# Chapter 7

## 2. What is a useful technique for identifying entities in a written description of a data modeling problem?

* One of the ways to identify entities is underline all the noun use in the client provided documentation or system proposal. In most cases nouns are the entities associated with the system.
* Most of the systems are centered around one critical entity, so it easier to identify that entity. For e.g. for loan origination system, loan is central entity around which other entities would be designed like users, loan payments etc. Once central entity is identified other entities revolving around can be designed easily.

## 6. Describe the different types of entities.

1. Independent Entity

It is core entity around which other entities are build. For e.g. for library management system Book is core entity. For loan origination system Loan is an independent entity. They have unique identifiers.

1. Weak or dependent entity

It relies on the existence of other main entity. For e.g. in you system you have City, State and Country. Then existence of city depends on state, city cannot exist without state. Same applies for state and country, state cannot exist without country.

1. Associative Entity

Typically this is mapping entity which exists because of many to many relationship between 2 entities. Associative entities hold the information that associates two or more entities that are related in a many-to-many relationship.

Entity STUDENTS, CLASSES

Each student takes many classes; each class has many students. So we have the associative entity STUDENT-CLASSES. The attributes might include dates.

1. Aggregate Entity

This entity is created when other types of entities (e.g. Independent entities) have common set of attributes. For e.g. in an ecommerce platform there are customer and vendors both need to have physical address. So address becomes an aggregate entity.

1. Subordinate Entity

It is an entity that stores information about an entity whose attributes vary for instances. For e.g. a dealer who sells car of various types like sedan and SUV. Now each of those types will have its own unique attributes.

## 8. Why might a data model contract?

* Generalization leads to data model contraction.
* It is relationship between general element and specific elements. For e.g. Car could general element and SUV, Sedan could be specific types of Cars which will have its own set of attributes.
* Specific type and generalized typed are tied together with foreign keys.
* As you generalize your structure more, you will contract the data model.

## 10. Data Modelling

A real estate investment company owns many shopping malls. Each mall contains many shops. To encourage rental of its shops, the company gives a negotiated discount to retailers who have shops in more than one mall. Each shop generates an income stream that can vary from month to month because rental is based on a flat rental charge and a negotiated percentage of sales revenue. Also, each shop has monthly expenses for scheduled and unscheduled maintenance. The company uses the data to compute its monthly net income per square meter for each shop and for ad hoc querying.

Below is the ER diagram

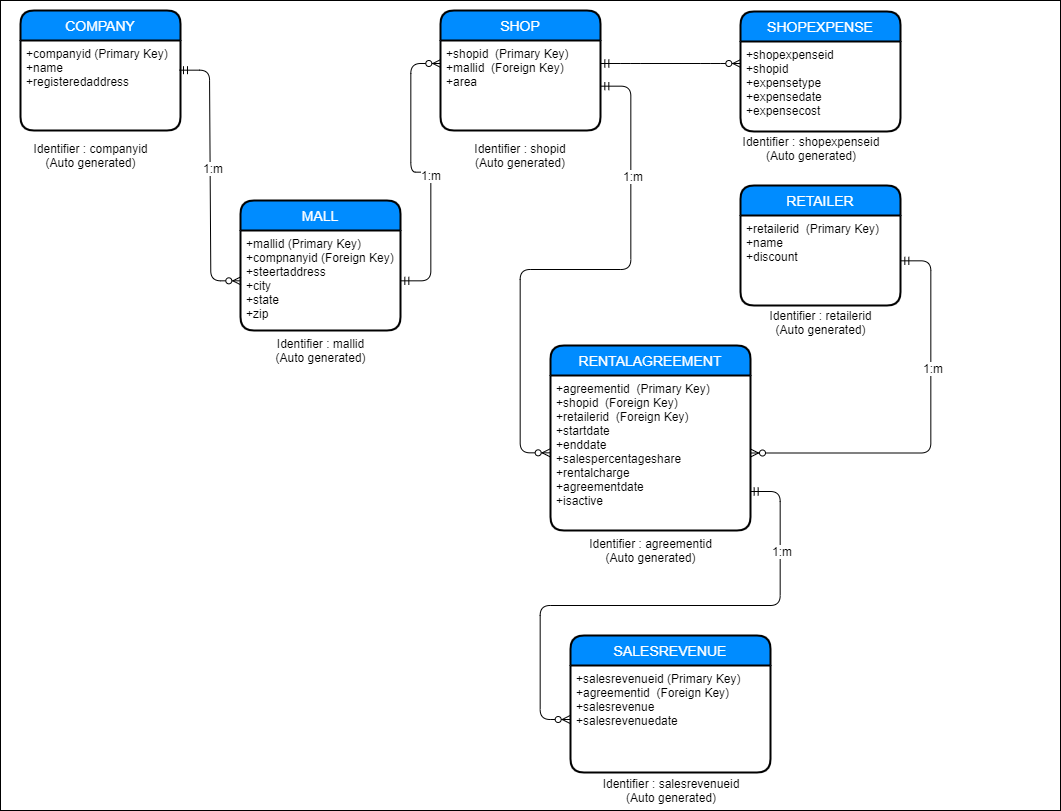


Figure 1: ER Diagram – Real Investment Company

Assumptions

* Retailers will report the sales revenue for every month for every shop, this value will get updated in the SALESREVENUE table. Month and year is getting captured via the salesrevenuedate field.
* Company will have at least one mall
* Mall will have at least one shop.
* Retailer and the mall owner company will sign a rental agreement for a particular shop or multiple shops for certain time frame like one year, 2 year etc.
  + The rental agreement will have rental dollar value to be paid, start date and end date
  + They will agree on the sales percentage share that retailer needs to pay to parent company for that shop.
  + For each retailer and shop there will one rental agreement active at any point of time. If there is not agreement active then shop is empty and no revenue is been generated.
* In case retailer is renting shops in multiple malls then in retailer table discount will get updated based on the negotiation between retailer and mall owner.
* Mall Owner Company will log the monthly scheduled/unscheduled maintenance cost in SHOPEXPENSE table. For each expense there will be an expense type.
* salespercentageshare and discount values would be between 0 to 1.

Data points about the design of the data model

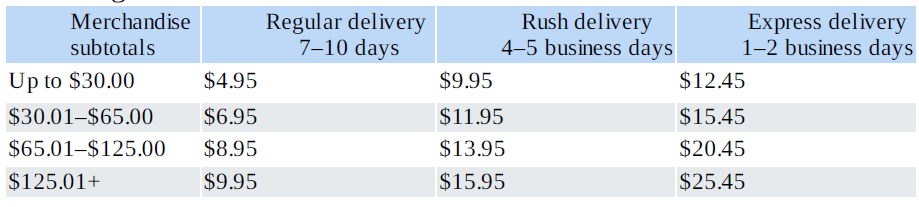
* All the primary key identifier I have made them as integer and auto generated as most of the modern databases support the same.
* In the RENTALAGREEMENT shopid, retailerid, startdate and enddate are unique (Composite unique key).
* IN SALESREVENUE agreementid, month, year are unique (Composite unique key).
* So for calculating monthly net income per square meter for each shop
  + You have to select a shop (shopid filter), join that with rental agreement that is active as well as with shop expense for maintenance cost, join with retailer and then join with SALES REVENUE table with filter of year and month on the salesrevenuedate column. From the shop expense you have to take sum of all expense costs with filter of year and month on expensedate.
  + Now you have following attributes based on above joins
    - shopId
    - area
    - expensecost (sum of all shop maintenance expenses)
    - salespercentageshare
    - rentalcharge
    - salesrevenue
    - retailerid
    - discount (retailer discount)
    - month
    - year
  + You can calculate monthly net income per square meter for each shop

(rentalcharge + (salesrevenue \* salespercentageshare) - expensecost) \* (1-discount)

area

## 12. Data Modelling

Draw a data model for the following freight table taken from a mail order catalog.



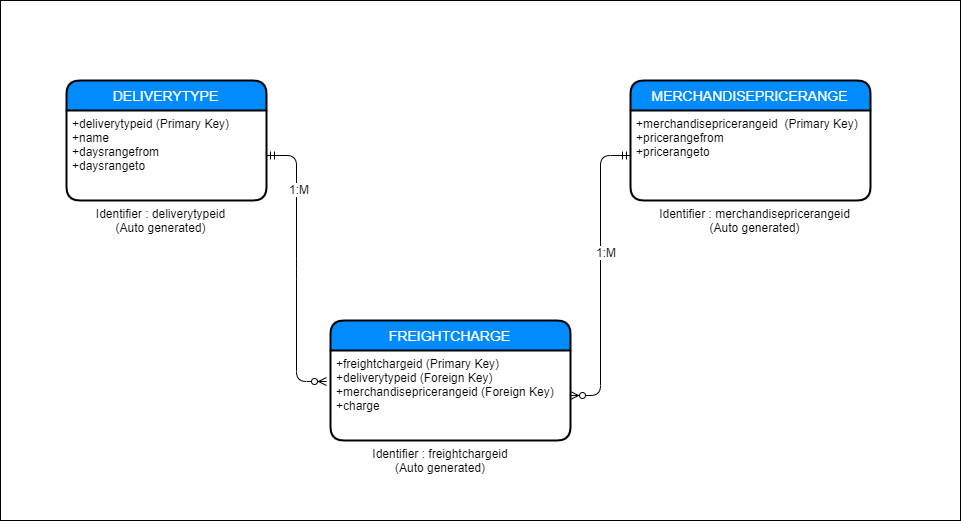


Figure 2: ER Diagram Freight Charges

Assumptions

* For Freight charges there needs to be merchandise price range and delivery type.

# Chapter 8

## 1. What is normalization, and what is its goal?

* Normalization is process of increasing the quality of your database design. Its goal is same as that of data modelling to create high fidelity design.
* Normalization is the process of efficiently organizing data in a database. There are 2 goals/objectives
  + **Eliminate redundant data** (storing the same data in more than one table)
  + Establishing proper data dependencies (storing related data in a table)

**Above goals help in reducing the amount of space a database consumes and ensure that data is logically stored.**

* **Normalization involves five forms which ranges from first normal form (1NF) to fifth normal form (5NF). After going through normalization you achieve following**
  + **Arranging data into logical groups such that each group describes a small part of the whole**
  + **Minimizing the amount of duplicated data stored in a database**
  + **Building a database in which you can access and manipulate the data quickly and efficiently without compromising the integrity of the data storage**
  + **Organizing the data such that, when you modify it, you make the changes in only one place**
* **In summary Normalization is part of successful database design. Without normalization, database systems can be inaccurate, slow, and inefficient and they might not produce the data you expect.**