Data Warehousing

Politecnico di Torino

Craft beer

Problem specifications

A company responsible for the production and supply of craft beer aims at analyzing the data retrieved in the last 5 years.

The company has different factories where beer is produced and some storages located across the european country. Each factory is characterized by a specific number of production lines. The database stores the information of the capacity of each storage, measured in cubic meters. The system also contains information about the cities where storages and factories are located.

The beer produced by the company is divided into types. Each type is characterized by the alcoholic content, the IBU (International Bitterness Units) index and the types of used hop. A beer type can be produced with one or more types of hop. There are 10 hop types that are established by the company (e.g. 'cascade', 'columbus', 'zeus' ...). Beer is stocked in packs that have three different formats: 1, 3 or 6 bottles.

The database stores the date of the production of the beer packs and the date of arrival at the storage.

The company managers are interested in analyzing the number of liters of produced beer and the average revenue for each bottle during the last five years. The analysis must be carried on based on:

- factory, number of production lines, city, province, region, state
- storage, capacity, city, province, region, state where the storage is located
- beer type, alcoholic content, IBU, list of hop types
- pack format (1, 3, 6 bottles)
- production date, month, year
- date of arrival at the storage, day of the week, month of the year, 2-months, trimester, 4-months, 6-months, year

Design

Design the data warehouse to address the specifications and to efficiently answer to the provided frequent queries. Draw the conceptual schema of the data warehouse and the logical schema (fact and dimension tables).

Query

Write the following frequent queries using the extended SQL language.

a) Consider beers distributed in packs with the format: 6 bottles. Separately for production month and factory, analyze: the cumulative monthly number of produced liters from the beginning of the year, the monthly percentage of liters produced with respect to the total annual liters for the considered factory, assign a rank to the factories according to the total number of produced liters in decreasing order, separately for each production month.

b) Consider beers produced in 2018. Run the analysis separately for beer type, storage and month of arrival at the storage. Analyze: the average revenue for a bottle, the average daily (consider the day of arrival at the storage) number of produced bottles, the percentage of bottles of the considered beer type with respect to the total considering all the beer types.