

**Data Management and Database Design**

**INFO 6210**

**Fall 2016**

**Assignment 3**

**Submission Date – Oct/05/2016 Wednesday**

**Student Name: Malick Fairoz Sayeed Abuthahir**

**NUID: 001235450**

**Program: MS in Information Systems**

**Professor Name: Yusuf Ozbek**

**College: College Of Engineering**

**University: Northeastern University**

## Book-1-A

### 1. Define each of the following terms.

#### a. Supertype

Supertype is generic entity type that has relationship with one or more subtypes. For example "student" is a generic entity which is a supertype and the undergraduate student and graduate student are the subtypes which come under the supertype.

#### b. Subtype

Subtype is a subgrouping of entities in an entity type that has meaningful to the organization and shares common attributes from other subgroupings.

#### c. Specialization

Specialization is the process of defining one or more subtypes of the supertype and forming a supertype relationship. It is a top-down process which is opposite to generalization process.

#### d. Entity cluster

An entity cluster is a set of one or more entity types and associated relationship grouped into a single abstract entity type.

#### e. Completeness constrain

This explains whether an instance of a supertype must be a member of at least any one of the subtypes. This has been classified into two rules they are total specialization rule and partial specialization rule.

#### f. Enhanced Entity relationship (EER) model

EER is used identify the model that has resulted in extending the original ER model with new modeling techniques. This extension makes EER similar to object oriented data modeling.

#### g. Subtype discriminator

It is an attribute of the supertype whose values determine the target subtype.

#### h. Total specialization rule

This rule explains that each entity instance from a supertype must be member of some subtype in the relationship.

#### i. Generalization

Generalization is a process of defining more general entity types from the set of more specified entity types. Thus it is known as bottom-up process.

j. Disjoined rule

It is a rule that specifies that an instance of a supertype may not be a member of two or more subtypes simultaneously.

k. Overlap rule

It is a rule that specifies that an instance of a supertype may simultaneously member of two or more subtypes.

l. Partial specialization rule

This rule specifies that the entity instance of the supertype are not allowed to be member of any subtype.

m. Universal data model

Universal data model is not a correct data model but it is successful starting point for developing an excellent data modeling project for an organization.

2. Match the following terms and definitions

e	Supertype	a	Subset of supertype
f	Entity cluster	b	Entity belong to two subtypes
a	Subtype	c	Subtype get supertype attributes
c	Specialization	d	Generalized entity type
g	Subtype discriminator	e	Creating subtype for an entity type
d	Attribute inheritance	f	A group of associated entity type and relationship
b	Overlap rule	g	Locates target subtype of an entity

3. Contrast the following terms

a)	<b>Supertype</b>	<b>Subtype</b>
	It is a generic entity type and has relationship with one or more subtype	It is a subgrouping of entity that has meaningful to organization and relationship distinct from other subgroupings
b)	<b>Generalization</b>	<b>Specialization</b>
	It defines more general entity types from a set of more specialized entity types	It defines one or more subtypes of a supertypes
c)	<b>Disjoint rule</b>	<b>Overlap rule</b>
	The Instance of the supertype may not simultaneously member of two or more subtypes	The Instance of the supertype may simultaneously member of two or more subtypes
d)	<b>Total specialization rule</b>	<b>Partial specialization rule</b>
	It specifies each instance of a supertype must be a member of at least one subtype	It specifies each instance of a supertype must not be a member of any subtype
e)	<b>Party</b>	<b>Party role</b>
	The core structure of the entity types are known as party, which generates persons or organization as the actors of the enterprise	The associated entity types are known as party role, which generalized various roles party can play at different times.

4. State two conditions that indicate when a database designer should consider using supertype/subtype relationship.

Ans)

- If the attributes that apply to some instance of an entity types
- If the instance of the subtype participate that are unique relationship to the subtype

5. State the reason for entity clustering

Ans)

The main reason for entity clustering is to present a data model in a useful way for a large and complex organization. It is a hierarchical decomposition of a macro level view of the data model into finer views, eventually resulting in a detailed data model.

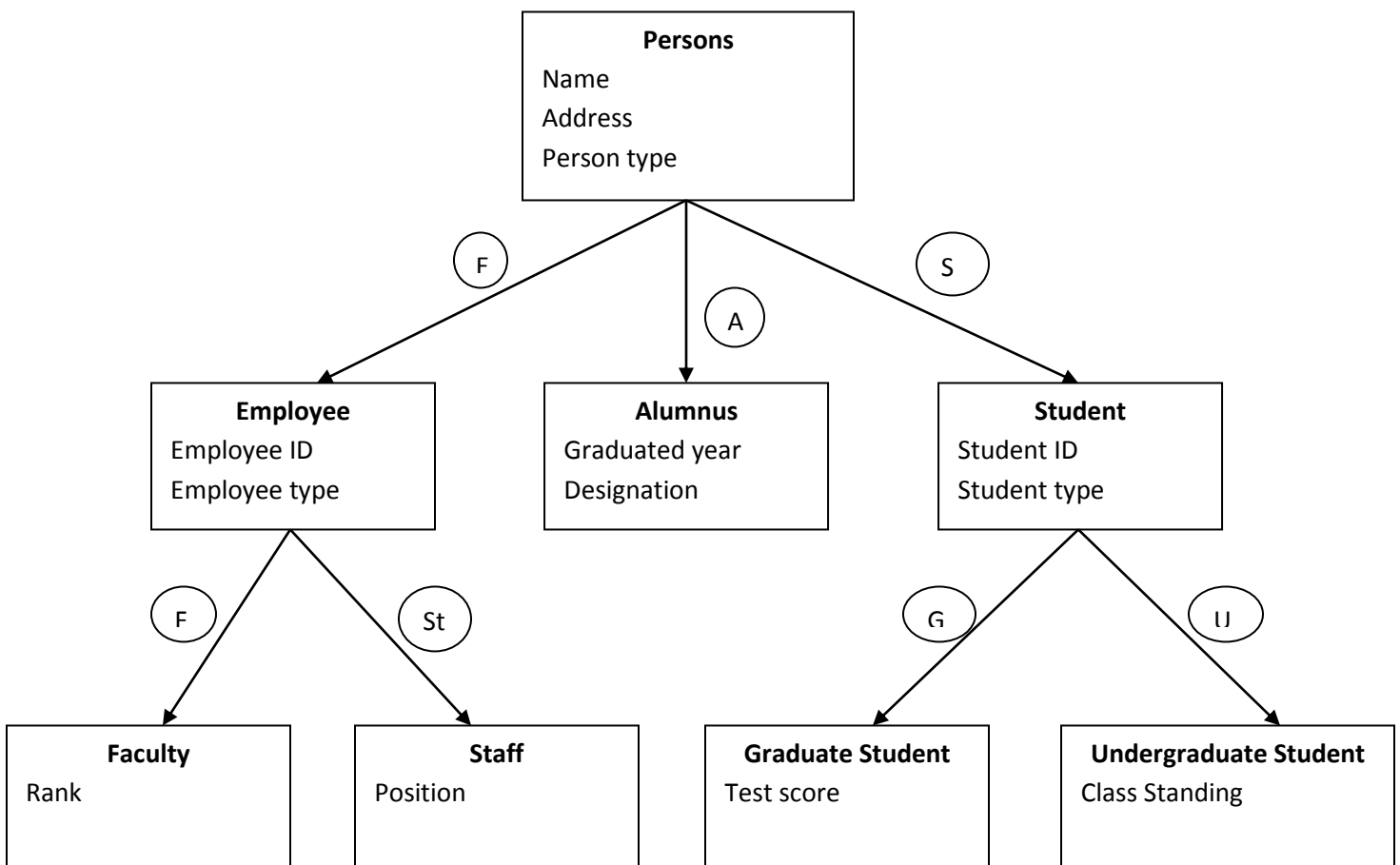
## Book-1-B

### Answers only for exercise problems

1. I come under the subtype Graduate student and the list of attributes are mentioned below

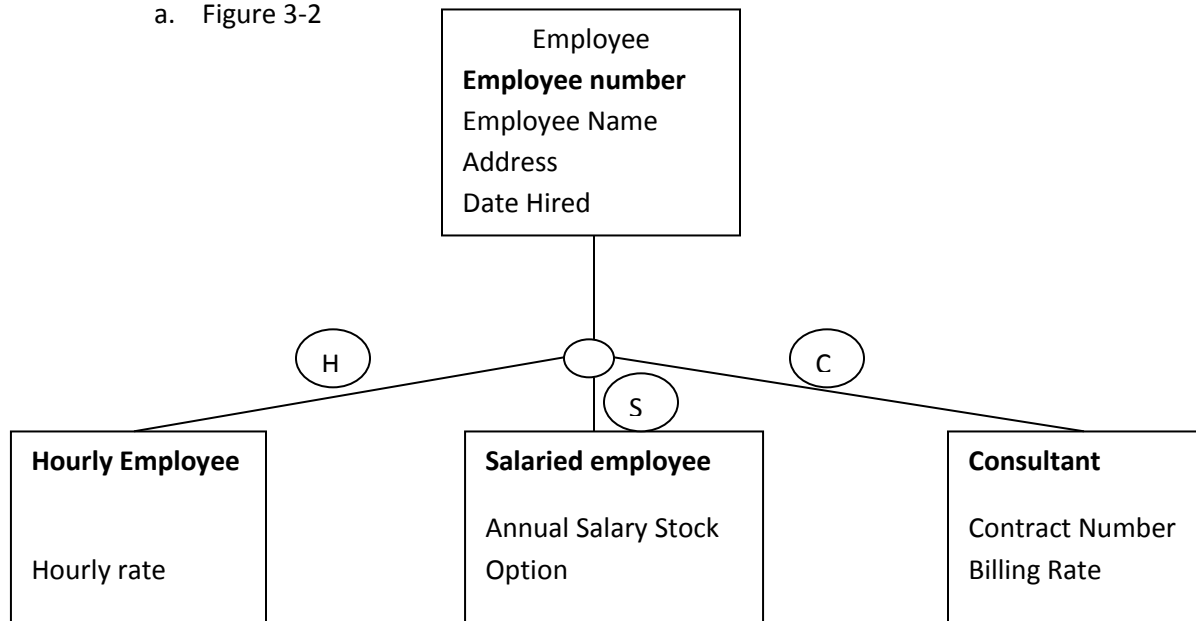
- a. Name - Malick
- b. Student ID - 001235450
- c. Department Information System
- d. Grade 3.8
- e. Address 693 Parker Street
- f. Gender Male
- g. DateOfBirth 1988-10-02

2. The subtype discriminator has been added each super types and the discriminator values are assigned to each subtypes in the below diagrams.

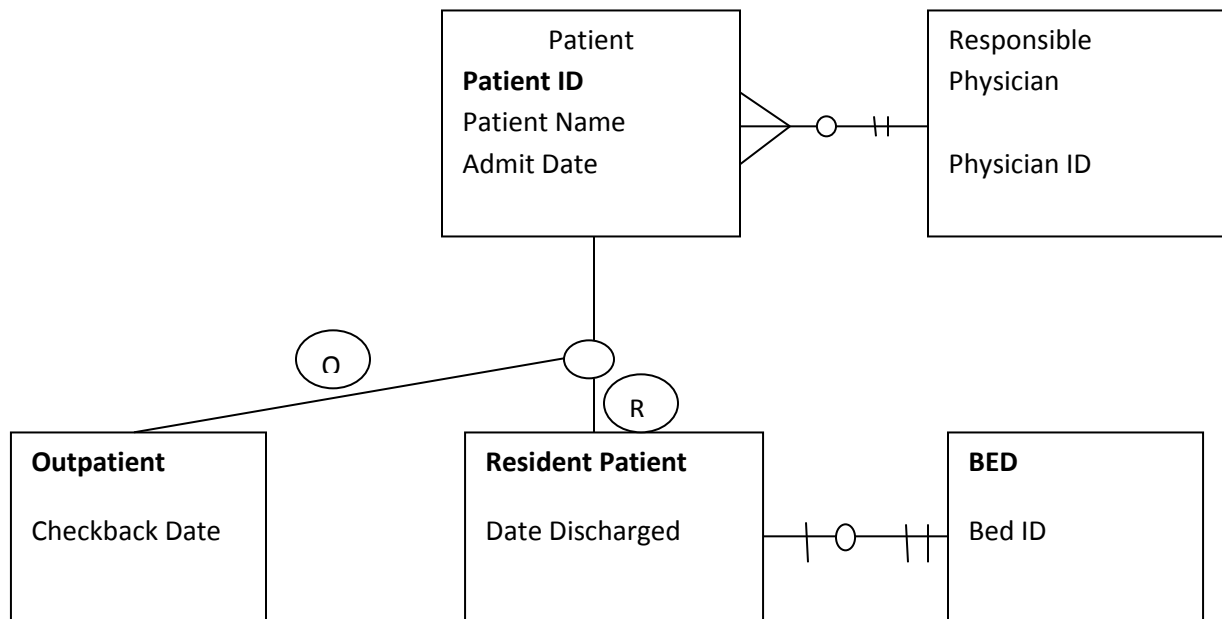


3. Subtype discriminator

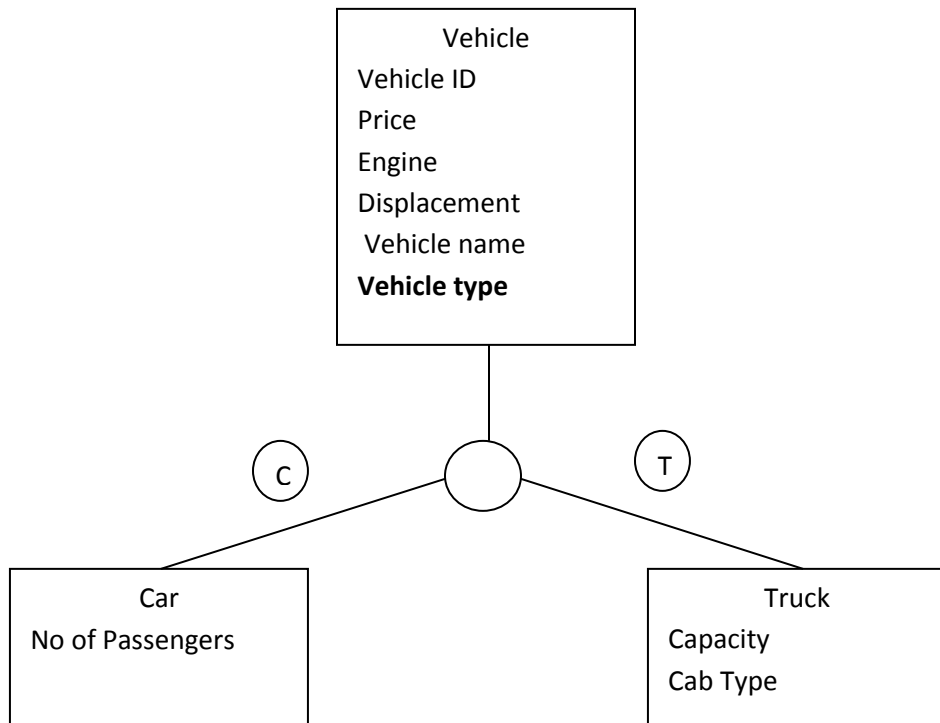
a. Figure 3-2



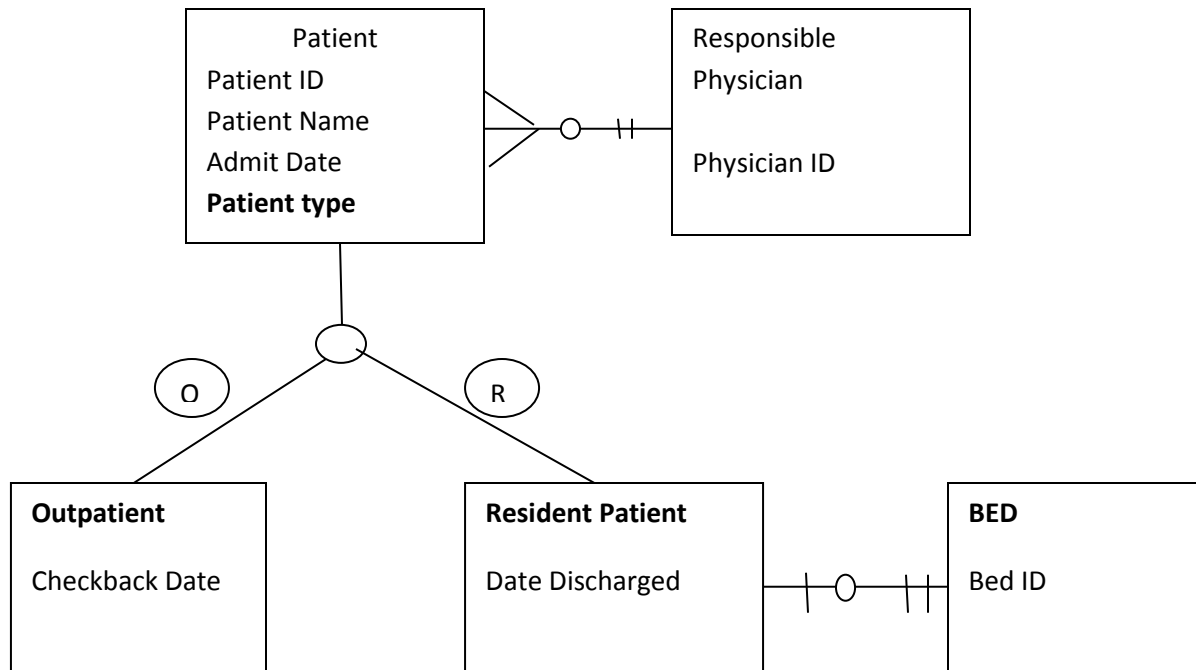
b. Figure 3-3



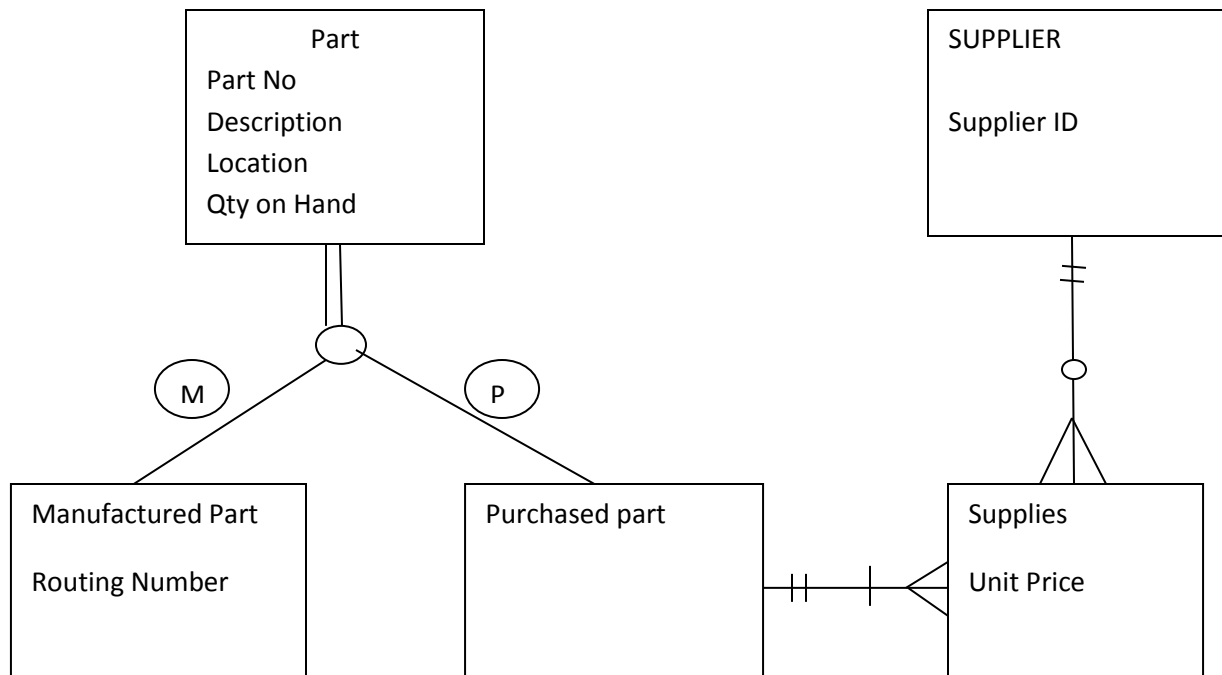
c. Figure 3-4b



d. Figure 3-7a



e. Figure 3-7b



4. The EER diagram explains that the employee entity is the supertype which has the attributes Employee Number, Employee Name, Address and Date Hired. The hourly employee, salaried employee and consultant are the subtype of the employee entity, which has the hourly rate as the attribute in the hourly employee subtype, and then annual salary and stock option are the attribute in the salaried employee subtype, similarly contact number and billing rate are the attributes in the consultant subtype. This flow type is a disjointed rule. In these three subtypes they have their unique attributes and to differentiate them we use discriminator.
5. In this EER diagram the patient entity is the supertype and the outpatient and resident patient are the subtype from the patient entity. This flow is known as disjointed rule. For each patient one physician is associated and for each physician many patients are associated. Here the patient type is the discriminator values.



## Book-2

1. When is the best time to use an RDBMS program's design tools?

Ans)

After designing the logical structure of the database which describes the size, shape and necessary systems for the database, it is a good time to use an RDBMS program to design the physical structure of the database.

2. True or False: Design is crucial to the consistency, integrity, and accuracy of data.

Ans)

True.

3. What is the most detrimental result of improper database design?

Ans)

Inaccurate information is the most detrimental result of improper database design.

4. What fact makes the relational database structurally sound and able to guarantee accurate information?

Ans)

Implementing the database with good design is the important fact that makes the relational database structurally sound and can able to guarantee accurate information.

5. State two advantages of learning a design methodology

Ans)

- a) It will give you the skill to design a good database structure.
- b) It will reduce the amount of time spend on designing the database and makes the process easier.

## Book-3

1. Why is terminology important?

Ans)

Terminology is important because they define the special ideas and concepts of relational database model. They are used to define the database design process itself and also can be used anywhere in the RDBMS.

2. Name the four categories of terms.

Ans)

The four categories of terms are

- a) Value related
- b) Structure related
- c) Relationship related
- d) Integrity related

3. What is the difference between data and information?

Ans)

Data is the value you store in the database and the information is the data that you process in a manner that makes it meaningful and useful when you retrieve it or view it.

4. What does a null represent?

Ans)

The null represent the missing or unknown value. It is not zero or blank space. Because both are consider to be meaningful in a query language like SQL.

5. What is a null's major disadvantage?

Ans)

The null's major disadvantage is they have adverse effect on mathematical operations. Like whenever a null value present in a mathematical operation then the result will also be the null value only which is unknown.

## Book-4

1. Why is it important to complete the design process thoroughly?

Ans)

Database design is considered to be very important since the structural integrity and data integrity in the database is directly proportional how thoroughly you follow the design process. It is well known that attempting to design a database without employing the thorough database design process is a bad idea.

2. True or False: The level of structural integrity is in direct proportion to how thoroughly you follow the design process.

Ans)

True.

3. What is the purpose of a mission statement?

Ans)

When defining the purpose of the database in the mission statement, it will ensure that you develop an appropriated database structure and gather the data necessary to support the intended purpose of the database.

4. What are mission objectives?

Ans)

The mission objectives are the statements that represent the general tasks that user can perform against the data on the database. It supports the mission statements to identify the various aspect of the database structure.

5. What constitutes your organization's fundamental data requirements?

Ans)

On interviewing both the user and management to indentify how they interact with the database, we will gathered the information and analyze them to compile the initial lists of fields. We then refine this list by removing the calculated list which will be used later in the design process. This refined list of fields constitutes the organizational fundamental data requirements.