Objectives

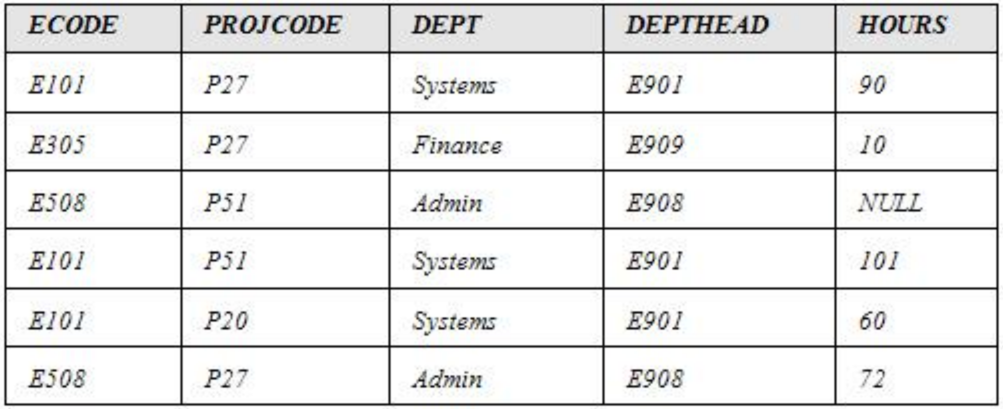
- In this session, you will learn to:
  - Describe data redundancy

Definition of Normalization

- Second Normal Form (2NF):
  - A table is said to be in 2NF when:
    - It is in the 1NF, and
    - No partial dependency exists between non-key attributes and key attributes.
  - The guidelines for converting a table into 2NF are:
    - Find and remove attributes that are functionally dependent on only a part of the key and not on the whole key. Place them in a different table.
    - Group the remaining attributes.

Definition of Normalization (Contd.)

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

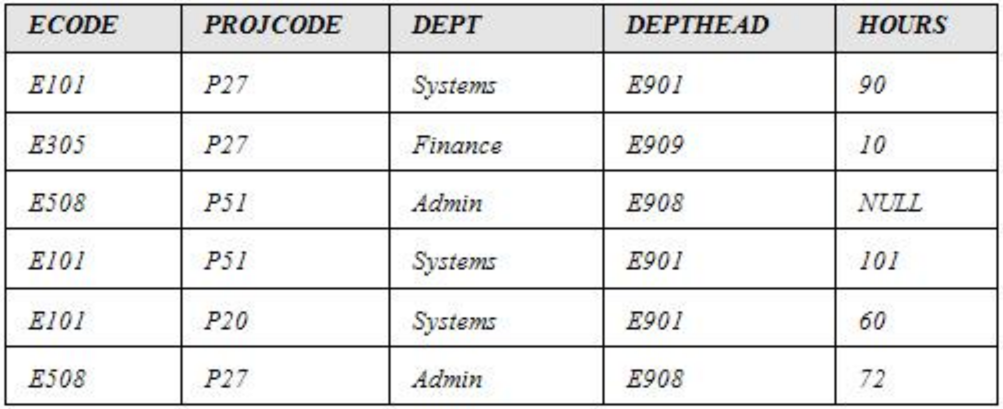


ECODE	PROJCODE	DEPT	DEPTHEAD	HOURS
E101	P27	Systems	E901	90
E305	P27	Finance	E909	10
E508	P51	Admin	E908	NULL
E101	P51	Systems	E901	101
E101	P20	Systems	E901	60
E508	P27	Admin	E908	72

- The preceding table could lead to the following problems:
  - Insertion
  - Updation
  - Deletion

Definition of Normalization (Contd.)

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



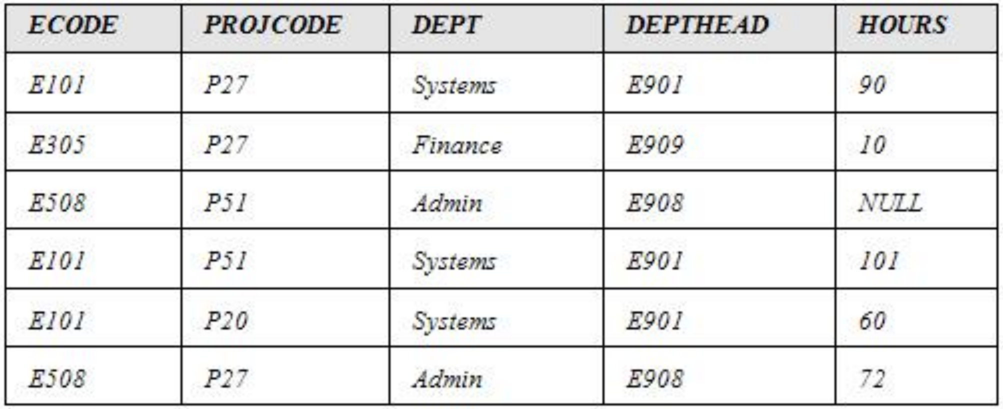
ECODE	PROJCODE	DEPT	DEPTHEAD	HOURS
E101	P27	Systems	E901	90
E305	P27	Finance	E909	10
E508	P51	Admin	E908	NULL
E101	P51	Systems	E901	101
E101	P20	Systems	E901	60
E508	P27	Admin	E908	72

- The preceding table could lead to the following problems:
  - Insertion
  - Updation
  - Deletion

The department of an employee cannot be recorded until the employee is assigned a project.

Definition of Normalization (Contd.)

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



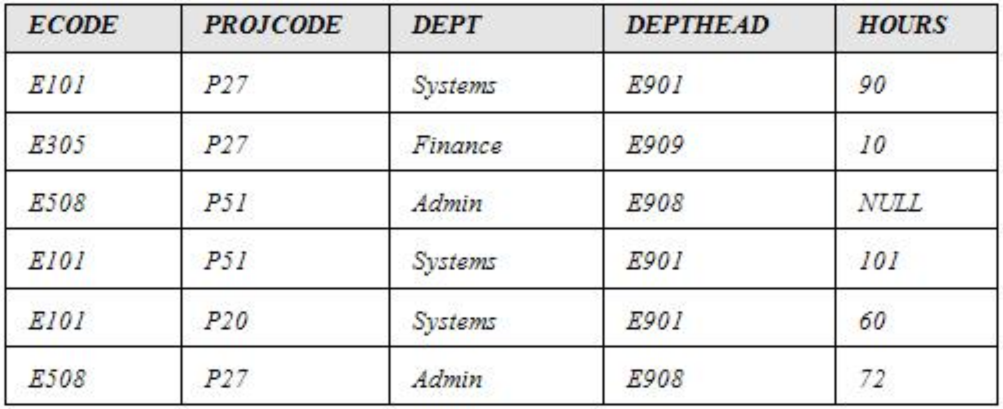
ECODE	PROJCODE	DEPT	DEPTHEAD	HOURS
E101	P27	Systems	E901	90
E305	P27	Finance	E909	10
E508	P51	Admin	E908	NULL
E101	P51	Systems	E901	101
E101	P20	Systems	E901	60
E508	P27	Admin	E908	72

- The preceding table could lead to the following problems:
  - Insertion
  - Updation
  - Deletion

For an employee, ECODE, DEPT, and DEPTHEAD are repeated. Any change will have to be recorded in every row of the EMPLOYEE table.

Definition of Normalization (Contd.)

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

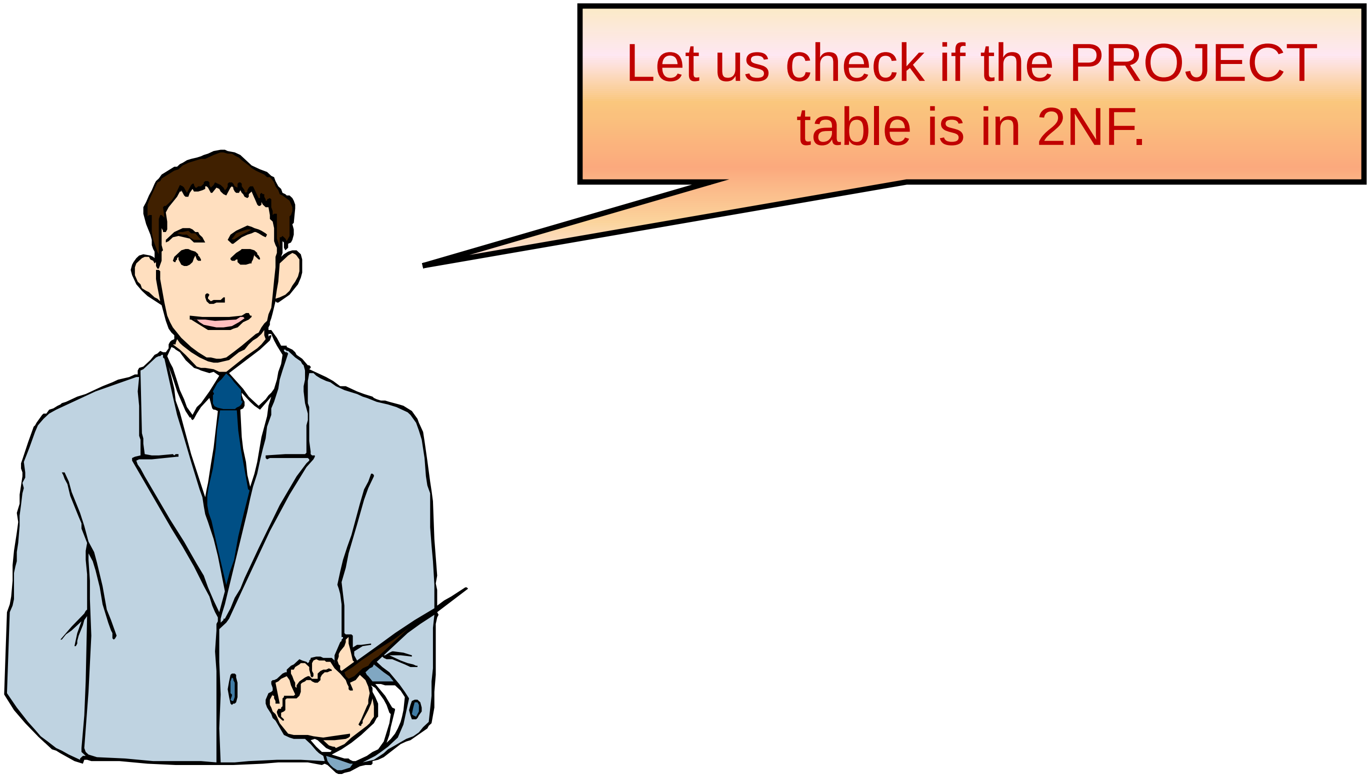


ECODE	PROJCODE	DEPT	DEPTHEAD	HOURS
E101	P27	Systems	E901	90
E305	P27	Finance	E909	10
E508	P51	Admin	E908	NULL
E101	P51	Systems	E901	101
E101	P20	Systems	E901	60
E508	P27	Admin	E908	72

- The preceding table could lead to the following problems:
  - Insertion
  - Updation
  - Deletion

When the project finishes, the employee details are deleted. This leads to loss in information about the department to which employee belongs.

Definition of Normalization (Contd.)

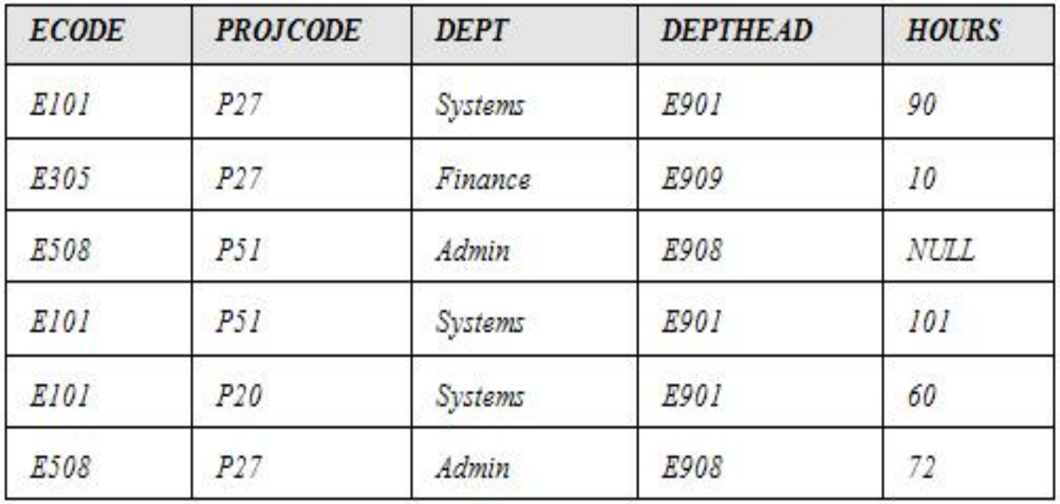


Let us check if the PROJECT table is in 2NF.

Definition of Normalization (Contd.)

Composite key

Hours is functionally dependent on the whole key, ECODE+PROJCODE.

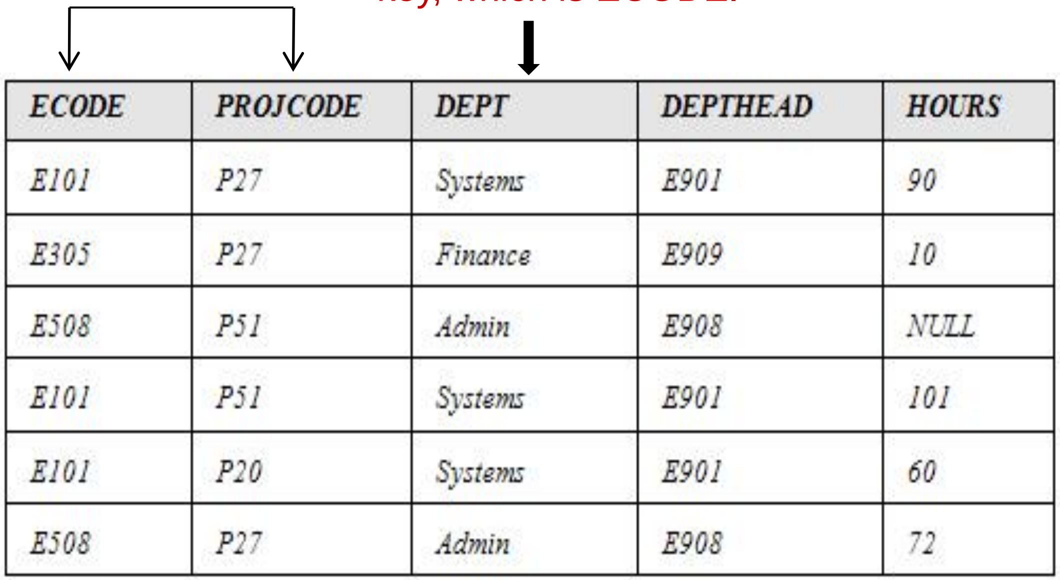


ECODE	PROJCODE	DEPT	DEPTHEAD	HOURS
E101	P27	Systems	E901	90
E305	P27	Finance	E909	10
E508	P51	Admin	E908	NULL
E101	P51	Systems	E901	101
E101	P20	Systems	E901	60
E508	P27	Admin	E908	72

Definition of Normalization (Contd.)

Composite Key

Dept is functionally dependent on part of the key, which is ECODE.

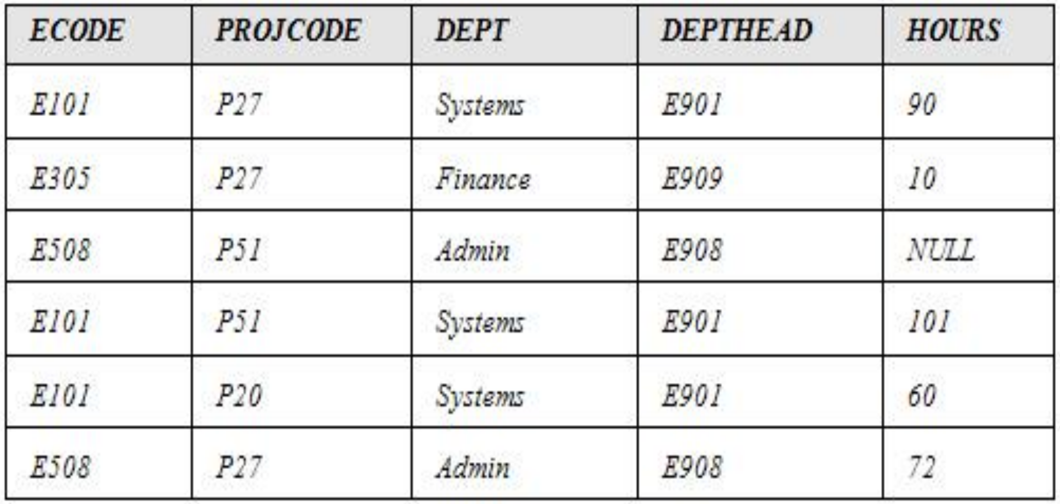


ECODE	PROJCODE	DEPT	DEPTHEAD	HOURS
E101	P27	Systems	E901	90
E305	P27	Finance	E909	10
E508	P51	Admin	E908	NULL
E101	P51	Systems	E901	101
E101	P20	Systems	E901	60
E508	P27	Admin	E908	72

Definition of Normalization (Contd.)

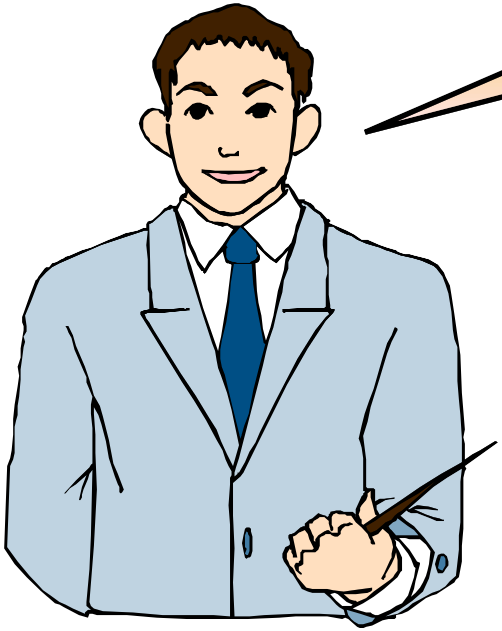
Composite Key

DEPTHEAD is functionally dependent on ECODE; however, it is not dependent on the attribute, PROJCODE.



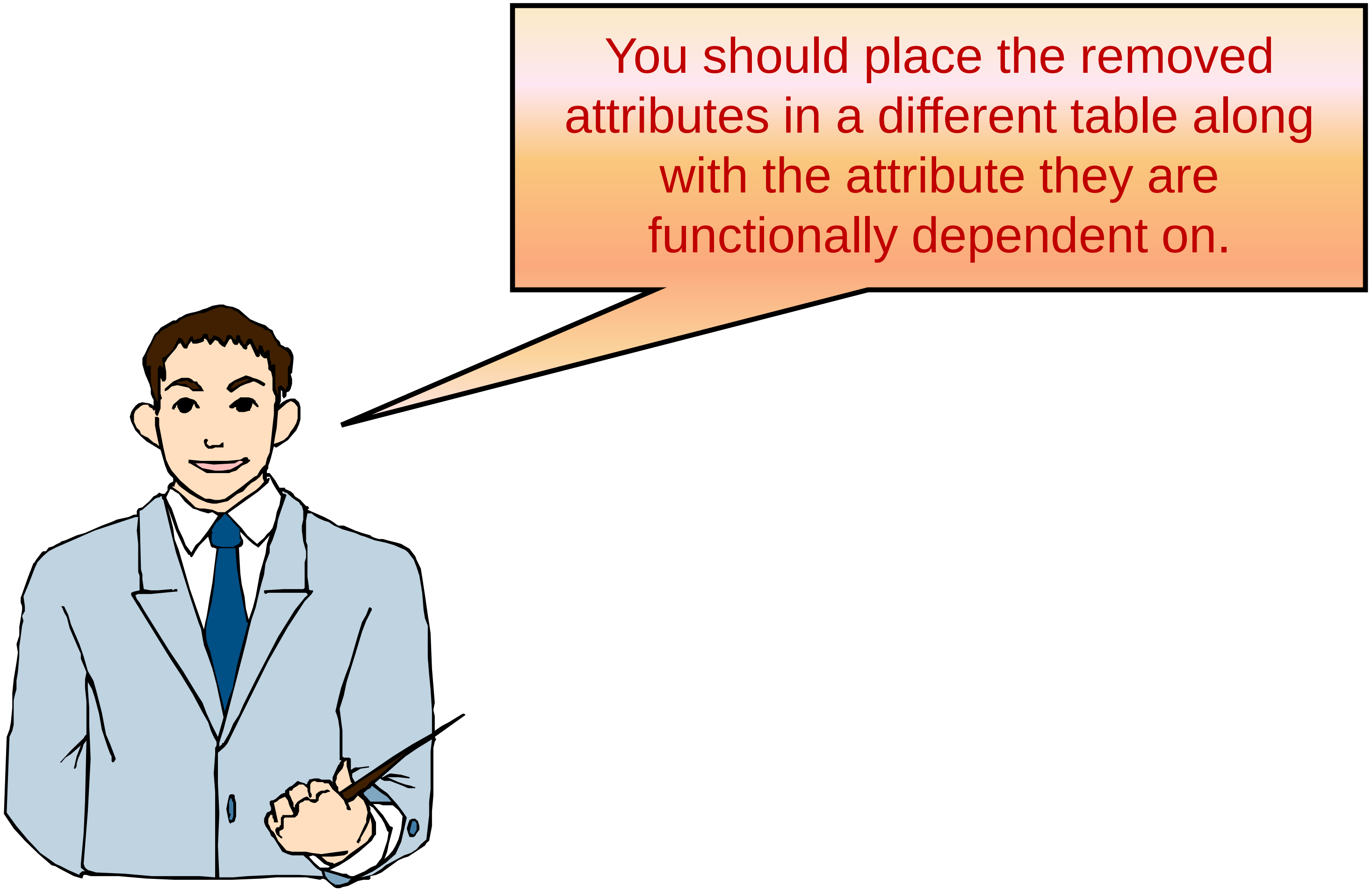
ECODE	PROJCODE	DEPT	DEPTHEAD	HOURS
E101	P27	Systems	E901	90
E305	P27	Finance	E909	10
E508	P51	Admin	E908	NULL
E101	P51	Systems	E901	101
E101	P20	Systems	E901	60
E508	P27	Admin	E908	72

Definition of Normalization (Contd.)



To convert the PROJECT table into 2NF, you must remove the attributes that are not functionally dependent on the whole key.

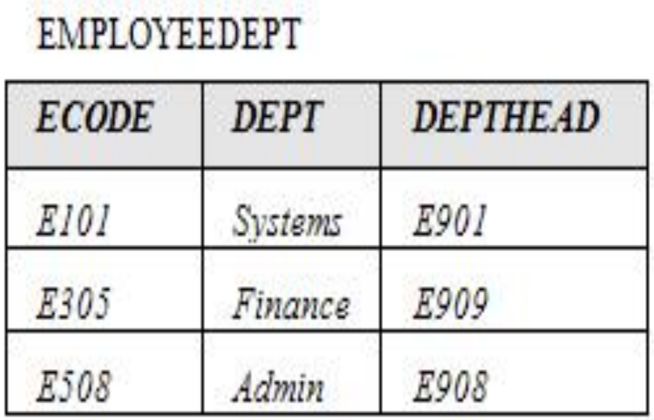
Definition of Normalization (Contd.)



You should place the removed attributes in a different table along with the attribute they are functionally dependent on.

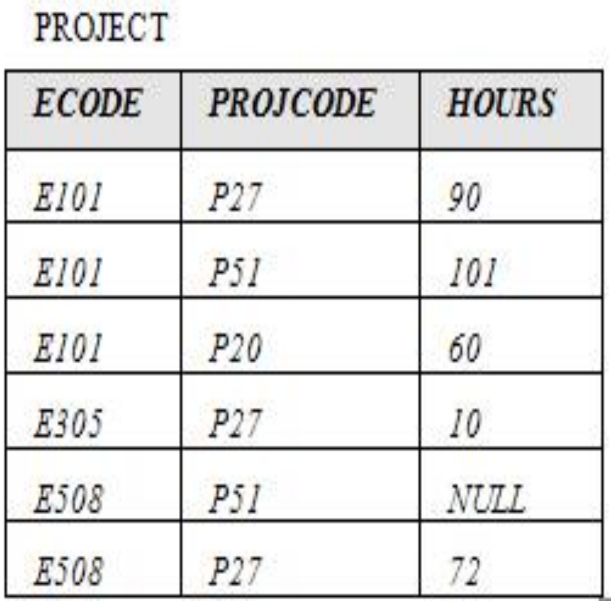
Definition of Normalization (Contd.)

- The EMPLOYEEDEPT and PROJECT tables are in 2NF, as shown in the following diagram.



EMPLOYEEDEPT

ECODE	DEPT	DEPTHEAD
E101	Systems	E901
E305	Finance	E909
E508	Admin	E908



PROJECT

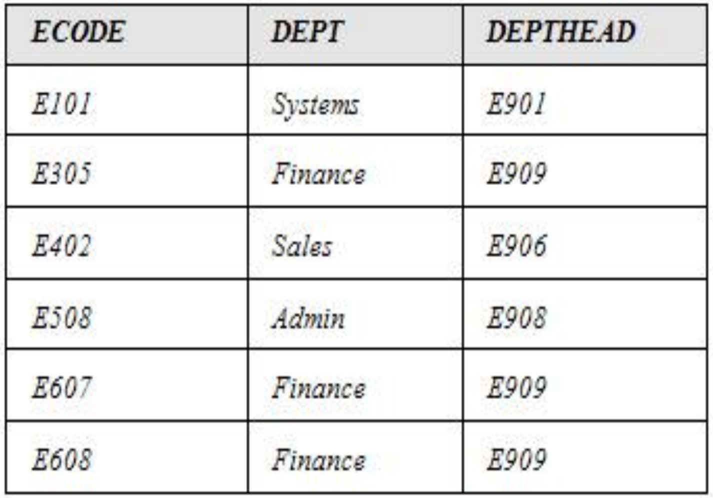
ECODE	PROJCODE	HOURS
E101	P27	90
E101	P51	101
E101	P20	60
E305	P27	10
E508	P51	NULL
E508	P27	72

Definition of Normalization (Contd.)

- Third Normal Form (3NF):
  - A relation is said to be in the 3NF if and only if:
    - It is in 2NF, and
    - No transitive (indirect) dependency exists between non-key attributes and key attributes.
  - The guidelines for converting a table into 3NF are:
    - Find and remove non-key attributes that are functionally dependent on attributes that are not the primary key. Place them in a different table.
    - Group the remaining attributes.

Definition of Normalization (Contd.)

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

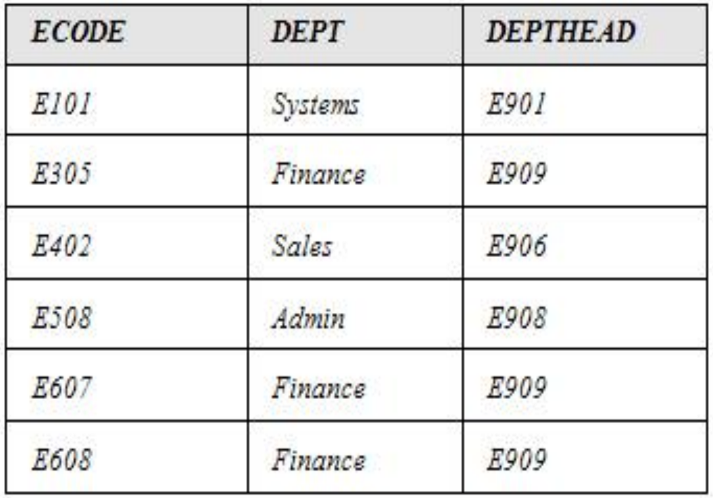


ECODE	DEPT	DEPTHEAD
E101	Systems	E901
E305	Finance	E909
E402	Sales	E906
E508	Admin	E908
E607	Finance	E909
E608	Finance	E909

- The preceding table could lead to the following problems:
  - Insertion
  - Updation
  - Deletion

Definition of Normalization (Contd.)

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



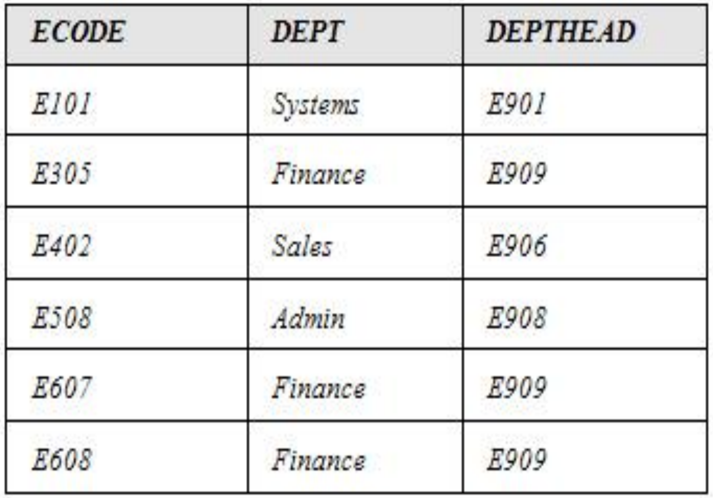
ECODE	DEPT	DEPTHEAD
E101	Systems	E901
E305	Finance	E909
E402	Sales	E906
E508	Admin	E908
E607	Finance	E909
E608	Finance	E909

- The preceding table could lead to the following problems:
  - Insertion
  - Updation
  - Deletion

The department head of a new department that does not have any employees at present cannot be entered in the DEPTHEAD column.

Definition of Normalization (Contd.)

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



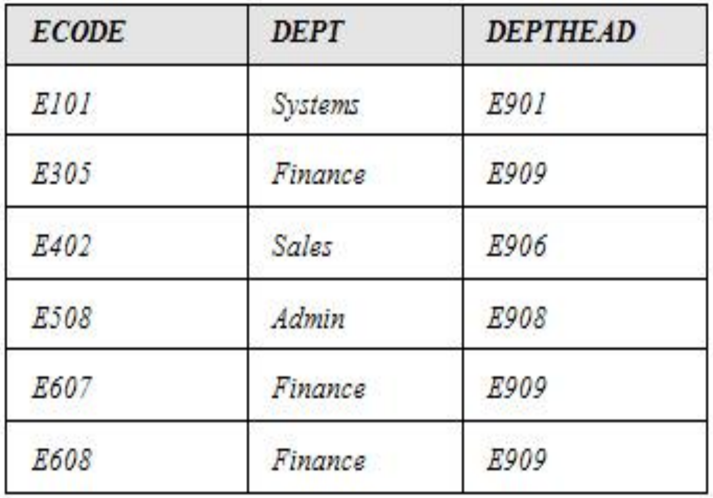
ECODE	DEPT	DEPTHEAD
E101	Systems	E901
E305	Finance	E909
E402	Sales	E906
E508	Admin	E908
E607	Finance	E909
E608	Finance	E909

- The preceding table could lead to the following problems:
  - Insertion
  - Updation
  - Deletion

For a department, DEPTHEAD is repeated. Any change will have to be made consistently across the table.

Definition of Normalization (Contd.)

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

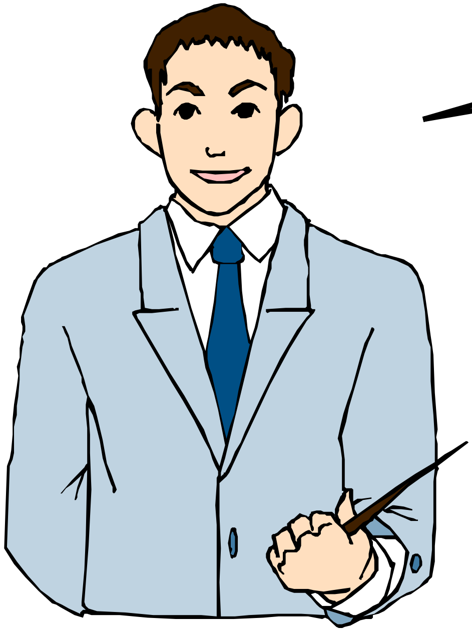


ECODE	DEPT	DEPTHEAD
E101	Systems	E901
E305	Finance	E909
E402	Sales	E906
E508	Admin	E908
E607	Finance	E909
E608	Finance	E909

- The preceding table could lead to the following problems:
  - Insertion
  - Updation
  - Deletion

If an employee record is deleted, the information about DEPTHEAD will also be deleted.

Definition of Normalization (Contd.)

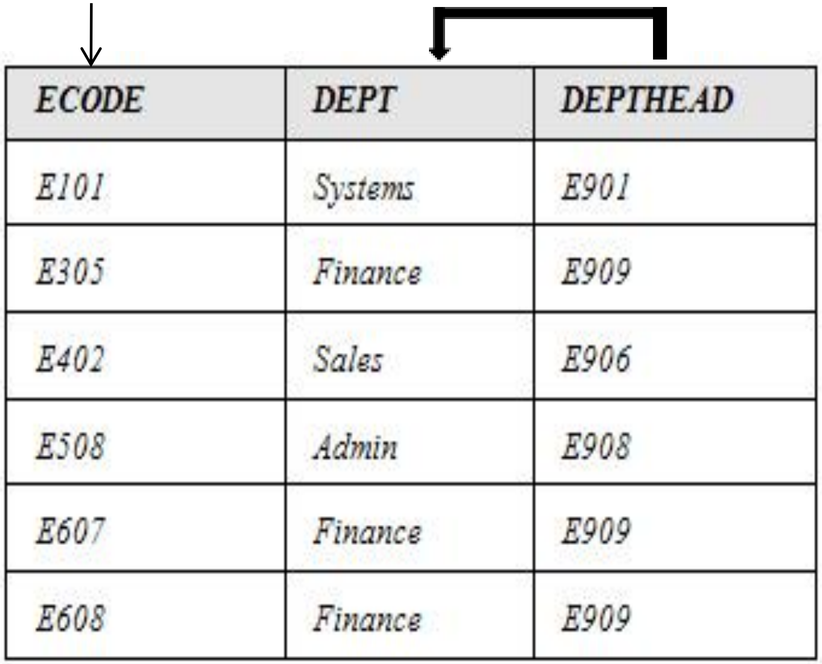


Let us check if the EMPLOYEE table is in 3NF.

Definition of Normalization (Contd.)

Primary key

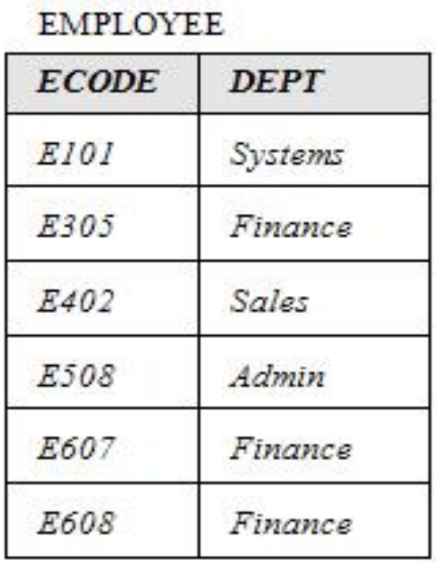
DEPTHEAD is functionally dependent on DEPT, which is not a primary key.



ECODE	DEPT	DEPTHEAD
E101	Systems	E901
E305	Finance	E909
E402	Sales	E906
E508	Admin	E908
E607	Finance	E909
E608	Finance	E909

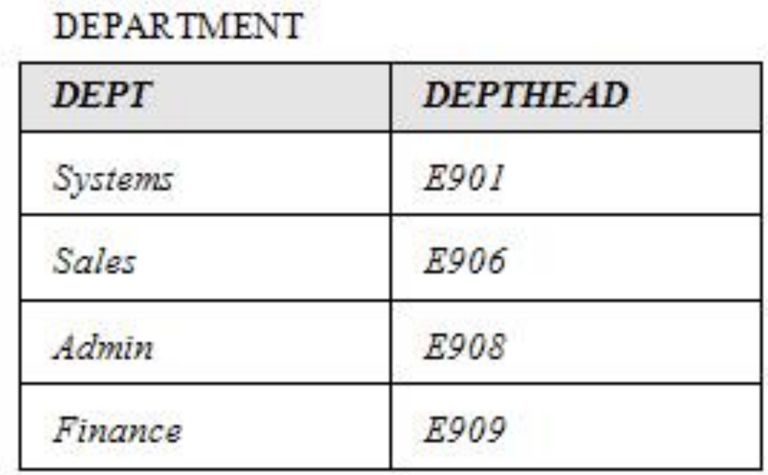
Definition of Normalization (Contd.)

- To convert the EMPLOYEE table into 3NF, you must remove the DEPTHEAD column and place it in another table, as shown in the following diagram.



EMPLOYEE

ECODE	DEPT
E101	Systems
E305	Finance
E402	Sales
E508	Admin
E607	Finance
E608	Finance



DEPARTMENT

DEPT	DEPTHEAD
Systems	E901
Sales	E906
Admin	E908
Finance	E909

Definition of Normalization (Contd.)

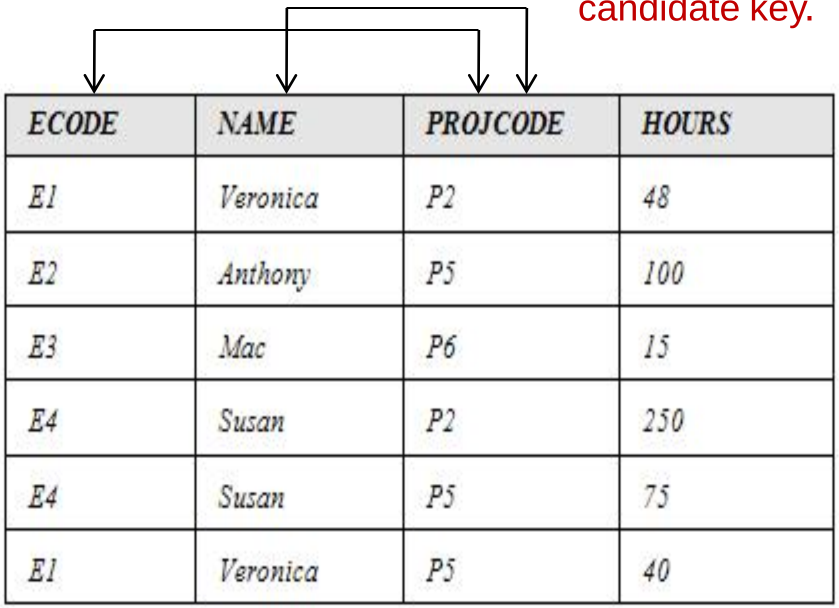
- Boyce-Codd Normal Form (BCNF):
  - The original definition of 3NF was not sufficient in some situations. It was not satisfactory for the tables:
    - That had multiple candidate keys.
    - Where the multiple candidate keys were composite.
    - Where the multiple candidate keys overlapped (had at least one attribute in common).
  - The guidelines for converting a table into BCNF are:
    - Find and remove the overlapping candidate keys. Place the part of the candidate key and the attribute it is functionally dependent on, in a different table.
    - Group the remaining items into a table.

Definition of Normalization (Contd.)

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

Primary Key

NAME+PROJCODE can also be chosen as the primary key and hence, is a candidate key.



ECODE	NAME	PROJCODE	HOURS
E1	Veronica	P2	48
E2	Anthony	P5	100
E3	Mac	P6	15
E4	Susan	P2	250
E4	Susan	P5	75
E1	Veronica	P5	40

PROJECT table is already in 3NF

Definition of Normalization (Contd.)

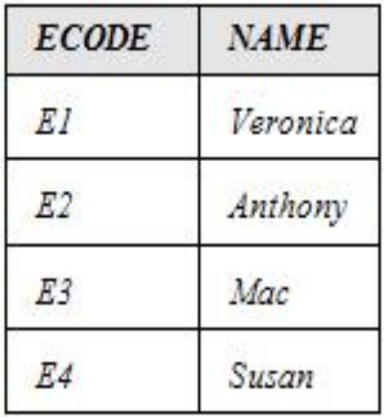
- The following points describe the functional dependencies in the PROJECT table:
  - HOURS is functionally dependent on ECODE+PROJCODE.
  - HOURS is also functionally dependent on NAME+PROJCODE.
  - NAME is functionally dependent on ECODE.
  - ECODE is functionally dependent on NAME.
- You will notice that the PROJECT table has:
  - Multiple candidate keys that are ECODE+PROJCODE and NAME+PROJCODE.
  - Composite candidate keys.
  - Candidate keys that overlap since the PROJCODE attribute is common between the two candidate keys.

Definition of Normalization (Contd.)

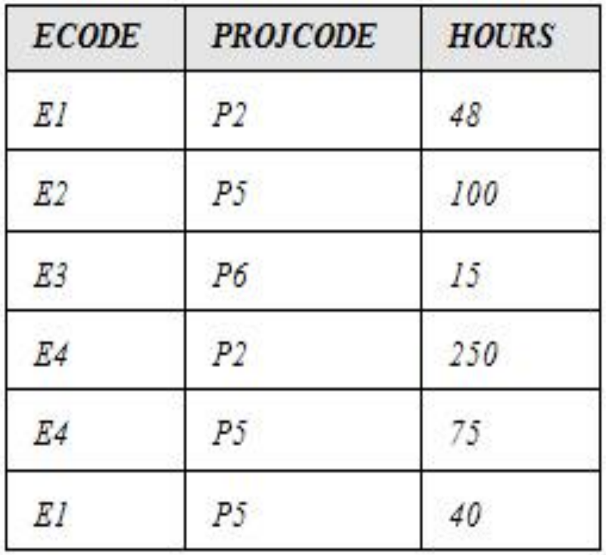
- The only non-key item is HOURS, which is dependent on the whole key, ECODE+PROJCODE or NAME+PROJCODE.
- ECODE and NAME are determinants since they are functionally dependent on each other.
- As per BCNF, the determinants have to be candidate keys.
- You can remove NAME and ECODE and place them in a different table.

Definition of Normalization (Contd.)

- You can remove NAME and ECODE and place them in a different table, as shown in the following diagram.



ECODE	NAME
E1	Veronica
E2	Anthony
E3	Mac
E4	Susan



ECODE	PROJCODE	HOURS
E1	P2	48
E2	P5	100
E3	P6	15
E4	P2	250
E4	P5	75
E1	P5	40

Just a minute

- Which of the following points helps in achieving a good database design?
  1. A table should store data for all the related entities together.
  2. Each table should have an identifier.
  3. Columns that contain NULL values should be created.

- Solution:
  2. Each table should have an identifier.

Just a minute

- In which normal form, you need to remove non-key attributes that are functionally dependent on non-primary key attributes?

- Solution:
  - Third normal form

Summary

- In this session, you learned that:
  - A table is said to be in 2NF when it is in the 1NF, and no partial dependency exists between non-key attributes and key attributes.
  - A relation is said to be in the 3NF if and only if it is in 2NF, and no transitive (indirect) dependency exists between non-key attributes and key attributes.
  - A relation is in BCNF if and only if every determinant is a candidate key.