3. Data Warehouse Design

To promote the excellence of the Made in Italy, the Italian government wants to analyze the exports of Italian wines, to gain insights into, for example, the most popular wines, the top exporter Italian areas, and the top importer foreign States.

The government has collected all purchasing orders from Italian wine producers. Each purchase order consists of: order date, amount in Euro, amount of wine in litres, type of wine, type of packaging (e.g., bottle, box, etc.), VAT number of the wine producer, foreign State of destination.

Government analysts want to create a datawarehouse from the purchase orders, to summarize and organize the data, and to answer some frequent queries.

In the datawarehouse, analysts want to integrate order data with additional context data they own, to enrich the analysis. In particular, each wine type is associated with the list of its quality certifications (e.g., DOC, DOP, DOCG). For instance, the "Barolo" wine is always associated with the DOCG quality certification, the "Nebbiolo" wine is always associated to the DOP and DOCG quality certifications. Each quality certification can be associated to one or more wine types. Some wine types can have no quality certifications.

Each purchase order is classified depending on the amount of wine: orders of less than 100 litres are considered small, up to 1000 litres are considered mid-sized, and the rest are big orders.

For each wine producer, analysts can get the full company data (name, address, province, region, geographical area of the region) from the VAT number.

The Italian government is interested in analyzing the total amount of wine (in litres) and the average price per litre depending on:

- month, two-month period, quarter, four-month period, half-year, year;
- province, region, geographical area (North, Centre, South Italy) of the wine producer;
- wine type
- quality certifications
- packaging type
- destination State and its continent
- order size (small, mid, big)

The datawarehouse includes data from year 2010 to 2013. The following are some of the frequent queries:

(a) Considering only wines with the "DOC" quality certification exported to Asia, select the average price per litre for each year, the percentage of litres exported in each year with respect to the total exported in all years, and the cumulative yearly amount of litres exported. The analisys must be performed separately for each packaging type.

```
SELECT
  PACKAGE, YEAR,
  SUM(EURO)/SUM(QTY),
  100*SUM(QTY) / SUM(SUM(QTY) OVER (PARTITION BY PACKAGE),
  SUM(SUM(QTY)) OVER (PARTITION BY PACKAGE
                      ORDER BY YEAR
                      ROWS UNBOUNDED PRECEDING)
FROM EXPORT E, PACKAGE P, TIME T, PLACE-Producer PL
WHERE E.CodPack=P.CodPack AND E.CodT=T.CodT AND
      P.CodPL=PL.CodPL AND DOC=TRUE AND CONTINENT='ASIA'
GROUP BY PACKAGE, YEAR
```

(b) For each region, select the average price per litre, the average number of litres exported per province, and the percentage of litres exported from each region with respect to the total of its geographical area (North, Centre, South). The analisys must be performed separately for each year.

```
SELECT
  REGION, YEAR,
  SUM(EURO)/SUM(QTY) AS AVERAGE_PRICE,
  SUM(QTY)/COUNT(DISTINCT PROVINCE) AS PROVINCE_AVERAGE,
  100*SUM(QTY) / SUM(SUM(QTY) OVER (PARTITION BY GEO_AREA, YEAR) AS PERC_GEO,
FROM EXPORT E, TIME T, PLACE-Producer PL
WHERE E.CodT=T.CodT AND P.CodPL=PL.CodPL
GROUP BY REGION, YEAR, GEO_AREA
```

(c) Rank all destination States for decreasing total quantity of wine. The analisys must be performed separately for each wine type, and considering only big-sized orders.

Design

(a) (6 points) Design the data warehouse to address the specifications and to efficiently answer to all the provided frequent queries.

(b) (8 points) Write the frequent queries (a) and (b) using the extended SQL language.