

## Assessment-2

1.

a) Category of a product may change over a period of time.

Historical category information (current category as well as all old categories) has to be stored. Which SCD type will be suitable to implement this requirement? What kind of structure changes are required in a dimension table to implement SCD type 2 and type 3.



Category of a product that changes over a period of time are kept in a dimension called

slowly changing dimension(SCD).It is of 3 types:

1.SCD1

2.SCD2

3.SCD3

1.In SCD1 we overwrite the new value in place of old value.But there are drawbacks doing this.We cannot store Historical Data

2.In SCD2 history is also kept,and we increases number of rows to keep historical data as well as current data.

3.In SCD3 history is kept by increasing number of columns ,one column is for previous and one for current.

PM_PRIMARY_KEY	EMPLOYEE_ID	NAME	LOCATION
100	1001	STEVE	USA
101	1002	David	Russia

### Example:

If location of steve changed to India:

### SCD1:

PM_PRIMARY_KEY	EMPLOYEE_ID	NAME	LOCATION
100	1001	STEVE	India
101	1002	David	Russia

### SCD2:

PM_PRIMARY_KEY	EMPLOYEE_ID	NAME	LOCATION
----------------	-------------	------	----------


100	1001	STEVE	USA
101	1002	David	Russia
102	1001	STEVE	India

### SCD3:

PM_PRIMARYKEY	EMPLOYEE_ID	NAME	LOCATION	PREV_LOCATION
100	1001	STEVE	India	USA
101	1002	David	Russia	Null

1.

#### **b) What is surrogate key? Why it is required?**

 A surrogate key is a key which does not have any contextual or business meaning. It is manufactured “artificially” and only for the purposes of data analysis. The most frequently used version of a surrogate key is an increasing sequential integer or “counter” value (i.e. 1, 2, 3).


Basically, surrogate key is an artificial key that is used as a substitute for natural key (NK) defined in data warehouse tables.

We can use natural key or business keys as a primary key for tables. However, it is not recommended because of following reasons:

Natural keys (NK) or Business keys are generally alphanumeric values that is not suitable for index as traversing become slower. For example, prod123, prod231 etc

- Business keys are often reused after sometime. It will cause the problem as in data warehouse we maintain historic data as well as current data.

**c) Stores are grouped in to multiple clusters. A store can be part of one or more clusters. Design tables to store this store-cluster mapping information.**

 Here we will create two tables , one for cluster and another for Stores. Between cluster and stores there will be cardinality One to Many.

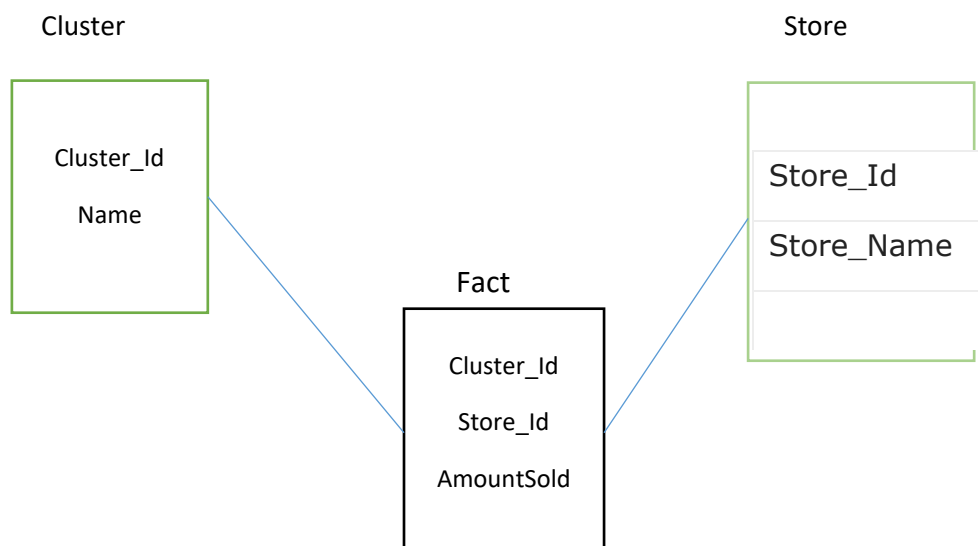
## Cluster

Cluster_Id	Name	Store_Id
100	C1	1
101	C2	1
102	C3	3
103	C4	2

## Store

Store_Id	Store_Name	AmountSold
1	S1	10000
2	S2	5000
3	S3	8000

## Dimension Modelling:



**d) What is a semi-additive measure? Give an example.**



Semi Additive measures are values that you can summarise across any related dimension except time.

For example, Sales and costs are fully additive; if We sell 100 yesterday and 50 today then we've sold 150 in total. We can add them up over time.

Stocks of a product in a store can be considered a semi additive measure because as stock decreases in time. if we had 100 in stock yesterday, and 50 in stock today, we're total stock is 50, not 150. It doesn't make sense to add up the measures over time, you need to find the most recent value.