **Primary Key (PK)**
  - Identify all **dependencies**

A table in 1NF

PK: PROJ_NUM + EMP_NUM

PROJ_NUM	PROJ_NAME	EMP_NUM	EMP_NAME	JOB_CLASS	CHG_HOUR	HOURS
15	Evergreen	103	June E. Arbough	Elect. Engineer	84.50	23.8
15	Evergreen	101	John G. News	Database Designer	105.00	19.4
15	Evergreen	105	Alice K. Johnson *	Database Designer	105.00	35.7
15	Evergreen	106	William Smithfield	Programmer	35.75	12.6
15	Evergreen	102	David H. Senior	Systems Analyst	96.75	23.8
18	Amber Wave	114	Annelise Jones	Applications Designer	48.10	24.6
18	Amber Wave	118	James J. Frommer	General Support	18.36	45.3
18	Amber Wave	104	Anne K. Ramoras *	Systems Analyst	96.75	32.4
18	Amber Wave	112	Darlene M. Smithson	DSS Analyst	45.95	44.0
22	Rolling Tide	105	Alice K. Johnson	Database Designer	105.00	64.7
22	Rolling Tide	104	Anne K. Ramoras	Systems Analyst	96.75	48.4
22	Rolling Tide	113	Delbert K. Joenbrood *	Applications Designer	48.10	23.6
22	Rolling Tide	111	Geoff B. Wabash	Clerical Support	26.87	22.0
22	Rolling Tide	106	William Smithfield	Programmer	35.75	12.8
25	Starflight	107	Maria D. Alonzo	Programmer	35.75	24.6
25	Starflight	115	Travis B. Bawangi	Systems Analyst	96.75	45.8
25	Starflight	101	John G. News *	Database Designer	105.00	56.3
25	Starflight	114	Annelise Jones	Applications Designer	48.10	33.1
25	Starflight	108	Ralph B. Washington	Systems Analyst	96.75	23.6
25	Starflight	118	James J. Frommer	General Support	18.36	30.5
25	Starflight	112	Darlene M. Smithson	DSS Analyst	45.95	41.4

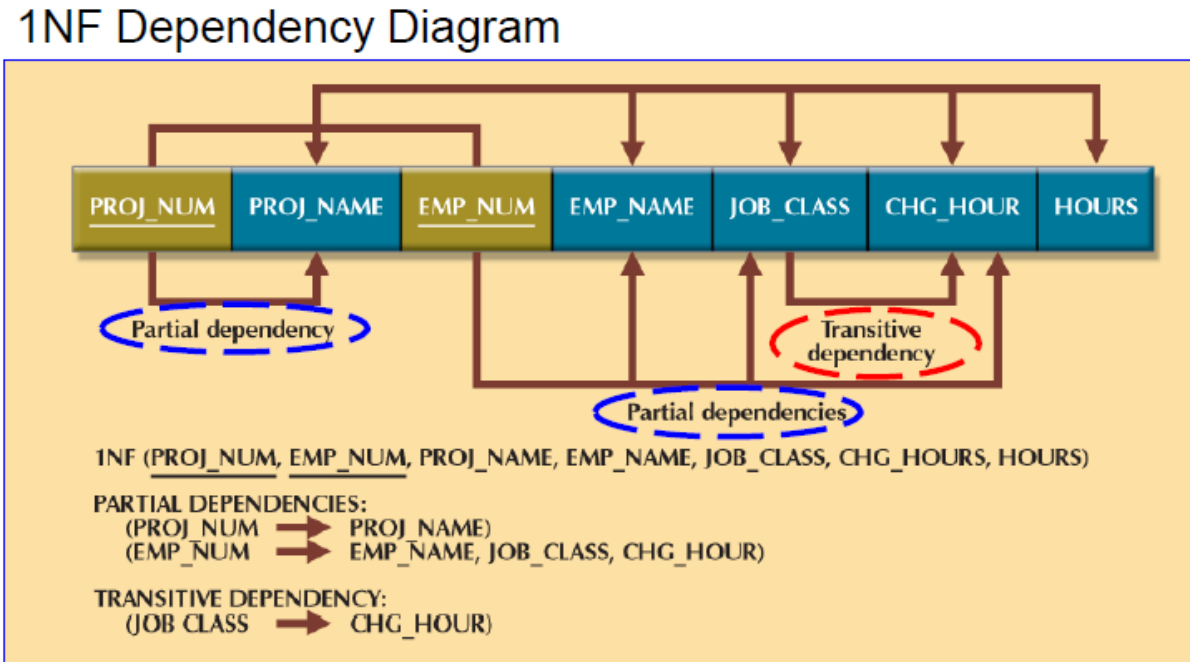
1NF Summary

- All key attributes are defined
- No repeating groups in table
- All attributes are dependent on the primary key(PK)

Functional Dependency

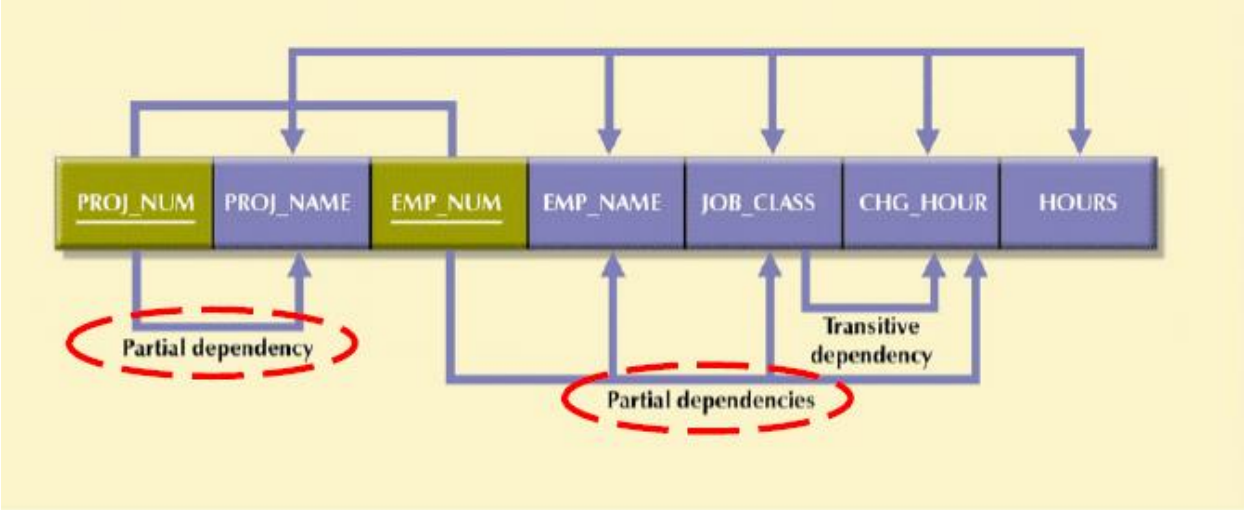
- The base of further normalization
- A particular relationship between two attributes
- **Definition**:
- For any relation R, attribute B is functionally dependent on attribute A if, for every valid instance of A, **the value of A uniquely determines the value of B**, represented as A → B. A is called ***determinant***.
- Example:
  - **Tax file number** → name, address, birthdate
  - **Vehicle identification No.** → make, model, colour
  - **ISBN** → book title

Visualizing dependencies



Second Normal Form (2NF)

What is the problem of 1NF?

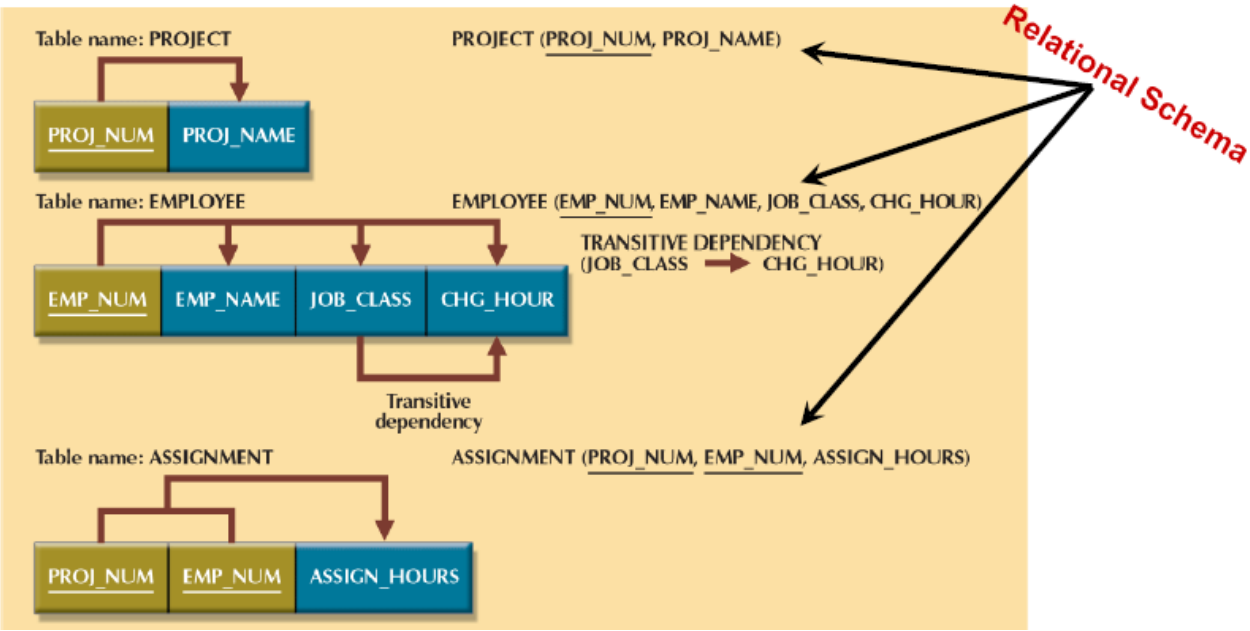


- There should be **NO partial dependencies** in a table
  - Part of the PK should not identify a subset of attributes in the same relation
- Goal of 2NF is to remove partial dependencies
- How?

Conversion to2NF

- To achieve 2NF, you should
  - Identify each **key component** and its corresponding **dependent attributes**
  - Each key component and its attributes form **a new table**
  - **Keep the key components** in the original table
- At this point, most anomalies are removed

A table in 2NF



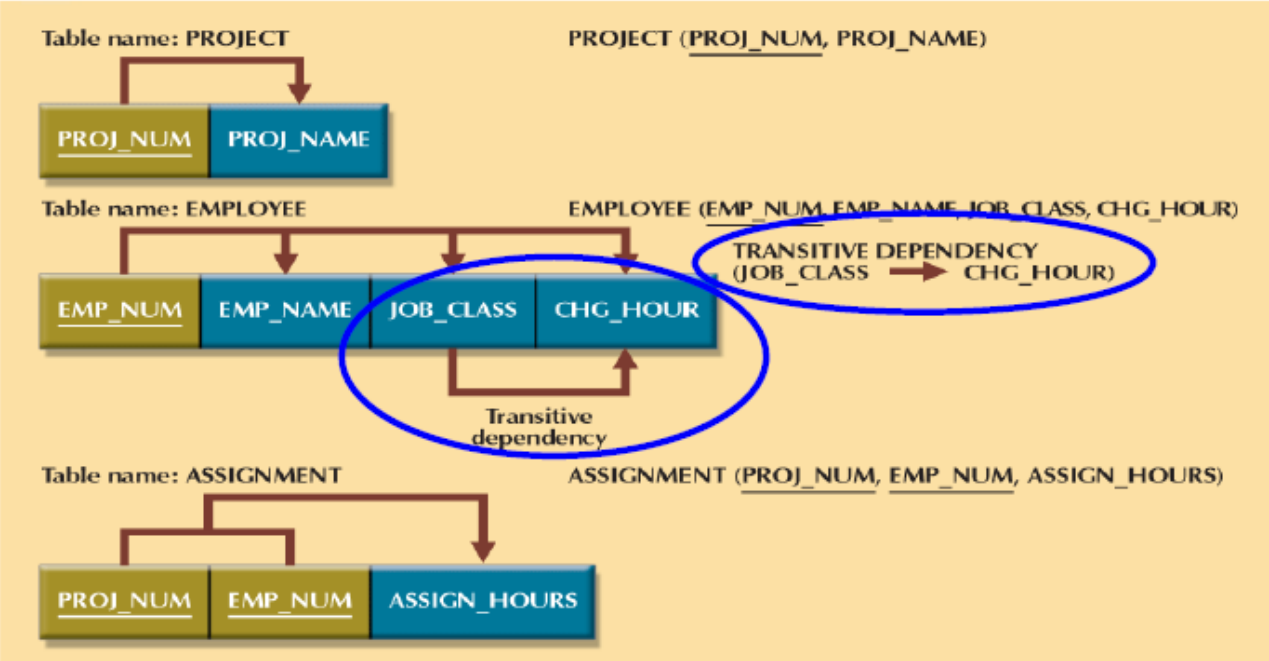
2NFSummary

- Already in1NF
- Includes no partial dependencies
  - No attribute dependent on **a portion** of primary key
- Still possible to exhibit **transitive dependency**
  - **Attributes may be functionally dependent on non-key attributes**

Question: If a table is already in 1NF and its PK contains only one attribute, is it already in2NF?

Third Normal Form (3NF)

What's the problem with 2NF?

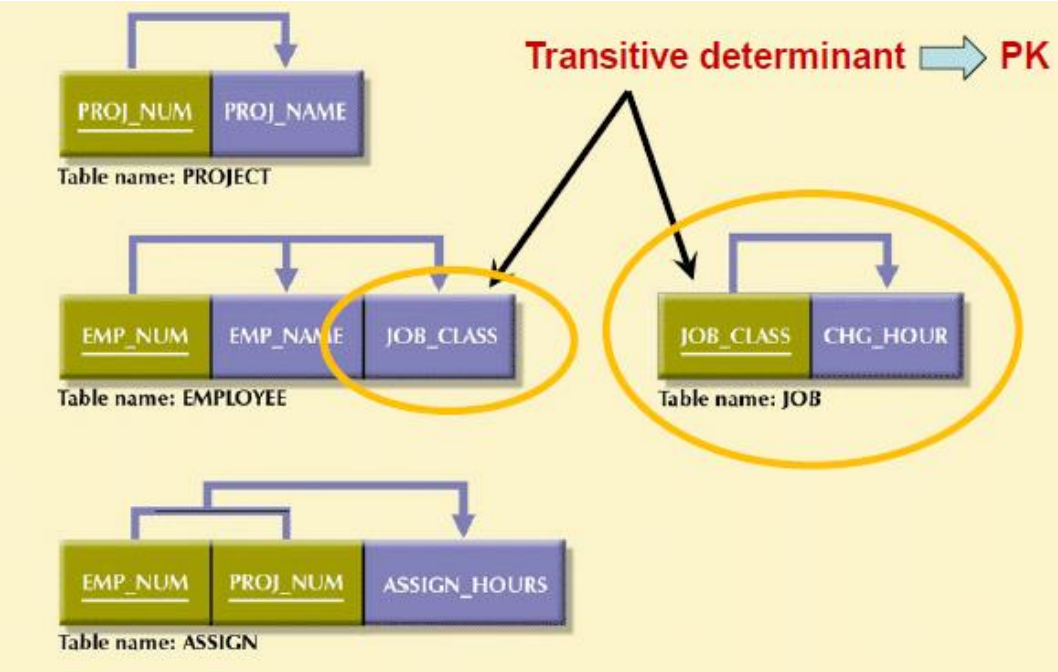


- There should be **NO transitive dependencies** in a table
  - The PK determines attribute **A**, which itself is a key to attribute **B** in the same table
- Goal of 3NF is to remove transitive dependencies
- How?

Conversion to3NF

- To achieve 3NF, you should
  - Identify the **transitive determinant** and its **dependent attributes**
  - Transitive determinant (**new PK**) and its dependent attributes form **a new table**
  - **Remove** the dependent attributes from the original table, and **keep** the transitive determinant in the original table

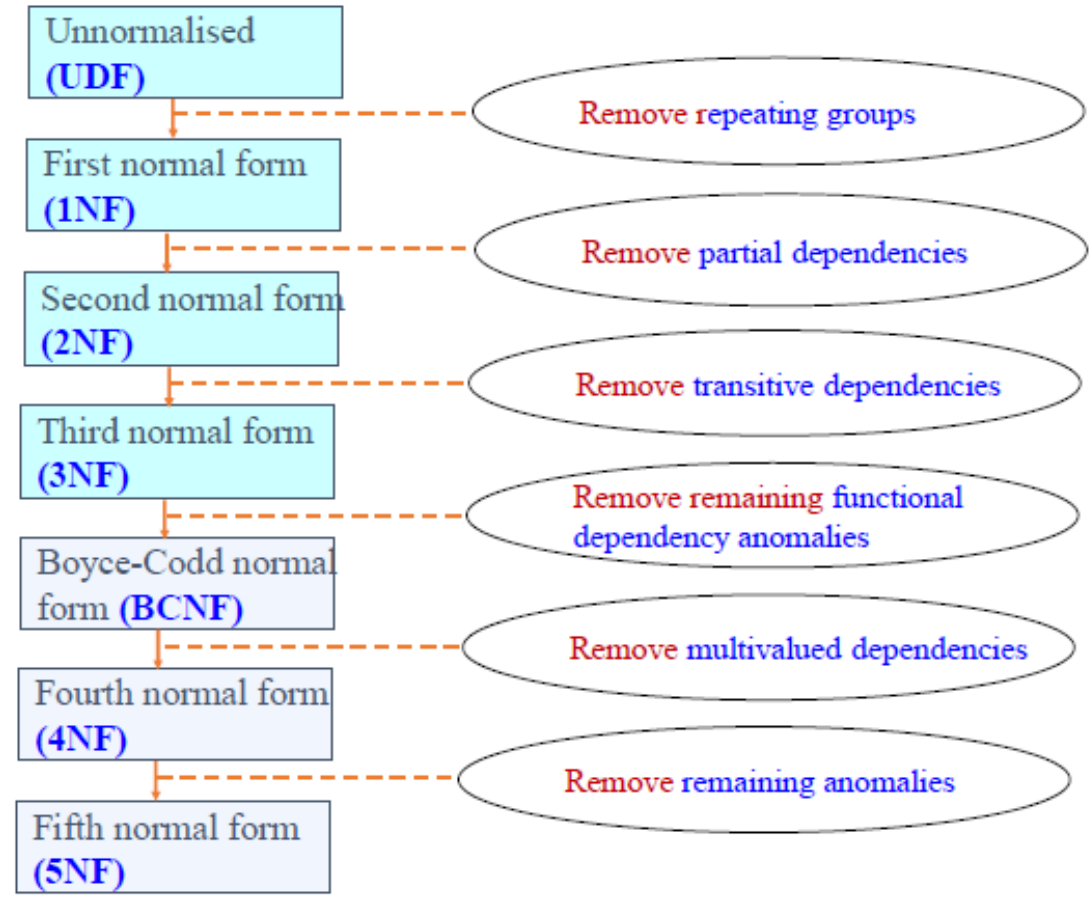
A table in 3NF



Higher Level Normal Form

- Tables in **3NF perform suitably** in business transactional databases
- Higher order normal forms **useful on occasion**
- Boyce Codd normal form (BCNF)
    - A **special case of 3NF**

Stages of Normalization



Reference

Lecture notes - Dr. Dilani Wickramaarachchi (University of Kelaniya)

Teachers Guide (2108)