###### Cairo University

**Faculty of Computers and Information Information Systems Department Database Systems 1**

**Section 8 (Normalization)**

* **Normalization** is a process for assigning attributes to entities. It reduces data redundancies and helps eliminate the data anomalies.
* Normalization works through a series of stages called normal forms:
  + **First normal form (1NF)**
  + **Second normal form (2NF)**
  + **Third normal form (3NF)**
* The highest level of normalization is not always desirable.

**Example: Suppose we have this table:**

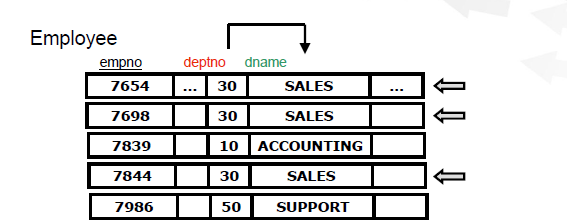
**Employee (empno, ename, deptno, dname, addr)**

What is wrong with this table?

**1. Redundancy:** deptno&dname.

**2. Extra Work:** if we changed the name of a department, we will have to do it in multiple places**.**

**3. Anomalies:** could change deptno without changing dname or vice versa.

****

Entities with the same value for deptno have the same value for dname.

•Including dname in the entity class is redundant, since it can be derived from deptno.

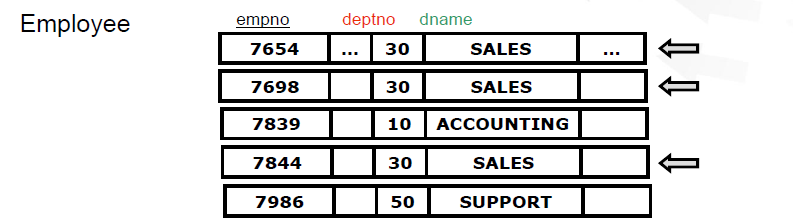
**Redundancy causes duplicate work:** suppose that the company wants to change deptno 30 to be the Sales & Marketing department, this change must be made to multiple employees.

***Redundancy can cause anomalies (inconsistencies) if modifications are not done carefully.***

**• Update Anomaly:** updating a value in a single cell can make the database inconsistent.

**• Insertion Anomaly:** adding an entity can make the database inconsistent.

**• Deletion Anomaly:** deleting some information can make the database inconsistent or cause unintended loss of information.

****

**Examples:**

**Modification Anomaly**: modify 7654's dname to 'SUPPORT' (without changing its deptno).

**Insert Anomaly**: insert a new employee with a deptno of 20, and a dname of 'SUPPORT'.

**Delete Anomaly**: delete employee 7986 (it’s the only employee in SUPPORT, and no other entity class keeps track that dept 50 is SUPPORT).

**Normalization:**

1. **First Normal Form (1NF) Definition:**
   * The term first normal form (1NF) describes the tabular format in which:
     + All the key attributes are defined.
     + There are no repeated groups in the table.
     + All attributes are dependent on the primary key.

**Conversion to First Normal Form**

* + A relational table must not contain repeated groups. That is each row/column intersection can contain only one value, not a set of values.
  + Repeated groups can be eliminated by adding the appropriate entry in at least the primary key column(s).

**Functional Dependencies:**

* A *Functional Dependency* describes a relationship between *attributes* in a single relation.
* An attribute is *functionally dependant* on another if we can use the value of one attribute to determine the value of another.
* We use the symbol ‘🡪' to indicate a functional dependency:   
  X 🡪 Y.
* Read "X functionally determines Y" or "Y functionally depends on X", more simply “*X* determines Y“or “Y depends on X".

**Examples:**

* Student\_ID 🡪Saddress, SDoB
* Student\_ID, Course# 🡪 Grade
* Model, Year 🡪CarPrice
* Course\_No, Section 🡪 Professor, Classroom, Number of Students

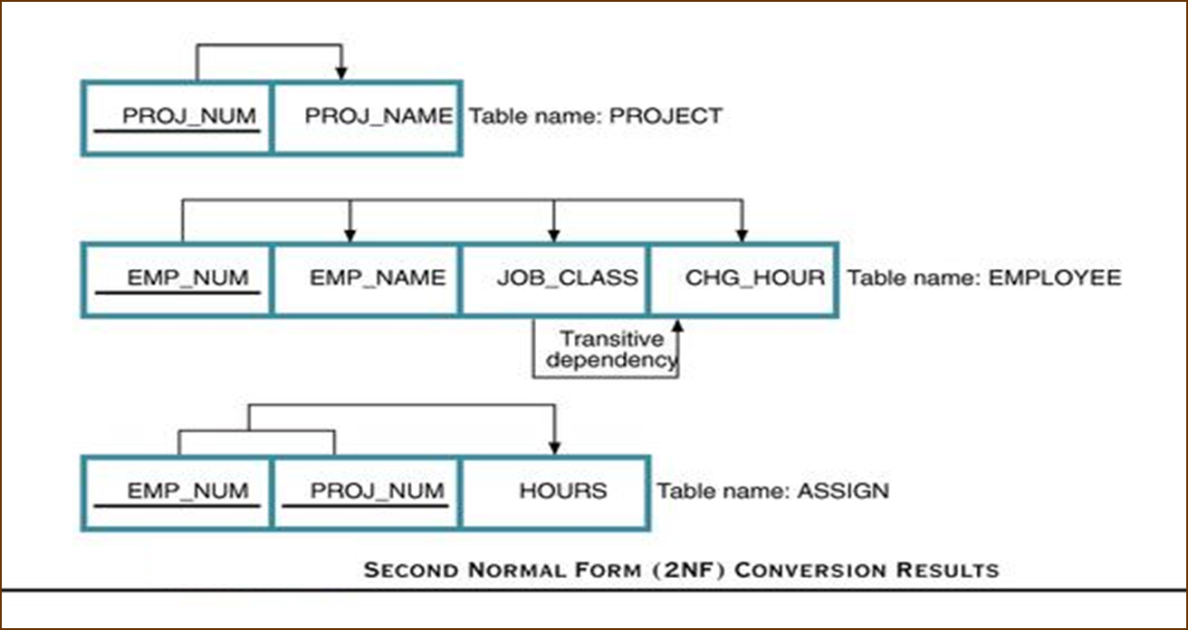
1. **Second Normal Form**

* Conversion to Second Normal Form: starting with the 1NF format, the database can be converted into the 2NF format by:
* Writing each key component on a separate line, then
* Writing the original key on the last line and
* Writing the dependent attributes after each new key.

PROJECT (PROJ\_NUM, PROJ\_NAME)

EMPLOYEE (EMP\_NUM, EMP\_NAME, JOB\_CLASS, CHG\_HOUR)

ASSIGN (PROJ\_NUM, EMP\_NUM, HOURS)

****

1. **Third Normal Form:**

A table is in 3NF if it is in 2NF and it contains no transitive dependencies.

* Conversion to third Normal Form:
* Create a separate table with attributes in a transitive functional dependence relationship.

PROJECT (PROJ\_NUM, PROJ\_NAME)

ASSIGN (PROJ\_NUM, EMP\_NUM, HOURS)

EMPLOYEE (EMP\_NUM, EMP\_NAME, JOB\_CLASS)

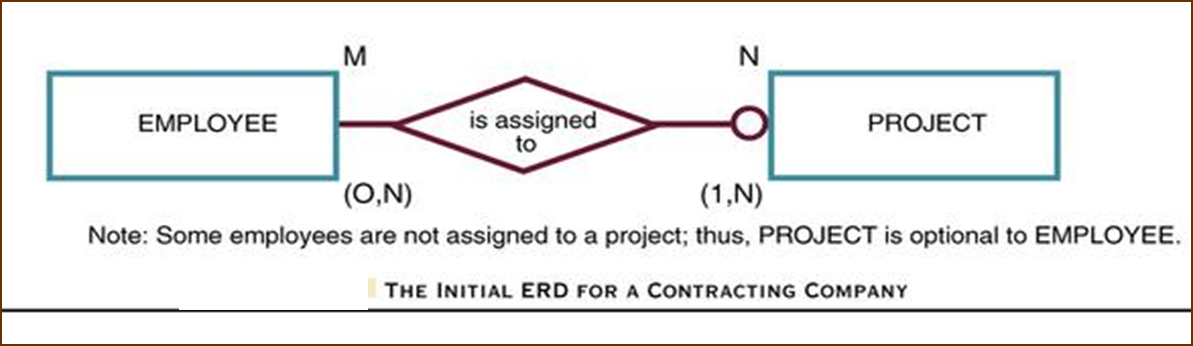
JOB (JOB\_CLASS, CHG\_HOUR)

**Example:**

Construction Company manages many projects. Each project requires the services of many employees. An employee may be assigned to several different projects. Some employees are not assigned to a project and perform duties not specifically related to a project. Some employees are part of a labor pool, to be shared by all project teams. Each employee has a (single) primary job classification. This job classification determines the hourly billing rate. Many employees can have the same job classification.

1. **Two Initial Entities:**

PROJECT (PROJ\_NUM, PROJ\_NAME)

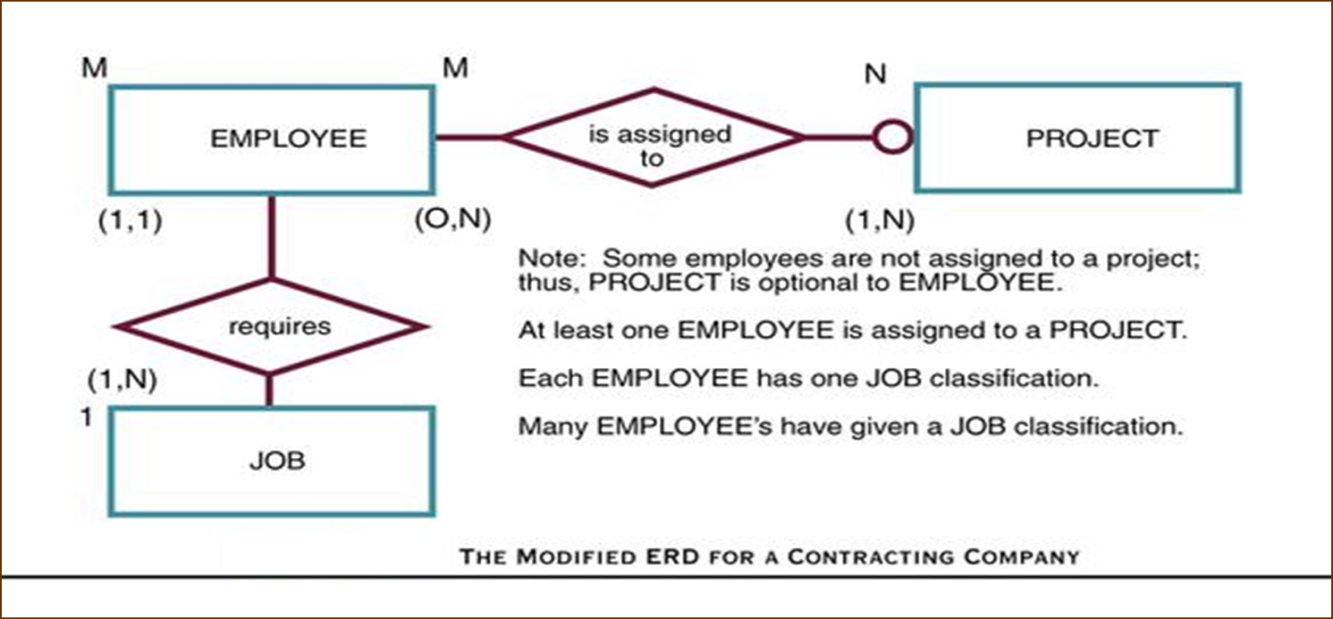
EMPLOYEE(EMP\_NUM,EMP\_LNAME,EMP\_FNAME,EMP\_INITIAL, JOB\_DESCRIPTION, JOB\_CHG\_HOUR)

1. **Three Entities After Transitive Dependency Removed**

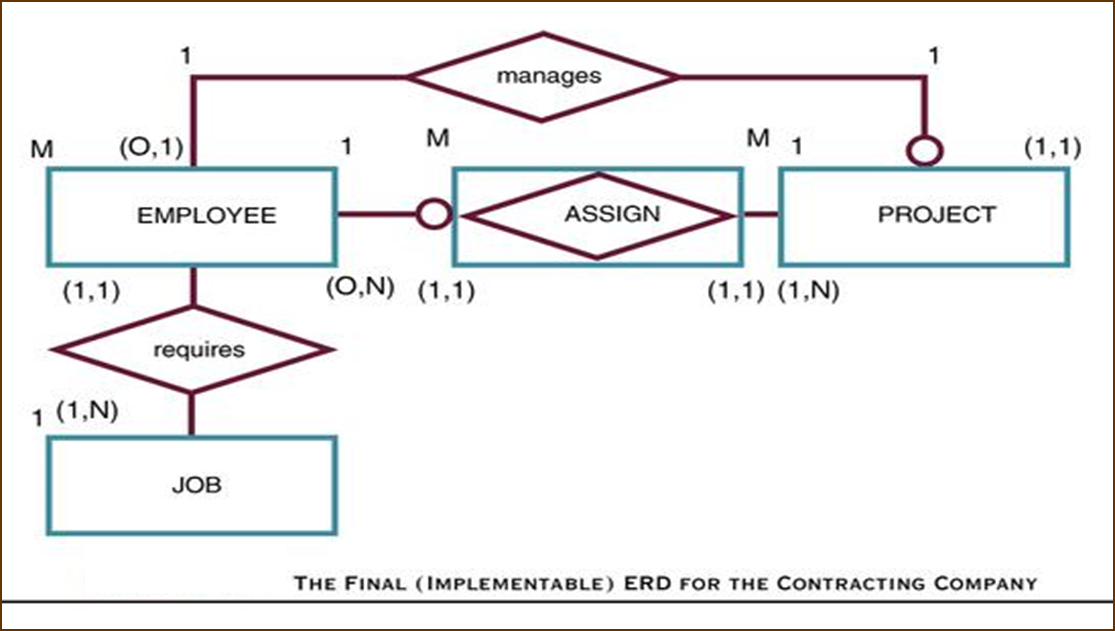
PROJECT (PROJ\_NUM, PROJ\_NAME)

EMPLOYEE (EMP\_NUM, EMP\_LNAME, EMP\_FNAME, EMP\_INITIAL, JOB\_CODE)

JOB (JOB\_CODE, JOB\_DESCRIPTION, JOB\_CHG\_HOUR)



1. **Composite Entity Assign**

****

Attribute ASSIGN\_HOUR is assigned to the composite entity ASSIGN.

“Manages” relationship is created between EMPLOYEE and PROJECT.

PROJECT (PROJ\_NUM, PROJ\_NAME, EMP\_NUM)

EMPLOYEE (EMP\_NUM, EMP\_LNAME, EMP\_FNAME, EMP\_INITIAL, EMP\_HIREDATE, JOB\_CODE)

JOB (JOB\_CODE, JOB\_DESCRIPTION, JOB\_CHG\_HOUR)

ASSIGN (ASSIGN\_NUM, ASSIGN\_DATE, PROJ\_NUM, EMP\_NUM, ASSIGN\_HOURS)