NORMALIZATION CONCEPT

CHAPTER 3.0 ENTITY RELATIONSHIP MODEL

PREPARED BY

ZHZ SFB NUR (2019)

Normal Form

- Normalization
  - Process of **decomposing** relations with **anomalies** to produce smaller, well structured relation.
- Database Normalization is a technique of organizing the data in the database.
- Normalization is a systematic approach of decomposing tables to eliminate data redundancy(repetition) and undesirable characteristics like **Insertion, Update and Deletion** Anamolies.
- It is a multi-step process that puts data into tabular form, removing duplicated data from the relation tables.

Update Anomalies

- Problems that occur when information is inserted, deleted or updated.
- Three types of update anomalies:
  - Insertion Anomalies
  - Deletion Anomalies
  - Modification Anomalies

Importance of normalization

- Define data correctly and the rest is much easier.
- Improve system performance and accuracy.
- Support integrity and consistency of the data
- Make it easier to expand database later
- Save space, minimize redundancy and eliminate anomalies.

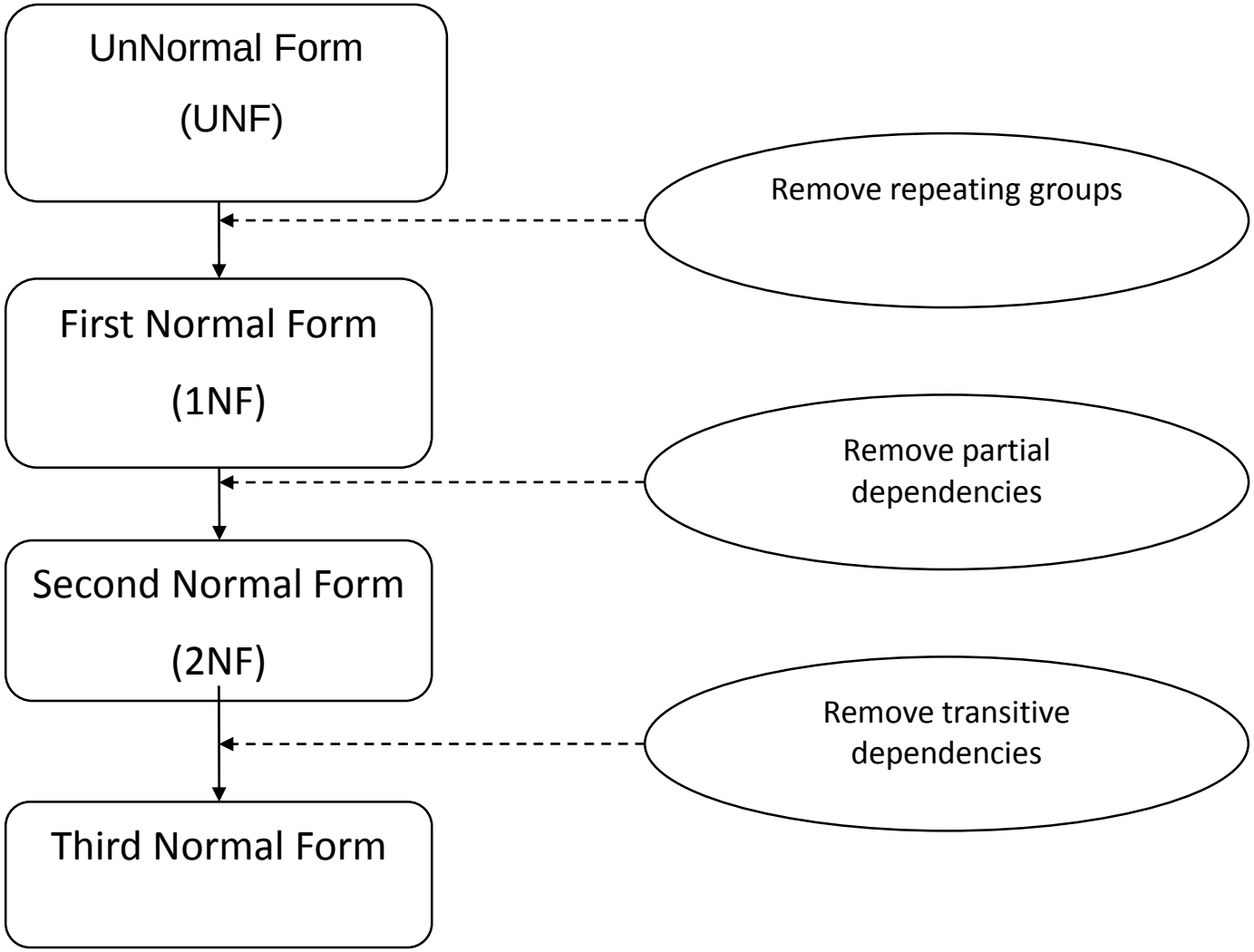
Stud_Id	Stud_Name	Stud_No	Major	Cour_Id	Cour_Title	Lect_Name	Lect_Room	Grade
333	Ali	1133	CSC	SA100	System Analysis	Rokiah	E1	B
				PR200	Bus Org	Hamid	E2	A
				IS100	Database	Zakiy	E3	C
777	Ahmad	9797	SUR	SUR10	Survey	Abdullah	E5	C
				IS100	Database	Zakiy	E3	A

Steps in Normalization

The most commonly used normal forms:

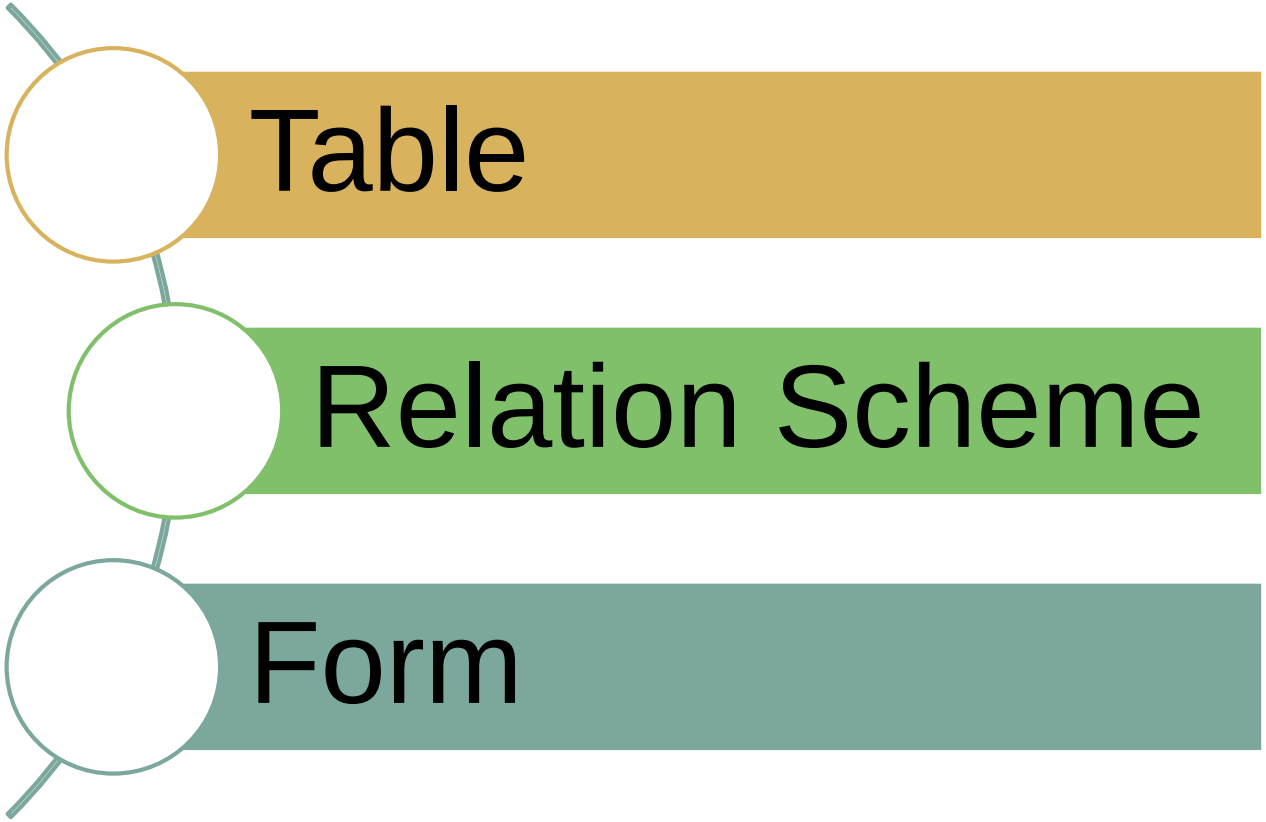
1. First normal form(1NF)
2. Second normal form(2NF)
3. Third normal form(3NF)
4. Boyce & Codd normal form (BCNF)

Normalization



Normalization

- Normalize the table until **3rd Normal Form** (*min*)



GRADE

Stud_Id	Stud_Name	Stud_No	Major	Cour_Id	Cour_Title	Lect_Name	Lect_Room	Grade
666	Sofia	1434	BU	BU100	Marketing	Roslan	E1	B
				BU200	Bus Org	Hasnah	E2	A
				IS100	Database	Zami	E3	C
777	Suria	6567	IS	IS200	Info Sys	Muthu	E4	B
				IS100	Database	Zami	E3	B
999	Zahid	9779	GIS	CSC10	Computer	Abdul	E5	C
				IS100	Database	Zami	E3	A

FAKULTI TEKNOLOGI MAKLUMAT DAN SAINS KUANTITATIF
UiTM SHAH ALAM
LAPORAN AGIHAN TUGAS SYARAHAN (ATS) PENSYARAH
SEMESTER NOV 2004 – MAC 2005

Nama: Musalmah Abd Rahman
No. Pekerja: 123654
Kod Jabatan: SS
Jabatan: Jabatan Sains Sistem

Kod kursus	Nama kursus	Kod program	Kumpulan		Beban ATS(jam)
ITS250	Pengenalan Pang. Data	CS110	A		4
ITS250	Pengenalan Pang. Data	CS111	A		4
CSC324	Pengurusan Pang. Data	AP121	A		6
				Jumlah	14

Nama: Norlin Mustafa
No. Pekerja: 456345
Kod Jabatan: CTN
Jabatan: Jabatan Rangkaian Sistem

Kod kursus	Nama kursus	Kod program	Kumpulan		Beban ATS(jam)
ITC100	Pengenalan Komputer	CS110	A		4
ITT300	Pengenalan Pang. Data	CS111	A		4
ITC260	Struktur Data	CS110	B		3
				Jumlah	11

Basic Normal Form

- Scenario :

  A company obtains parts from a number of suppliers. Each supplier is located in one city. A city can have more than one supplier located there and each city has a status code associated with it. Each supplier may provide many parts. The company creates a simple relational table to store this information that can be expressed in relational notation as:

  SUPPLIER-PART(s_id, status, city, p_id, qty)

Unnormalize Form

- A table that contains one or more repeating groups.

SUPPLIER(s_id, status, city, p_id, qty)

s_id	status	city	part_id	quantity
S1	20	London	P1	300
			P2	200
			P3	400
			P4	200
			P5	100
			P6	100
S2	10	Paris	P1	300
			P2	400
s3	10	Paris	P2	200
S4	20	London	P2	200
			P4	300
			P5	500

Repeating groups @ multivalued attributes

Unnormalize Form

- A table that contains one or more repeating groups.

SUPPLIER(s_id, status, city, p_id, qty)

s_id	status	city	part_id	quantity
S1	20	London	P1	300
S1	20	London	P2	200
S1	20	London	P3	400
S1	20	London	P4	200
S1	20	London	P5	100
S1	20	London	P6	100
S2	10	Paris	P1	300
S2	10	Paris	P2	400
S3	10	Paris	P2	200
S4	20	London	P2	200
S4	20	London	P4	300
S4	20	London	P5	500

Repeating groups @ multivalued attributes

First Normal Form

- Def : A relation in which the intersection of each row and column contains one and only one value.

s_id	status	city	part_id	quantity
S1	20	London	P1	300
S1	20	London	P2	200
S1	20	London	P3	400
S1	20	London	P4	200
S1	20	London	P5	100
S1	20	London	P6	100
S2	10	Paris	P1	300
S2	10	Paris	P2	400
S3	10	Paris	P2	200
S4	20	London	P2	200
S4	20	London	P4	300
S4	20	London	P5	500

SUPPLIER
(s_id, p_id ,status, city,qty)

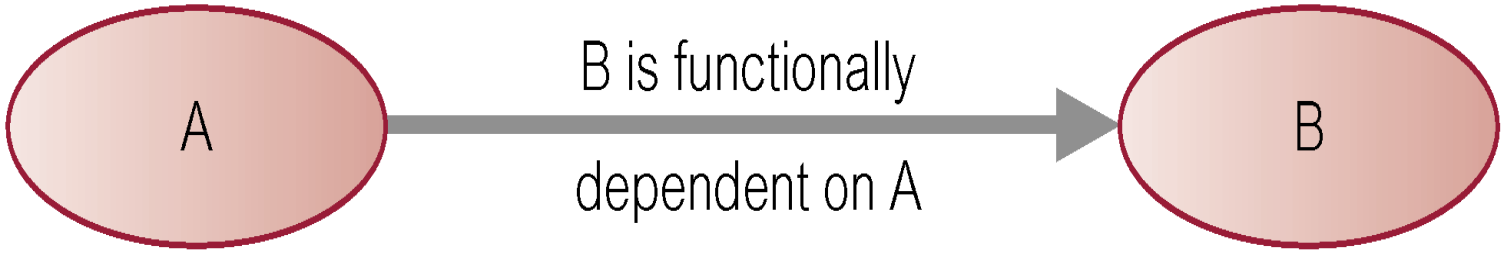
First Normal Form (1NF)

- First normal form still contains redundant data.
- Redundancy causes problem called update anomalies.
- As per the rule of first normal form, an attribute (column) of a table cannot hold multiple values.
- It should hold only atomic values.

Functional Dependencies

- Important concept associated with normalization.
- **Functional dependency describes relationship between attributes.**
- For example, if A and B are attributes of relation R, B is functionally dependent on A (**A → B**), if each value of A in R is associated with exactly one value of B in R.
- Sometimes called A functionally determines B

Functional Dependencies



- Determinant
  - the attribute or group of attributes on the left-hand side of the arrow of a functional dependency.

Second Normal Form (2NF)

- Def : A relation is in first normal form and every non-candidate key attribute is fully functionally dependent on any candidate key.
- A table that is in first normal form and every non-primary key attribute is fully dependent on the primary key.
- Involve in removing partial dependencies.

- **For a table to be in the Second Normal Form,**
  - **It should be in the First Normal form (1NF).**
  - **It should not have <u>Partial Dependency</u>.**

1NF to 2NF

1NF

Supplier (s_id, part_id, city, status, quantity

FD

s_id, part_id → city, status, quantiy

2NF

FD

s_id → city, status

s_id, part_id → quantiy

Supplier (s_id, city, status)

Part (s_id, part_id, quantity)

s_id	status	city	part_id	quantity
S1	20	London	P1	300
S1	20	London	P2	200
S1	20	London	P3	400
S1	20	London	P4	200
S1	20	London	P5	100
S1	20	London	P6	100
S2	10	Paris	P1	300
S2	10	Paris	P2	400
S3	10	Paris	P2	200
S4	20	London	P2	200
S4	20	London	P4	300
S4	20	London	P5	500

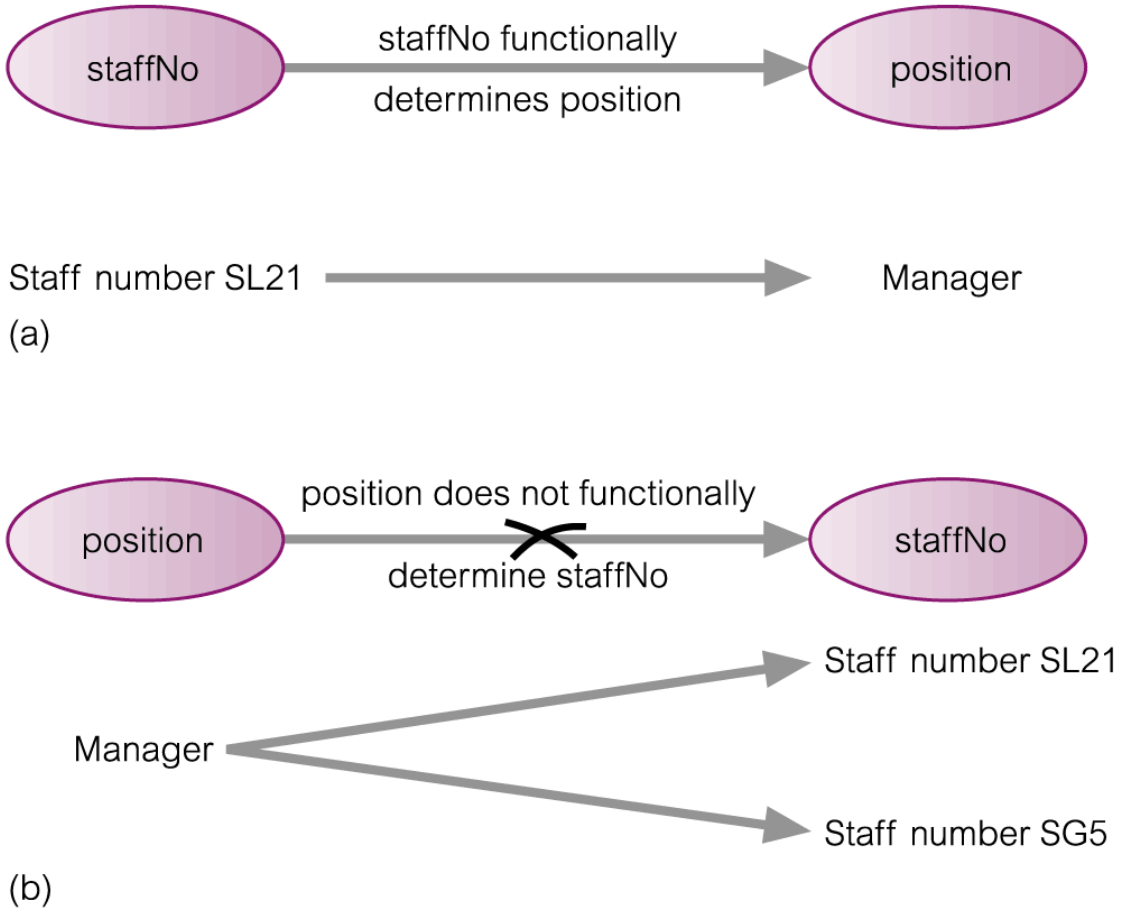
Second Normal Form

s_id	part_id	quantity
S1	P1	300
S1	P2	200
S1	P3	400
S1	P4	200
S1	P5	100
S1	P6	100
S2	P1	300
S2	P2	400
S3	P2	200
S4	P2	200
S4	P4	300
S4	P5	500

s_id	status	city
S1	20	London
S2	10	Paris
S3	10	Paris
S4	20	London

Supplier (s_id, city, status)
Part (s_id, part_id, quantity)

Example of Functional Dependencies



Example

Table : EMP_COURSE

Emp_ID	Course_Title	Date_Completed
100	SPSS	6/19/200X
100	Surveys	10/7/200X
140	Tax Acc	12/8/200X
110	SPSS	1/12/200X
110	C++	4/22/200X
150	SPSS	6/19/200X
150	Java	8/12/200X

Functional Dependency:

Emp_ID, Course_Title → Date_Completed.

Functional Dependency

- Full Functional Dependency
  - Indicates that if A and B are attributes of a relation, B is fully functionally dependent on A, but not on any proper subset of A.

Full Functional Dependency

staffNo	sName	position	salary	branchNo	bAddress
100	Syazana	Manager	30000	B005	Penang
200	Farah	Assistant	12000	B003	Kelantan
300	Zilla	Supervisor	18000	B003	Kelantan
400	Malathy	Assistant	9000	B007	Seremban
500	Sanghita	Manager	24000	B003	Kelantan
600	Syafiq	Assistant	9000	B005	Penang

Table : StaffBranch

FD : staffNo → sName, position, salary, brachNo, bAdress

The relation is not in full dependency because bAddress is fuctionally dependent on branchNo

staffNo → sName, position, salary, brachNo
brachNo → bAddress

Full Functional Dependency

staffNo	sName	position	salary	branchNo
100	Syazana	Manager	30000	B005
200	Farah	Assistant	12000	B003
300	Zilla	Supervisor	18000	B003
400	Malathy	Assistant	9000	B007
500	Sanghita	Manager	24000	B003
600	Syafiq	Assistant	9000	B005

branchNo	bAddress
B005	Penang
B003	Kelantan
B007	Seremban

Figure 1 :
Staff and Branch relations