**Objectives:**

Objective of this case study is to build a Data Warehouse as a Dimensional model in Hadoop for executing analytical queries.

**Data Source:**



Use the SQL Scripts in attached data files to create the mysql database.

**Assignment:**

1. Create and load an Operational Data Warehouse in MySQL using the scripts provided.
2. Add a last\_upd\_date column in Customer, Product, Employee Shipper tables to track changed records and update all historical records with current date.
3. Create a target Star Schema to store Orders data having following fact and dimensional tables in Apache Hive:

Orders Fact

Employee Dim

Customer Dim

Order Date Dim

Shipper Dim

1. Create Sqoop jobs for initial and incremental extracts from operational database and load to HDFS.
2. Create load scripts using hive to load tables as per below mentioned SCD logic

* Customer Dim & Shipper Dim tables will be stored using SCD Type1
* Employee Dim will have job title as SCD type 2 column and all other columns as SCD Type 1
* SCD Type 2 records tracked using eff\_begin\_date, eff\_end\_date and current\_version flag
* Product Dim will have all columns as SCD Type 2 columns

1. Test load by inserting / updating records in source tables and executing scripts to propagate the changes to target tables
2. Execute the sql commands using Spark SQL
3. Automate the load using oozie scheduler