**Problem Statement :**

We have sales data mart in MySQL and below are the table details :

DB name : labuser\_database

Table Details with ERD :

A close up of a map

Description automatically generated

Basic EDA details :

* Each supplier can supply multiple products, but one product can be supplied by only one supplier.
* Product table has a unit price which is relatable to buy price from the supplier.
* Order table has a unit price which is relatable to sell price which can be less or more than buy price of the product.

Below are the tasks which need to be performed:

* Import all these tables to hive tables. Hive DDLs can be inferred from table schemas in MySQL.
* For each sales month in the order table, get most profitable product, total profit from the product, profit % and its supplier Name and export this information to a Mysql table.

A product is profitable if sell price is greater than buy price from the supplier.

Final table should have these columns/values with storage as ORC.

* + Sales month
  + Most profitable product names
  + Supplier name
  + Total profit
  + Profit % up to 2 decimals
* For all the top 10 selling products (in terms of number of customers who have bought a product. Not total sales amount or total sales units of a product), provide 3 other product recommendations. Meaning, if a customer buys one of the top 10 selling product then get him 3 other product names which he may buy.
* If more than one product have same customer count then they will have same rank while getting top 10 selling products.
* If there is no sufficient data in orders table, then product recommendation may come with one or two recommendations.

Recommendation of products for each product can be derived from below logic.

* + Get all the customer IDs of a product.
  + Get all the product IDs which have been bought by above customers.
  + arrange the above products by total number of customers who bought them.
  + Select top 3 products using row\_number.

Final table should have below columns/values with storage as ORC:

* Product ID,
* Product Name
* Recommended product IDs ( a list of IDs)
* Recommended product Names (A list of names)