Assessment

1. 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.

We can either use SCD type 2 or SCD type 3 based on storing limited history of data or full history of data. If we choose to use SCD type 2, that means we can store the full history of old categories of products. If SCD type 3 is used, only limited history say last 3 changes in categories of product has been stored as history.

In SCD type 2, we can add columns like “fromDate”, “toDate” and surrogate key . fromDate and toDate can specify the time period when specific product had particular category. Once the category is modified, in previous row with older category will be updated with some value in toDate column and a new row is generated with new category with fromDate (as when category was changed to new category) and toDate is assigned as null until next change in category.

In SCD type 3, we can add columns like “previous\_category” and “new\_category” in dimension table, which will have only the immediate history as previous\_category.

1. What is surrogate key? Why it is required?

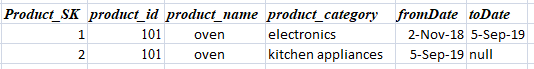
Surrogate key is sequentially generated key (can also be auto generated) attached with each and every record. They don’t have major meaning behind (since it doesn’t carry any business meaning regarding records its attached to)

It’s an artificial key used as substitute for natural key. For example, in above question given scenario, considering product\_id as a primary key, every time there is a change in a category of specific product\_id, a new row is generated with new category, implies in a single table we have duplicate product\_id so we use surrogate key as substitute of product\_id to determine the change in category.

Initially



After category changed



Here product\_SK is taken as surrogate key

1. 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.

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