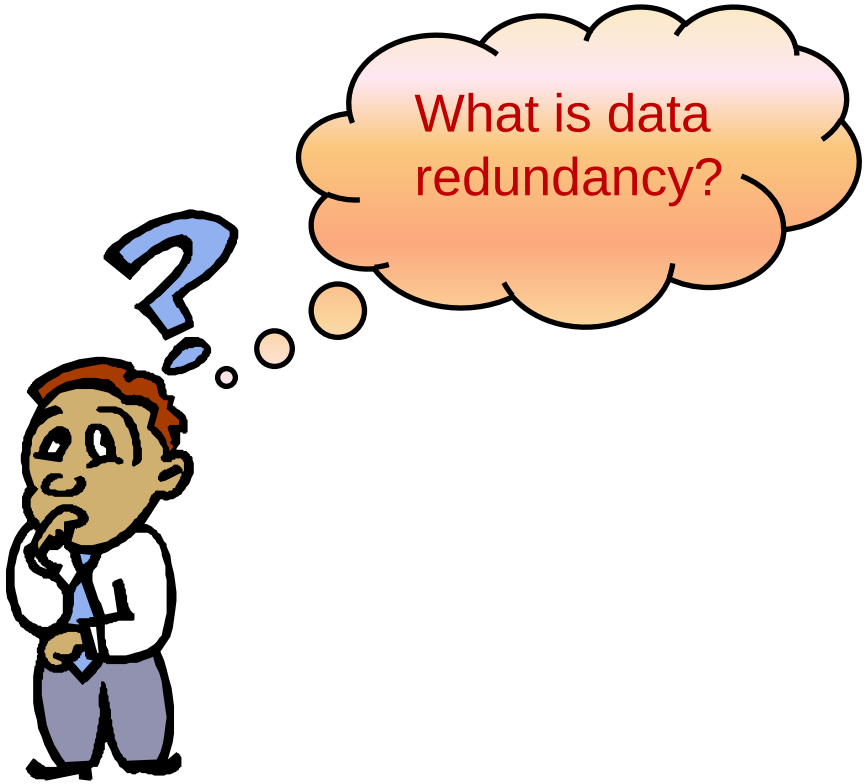
Objectives

- In this session, you will learn to:
  - Describe data redundancy

Understanding Data Redundancy



Understanding Data Redundancy (Contd.)

Understanding Data Redundancy (Contd.)

- Redundancy:
  - Increases the time involved in updating, adding, and deleting data.
  - Increases the utilization of disk space and hence, disk I/O increases.
- Redundancy can, therefore, lead to:
  - Insertion, modification, and deletion of data, which may cause inconsistencies.
  - Errors, which are more likely to occur when facts are repeated.
  - Unnecessary utilization of extra disk space.

Understanding Data Redundancy (Contd.)

- Consider the STUDENT table, as shown in the following diagram.

STUDENT
STUDENTID
STUDENTNAME
STUDENTBIRTHDATE
STUDENTADDRESS
STUDENTCITY
STUDENTZIP
STUDENTCLASS
STUDENTSEMESTER
STUDENTTEST1
STUDENTTEST2

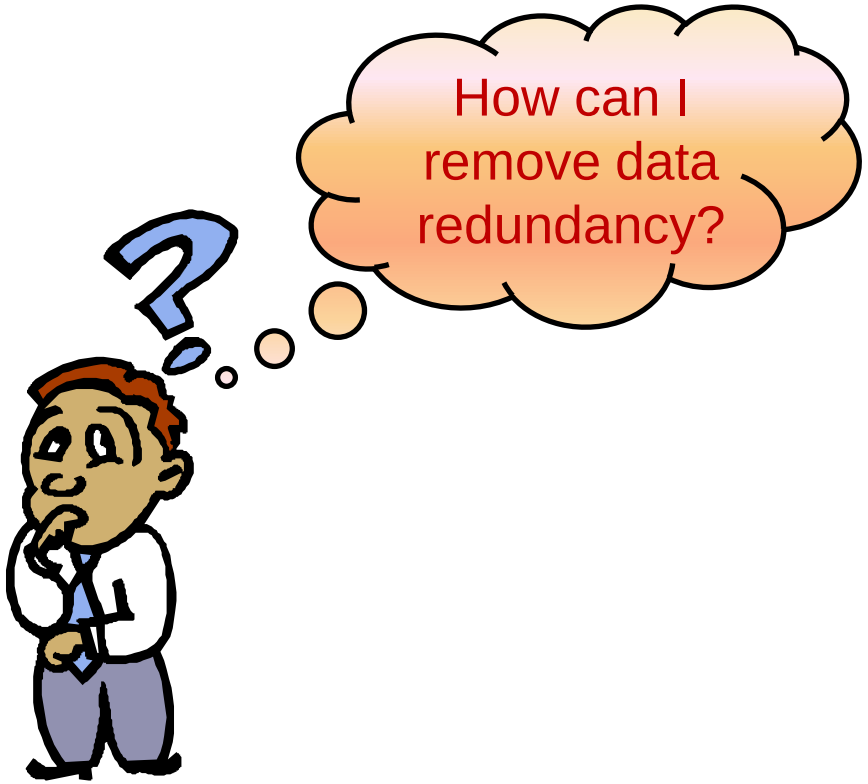
Understanding Data Redundancy (Contd.)

- The STUDENT table contains the values for each attribute, as shown in the following diagram.

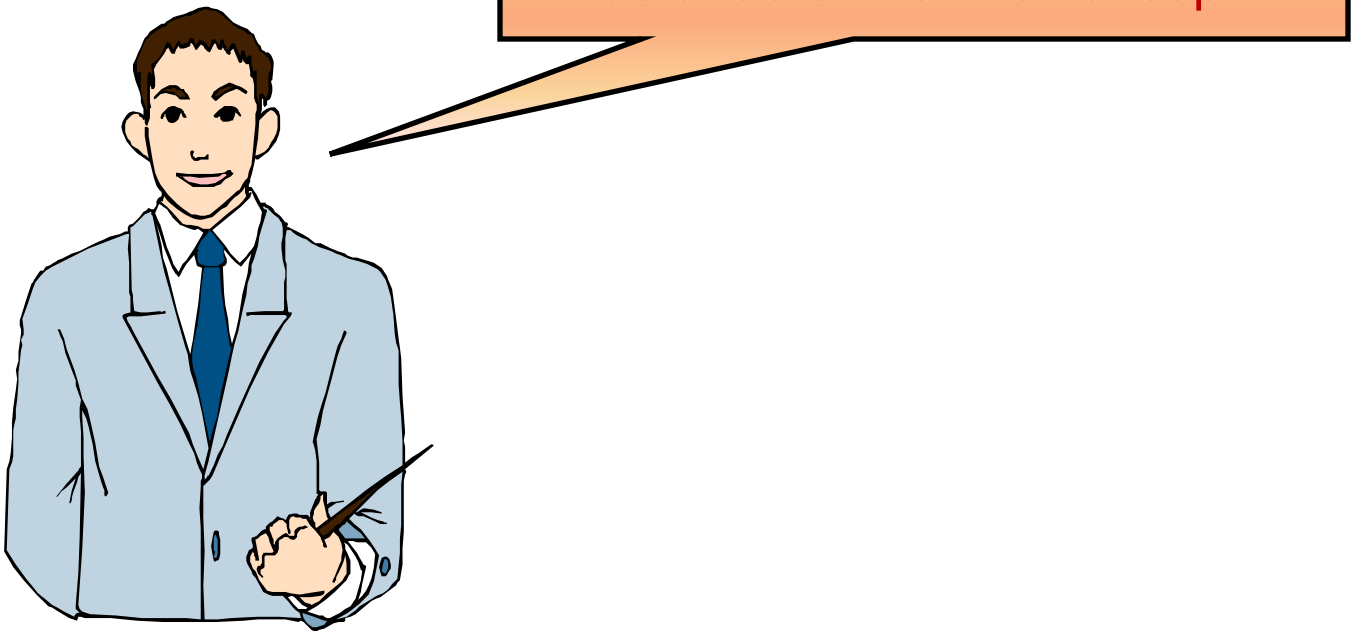
STUDENTID	STUDENTNAME	.....	STUDENTSEMESTER	STUDENTTEST1	STUDENTTEST2
001	Mary	.....	SEM-1	40	65
001	Mary	.....	SEM-2	56	48
002	Jake	.....	SEM-1	93	84
002	Jake	.....	SEM-2	85	90

The details of the students, such as STUDENTID and STUDENTNAME are repeated while recording marks of different semesters.

Definition of Normalization



Definition of Normalization (Contd.)



Definition of Normalization (Contd.)

- Normalization:
  - Is a method of breaking down complex table structures into simple table structures by using certain rules.
  - Has the following benefits:
    - It helps in maintaining data integrity.
    - It helps in simplifying the structure of tables, therefore, making a database more compact.
    - It helps in reducing the null values, which reduces the complexity of data operations.

Definition of Normalization (Contd.)

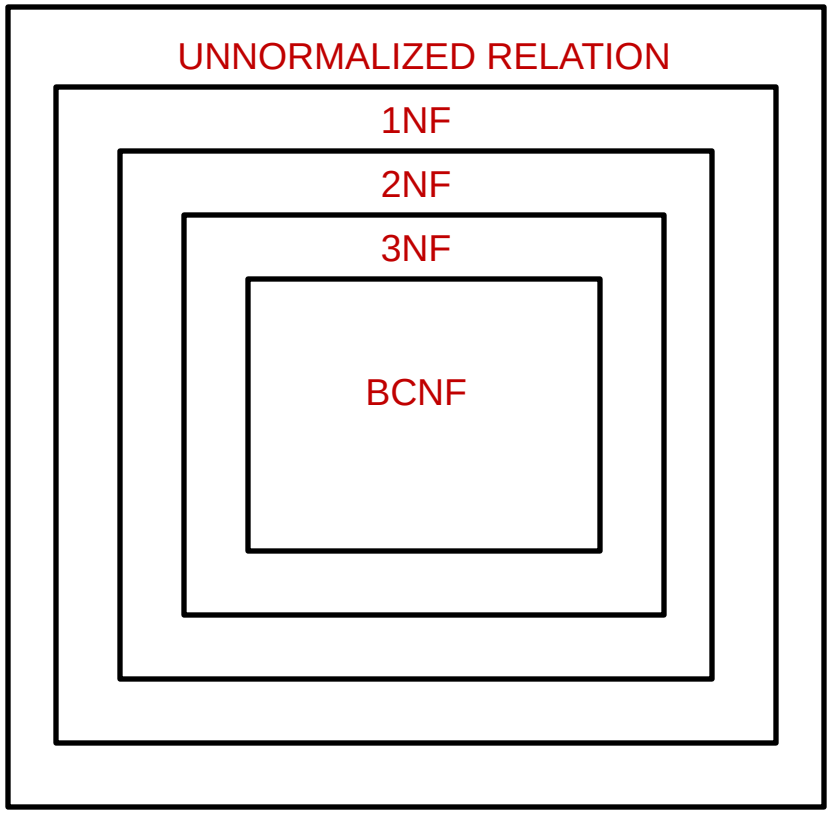
- Some rules that should be followed to achieve a good database design are:
  - Each table should have an identifier.
  - Each table should store data for a single type of entity.
  - Columns that accept NULLs should be avoided.
  - The repetition of values or columns should be avoided.

Definition of Normalization (Contd.)

- Normalization results in the formation of tables that satisfy certain normal forms.
- The normal forms are used to remove various types of abnormalities and inconsistencies from the database.
- The most important and widely used normal forms are:
  - First Normal Form (1NF)
  - Second Normal Form (2NF)
  - Third Normal Form (3NF)
  - Boyce-Codd Normal Form (BCNF)

Definition of Normalization (Contd.)

- The following diagram shows the different levels of normalization.



Definition of Normalization (Contd.)

- First Normal Form (1NF):
  - A table is said to be in 1NF when each cell of the table contains precisely one value.
  - The guidelines for converting a table into 1NF are:
    - Place the related data values in a table. Further, define similar data values with the column name.
    - There should be no repeating group in the table.
    - Every table must have a unique primary key.

Definition of Normalization (Contd.)

- Consider the PROJECT table, as shown in the following diagram.

Primary key

PROJECT table is not in first normal form because cells in PROJCODE and HOURS have more than one value.

ECODE	DEPT	DEPTHEAD	PROJCODE	HOURS
E101	Systems	E901	P27 P51 P20	90 101 60
E305	Sales	E906	P27 P22	109 98
E508	Admin	E908	P51 P27	NULL 72

Definition of Normalization (Contd.)

- By applying the 1NF definition to the PROJECT table, you arrive at the table, as shown in the following diagram.

ECODE	DEPT	DEPTHEAD	PROJCODE	HOURS
E101	Systems	E901	P27	90
E101	Systems	E901	P51	101
E101	Systems	E901	P20	60
E305	Sales	E906	P27	109
E305	Sales	E906	P22	98
E508	Admin	E908	P51	NULL

Definition of Normalization (Contd.)

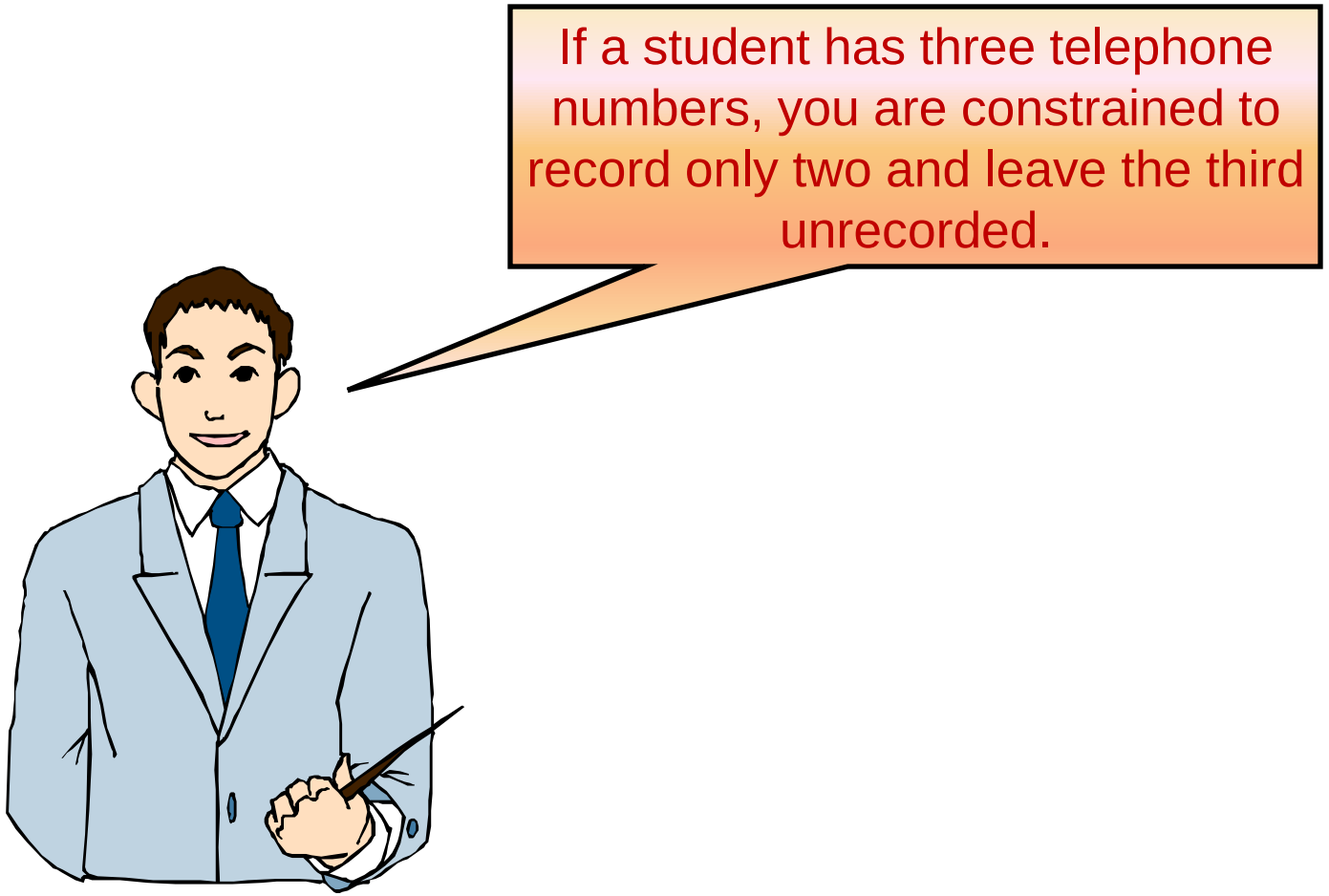
- Consider the STUDENT table, as shown in the following diagram.

Primary key

STUDENT table contains null values in PHONE NUMBER2 column, and the number of telephone numbers per student is restricted to two.

ROLL NUMBER	STUDENT FIRST NAME	STUDENT LAST NAME	PHONE NUMBER1	PHONE NUMBER2
R01	Luke	Thomas	234456	
R02	Rita	Wilson	245688	276453
R03	Tom	Gellar	256487	
R04	Jack	Miles	234789	2341543

Definition of Normalization (Contd.)



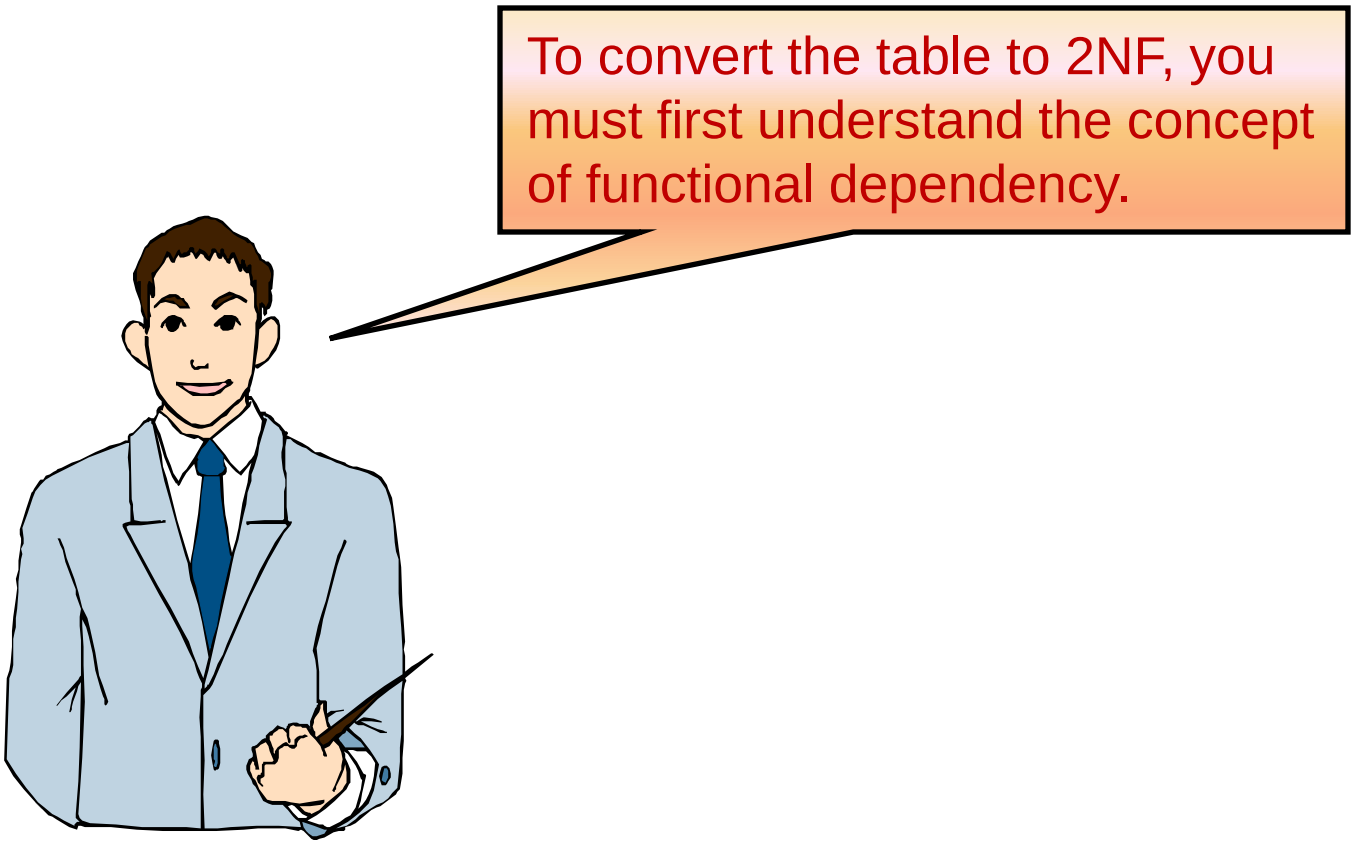
Definition of Normalization (Contd.)

- By applying the 1NF definition to the STUDENT table, you can arrive at the tables, as shown in the following diagram.

ROLL NUMBER	STUDENT FIRST NAME	STUDENT LAST NAME
R01	Luke	Thomas
R02	Rita	Wilson
R03	Tom	Gellar
R04	Jack	Miles

ROLL NUMBER	PHONE NUMBER
R01	234456
RO2	245688
R02	276453
R03	256487
R04	234789
R04	2341543

Definition of Normalization (Contd.)



Definition of Normalization (Contd.)

- Functional dependency:
  - Attribute A is functionally dependent on B if and only if, for each value of B, there is exactly one value of A.
  - Attribute B is called the determinant.

Definition of Normalization (Contd.)

- Consider the REPORT table, as shown in the following diagram.

Primary key

ROLL NUMBER	COURSE _CODE	COURSE NAME	T_NAME	ROOM_NUMBER	MARKS	GRADE
R001	C100	Java	James	301	88	A
R002	C101	C#.NET	Peter	302	75	B
R003	C101	C#.NET	Peter	302	60	C
R004	C100	Java	James	301	72	B

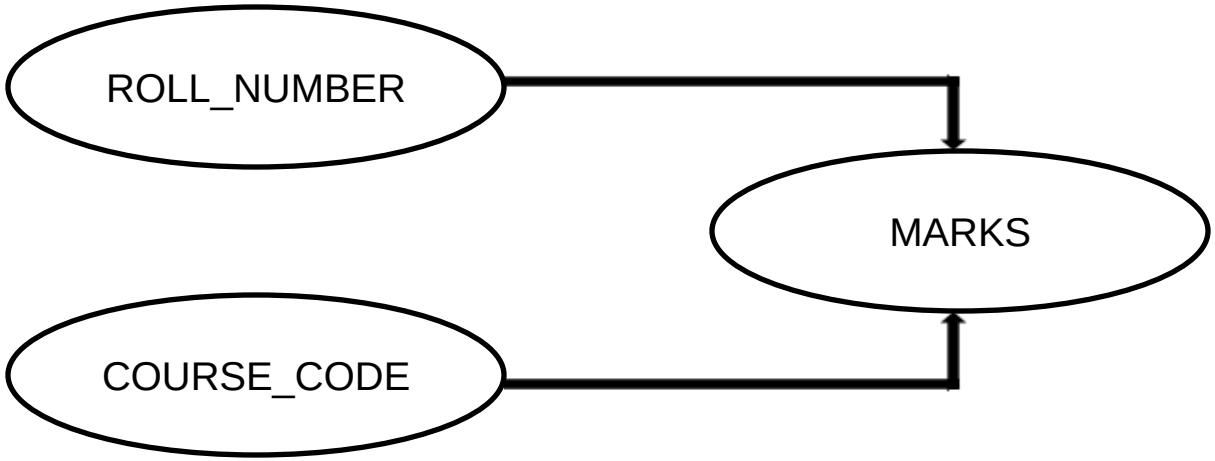
Definition of Normalization (Contd.)

- In the REPORT table:
  - For a particular value of ROLL_NUMBER+COURSE_CODE, there is precisely one corresponding value for MARKS.
  - Hence, MARKS is functionally dependent on ROLL_NUMBER+COURSE_CODE.
  - This can be symbolically represented as:

    (ROLL_NUMBER, COURSE_CODE)→MARKS.

Definition of Normalization (Contd.)

- The following diagram shows the functional dependency between MARKS and ROLL_NUMBER+COURSE_CODE.

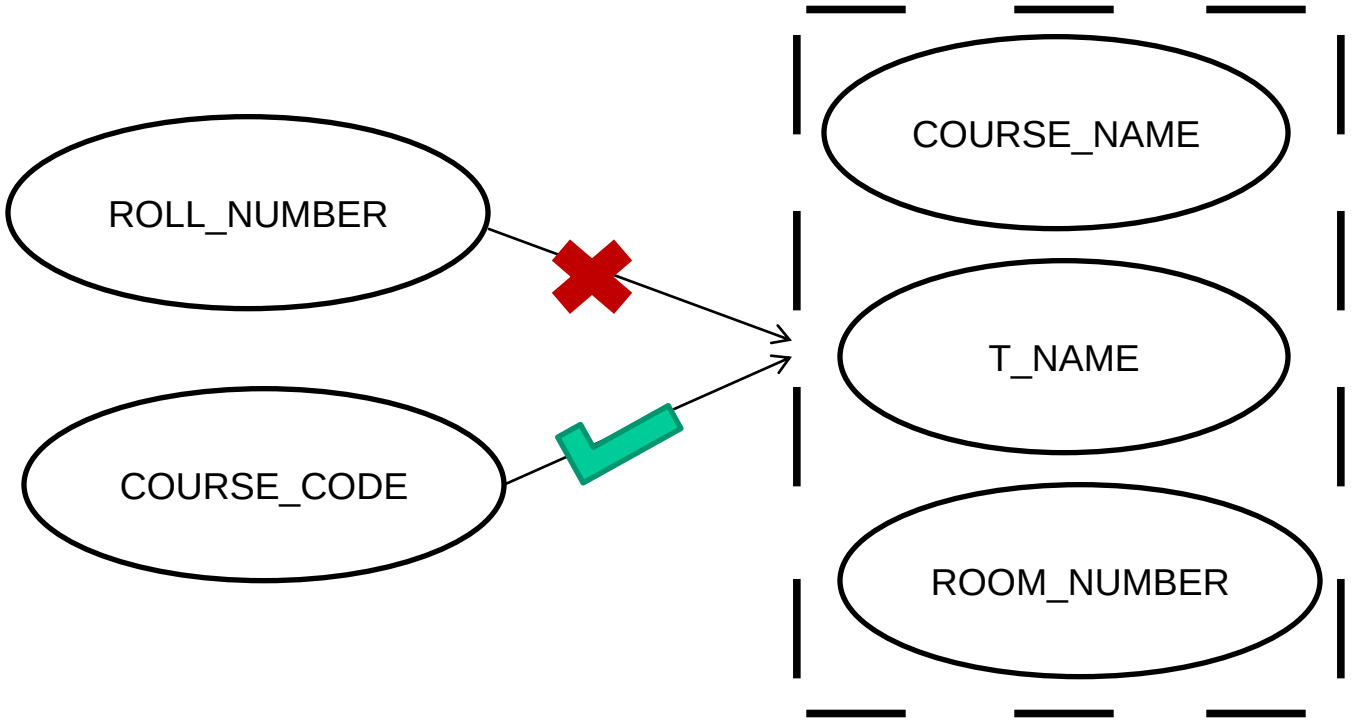


Definition of Normalization (Contd.)

- The other functional dependencies in the REPORT table are:
  - COURSE_CODE→COURSE_NAME
  - COURSE_CODE→T_NAME (Assuming one course is taught by only one teacher.)
  - T_NAME→ROOM_NUMBER (Assuming each teacher has his/her own, unshared room.)
  - MARKS→GRADE

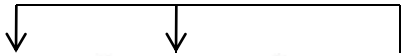
Definition of Normalization (Contd.)

- COURSE_NAME, T_NAME, and ROOM_NUMBER attributes are partially dependent on the whole key.
- This dependency is called partial dependency, as shown in the following diagram.



Definition of Normalization (Contd.)

- ROOM_NUMBER is dependent on T_NAME, and T_NAME is dependent on COURSE_CODE, as shown in the following diagram.

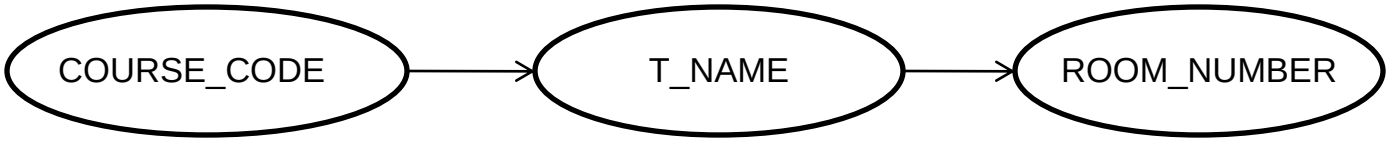


ROLL NUMBER	COURSE _CODE	COURSE NAME	T_NAME	ROOM_NUMBER	MARKS	GRADE
R001	C100	Java	James	301	88	A
R002	C101	C#.NET	Peter	302	75	B
R003	C101	C#.NET	Peter	302	60	C
R004	C100	Java	James	301	72	B

Definition of Normalization (Contd.)

Definition of Normalization (Contd.)

- The following diagram shows the transitive dependency.



Summary

- In this session, you learned that:
  - Normalization is used to simplify table structures.
  - Normalization results in the formation of tables that satisfy certain specified constraints, and represent certain normal forms. The normal forms are used to ensure that various types of abnormalities and inconsistencies are not introduced in the database. A table structure is always in a certain normal form.
  - The most important and widely used normal forms are:
    - First Normal Form (1NF)
    - Second Normal Form (2NF)
    - Third Normal Form (3NF)
    - Boyce-Codd Normal Form (BCNF)
  - A table is said to be in 1NF when each cell of the table contains precisely one value.

Summary (Contd.)

- The following dependencies are found in normalization:
  - Functional
  - Partial
  - Transitive