TPC-H

INSTRUCTIONS

1. This project is an individual project.

2. Submit the `answers.sql` with your answers in the space indicated to the "`Project 5: Submission`" folder in Luminus "`Files > Projects > Project 5: TPC-H`" by **Friday 17 April 2020, 18:30**.

3. There is no late submission possibility for this project.

In this project, we use the TPC-H benchmark[1] to create a data warehouse. TPC-H uses the example of a wholesale supplier business analytics application. You can read the selected excerpts from the TPC-H definition document in the "`tpch.pdf`" file available in Luminus "`Files > Projects > Project 5: TPC-H`".

The TPCH star schema is created by combining the LINEITEM and ORDERS tables into one FACT_LINEORDER fact table, by dropping the PARTSUPP table and dropping the COMMENT attributes in all tables, by adding a DIM_DATE dimension table, by merging the NATION and REGION tables into DIM_CUSTOMER and DIM_SUPPLIER tables, by dropping P_RETAILPRICE (this is likely to change too frequently to be in a dimension; the part price is better determined for an order as LO_EXTENDEDPRICE/LO_QUANTITY).

Download the "`table.zip`" and "`tpch.sql`" files from Luminus "`Files > Projects > Project 5: TPC-H`". Create a PostgreSQL database called "`cs4221_p5`".

Unzip "`table.zip`". The tables (files "`*.tbl`") are now located in a directory. Copy the location of the directory. For instance (in Windows): "`C:\Users\dcssb\Desktop\table`".

Open "`tpch.sql`" and find the following section.

```
COPY customer FROM 'customer.tbl' WITH DELIMITER '|';
COPY lineitem FROM 'lineitem.tbl' WITH DELIMITER '|';
COPY nation FROM 'nation.tbl' WITH DELIMITER '|';
COPY orders FROM 'orders.tbl' WITH DELIMITER '|';
COPY part FROM 'part.tbl' WITH DELIMITER '|';
COPY partsupp FROM 'partsupp.tbl' WITH DELIMITER '|';
COPY region FROM 'region.tbl' WITH DELIMITER '|';
COPY supplier FROM 'supplier.tbl' WITH DELIMITER '|';
```

Add the location of the directory in front of the tables names. For instance:

```
COPY customer FROM 'C:\Users\dcssb\Desktop\table\customer.tbl' WITH DELIMITER '|';
COPY lineitem FROM 'C:\Users\dcssb\Desktop\table\lineitem.tbl' WITH DELIMITER '|';
COPY nation FROM 'C:\Users\dcssb\Desktop\table\nation.tbl' WITH DELIMITER '|';
COPY orders FROM 'C:\Users\dcssb\Desktop\table\orders.tbl' WITH DELIMITER '|';
COPY part FROM 'C:\Users\dcssb\Desktop\table\part.tbl' WITH DELIMITER '|';
COPY partsupp FROM 'C:\Users\dcssb\Desktop\table\partsupp.tbl' WITH DELIMITER '|';
COPY region FROM 'C:\Users\dcssb\Desktop\table\region.tbl' WITH DELIMITER '|';
COPY supplier FROM 'C:\Users\dcssb\Desktop\table\supplier.tbl' WITH DELIMITER '|';
```

Run the SQL script "`tpch.sql`" in pgAdmin 4 Query Tool.

This creates a star schema (cf. the file `tpch_star_schema.pdf` available from Luminus) with a fact table `FACT_LINEORDER` and four dimension tables `DIM_PART`, `DIM_SUPPLIER`, `DIM CUSTOMER`, `DIM_DATE`. Note the difference between Gregorian calendar and ISO calendar[2] in the different attributes in the `DIM_DATE` table. Unless otherwise indicated, use the date or date component in Gregorian calendar.

---

[1] `http://www.tpc.org/tpch`
[2] `https://en.wikipedia.org/wiki/ISO_week_date`

Answer the following questions in SQL. Download the `answers.sql` from Luminus "`Files > Projects > Project 5: TPC-H`". Insert your answers in the space indicated in the file.

Question 1 [1 marks]

Display the order key, part name, supplier name, order date and extended price for each line order placed strictly before January 1994 by customer "`Customer#000000001`".

Question 2 [3 marks]

Print the total sum of the extended prices of all line orders placed by customers in Africa for each grouping set of customer nation and customer market segment for brand "`Brand#13`".

Question 3 [3 marks]

Print in ascending order of year, quarter, month and week of year of the order date the corresponding total sums of the extended prices of line orders placed between 1991 and 1992 (inclusive), the corresponding total sums of their discounted extended prices, the corresponding total sums of their discounted extended prices plus tax for each grouping set of year, quarter, month and week of year of the order date in this hierarchical order.

Question 4 [3 marks]

For each priority, print the number of orders in which at least one line item was received by the customer later than its committed date (`receiptdate > commitdate`) in order of the priorities (1 to 5).

– END OF PAPER –