Dimension tables:

* + Dim\_Store ( Table for Store related information )
  + Dim\_Region (Table for Geo related Information)
  + Dim\_Item ( Table for Items related Information)
  + Dim\_Calendar ( Date Dimension)

Facts :

* + Fact\_instk\_dly ( In stock related information. What is the number of items in stock in a store on a given date and location)
  + fact\_scanv\_dly (Total number of items sold in a store on a give date and location)

Problem Statement :

We need to find stores which are going out of stocks on a day due to less number of in stocks maintained and more number of sales.

Store this information into a table in hive and export it to MYSQL.

Prerequisites :

All the target tables in Hive must be partitioned on some columns with ORC storage format. All the tables should have additional row\_insertion\_dttm column which will store timestamp of a record when it was loaded to the table. You can create stage tables in hive if required.

Refer attached excel sheet for detailed source /Target and problem statement information.



**ETL flow :**

A screenshot of a computer

Description automatically generated