

Dept. of CSE, BUET
Assignment on Big Data analytics (Data Warehouse)
CSE453, Jan 2021
Full marks: 20
Submission deadline: June 15, 2021

Design an e-commerce data warehouse for a nationwide chain of superstores of Bangladesh. There are many suppliers for the chain shop. The suppliers has sup-id, name, type of products to supply (clothes, machineries, food etc.), address (street, city, district). The chain shops serves the customers physically from the superstores and each superstore has its own system for all kinds of transactions (sale, procurement from the suppliers etc.). There are registered customers and each customer has customer id, name, NID, address (House no., street, thana, city, district, division and age-group. A customer can purchase many items in a single transaction with transaction id, transaction type (cash or card), timestamp id, time of the day, day of the week, date, week, month, year, quantity, unit price and total price. Each item has an item id, name, type, country of manufacture. You have to design and implement a warehouse for this chain of superstores to support policy decision process and knowledge discovery. The following reports will be generated from the warehouse.

Task 1: Design the architecture of the warehouse and explain the sources, preprocessing, noise reduction, transformation and uploading.

Task 2: Design the star schema for the warehouse **using the scenario and the data set** and explain how the data of the superstore database will be collected to the DW (source driven or destination driven).

Task 3: Implement the star schema in PostgreSQL and upload the given data into the database.

Task 4: three different cross tabulation for three different dimensions using total price/ quantity. Write SQL to find the cross-tabs.

Task 5: Find at least 5 important DSS reports (one for each dimension) as bar chart. Use cube operation in SQL to find the reports data.

Data set for the DW has been uploaded in the Google drive shared folder given below.

Submission instruction:

1. Upload the hand-written report describing tasks 1-5 in pdf format to moodle.
2. Add the system presentation video shared link to your Google drive into the shared Google sheet in

https://drive.google.com/drive/folders/1DGAO_WdmoEYx1sxyWI9EVRsffs11fWil