TÉCNICO LISBOA

Sistemas de Informação e Bases de Dados

Lab 12: Decision Support

A university campus with multiple buildings has installed a metering infrastructure to measure energy consumption. In order to analyse evolutions of consumption, data obtained is stored in a Data Warehouse with a star schema as shown in Figure 1.

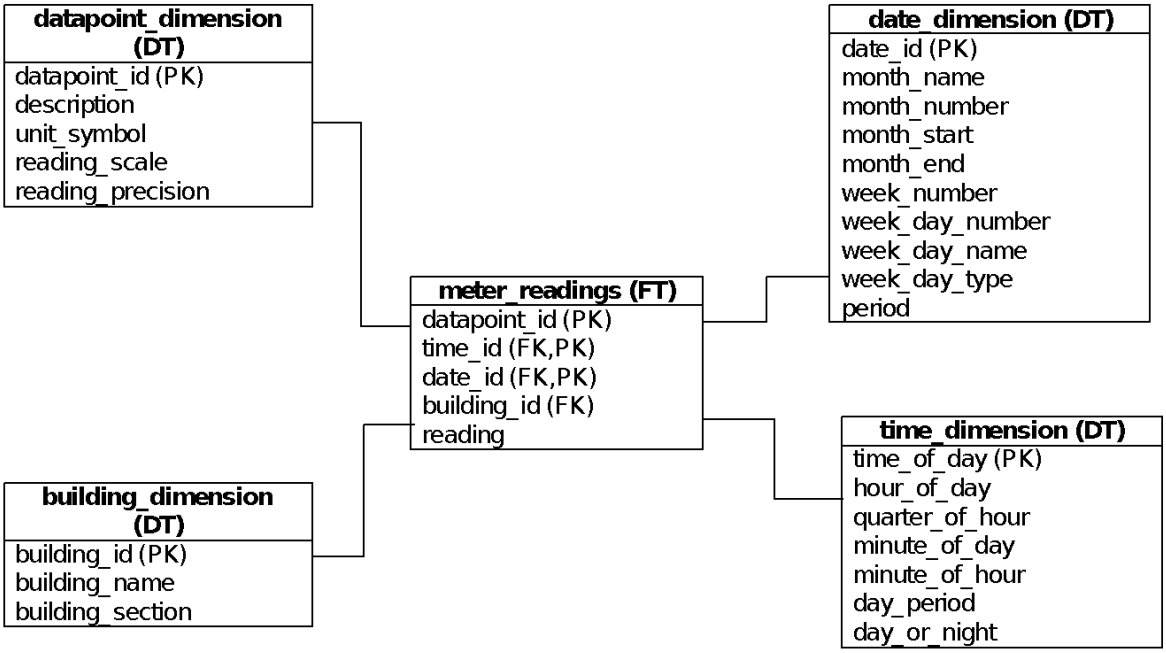

Figure 1. Star schema of the energy consumption measurements database

The fact table is the **meter_readings** table that contains the instantaneous watt-hour measurement events of each meter. The ***building_dimension*** table contains the details of each building. Each measurement refers to a single meter (a datapoint), and it refers to a building at a given instant in time. The ***time_dimension*** table characterizes every minute of a day. The ***date_dimension*** table characterizes each day over a year.

Database loading

The **energy-data-backup.sql** file contains a set of SQL instructions to create and load a database of meter readings of power consumption in the university campus.

To load the data into the Data Warehouse, you should follow these steps:

1. Transfer the **energy-data-backup.sql** file into Sigma and open a Postgres session by following an earlier lab guide (the guide for Lab 01 for example).
2. Execute the instructions contained in the file through the **\i energy-data-backup.sql**

The system outputs some 'OK' messages as it executes the instructions in the file.

⚠ Due to the size of the file, it is normal for the loading to take some minutes.

OLAP Queries

1. Create a query to determine the average consumption (campus-wide) per day of the week. Which days of the week have the highest and/or lowest consumption?

2. Create a query to determine the average consumption per building and per week during the last three weeks of the year.

3. ROLLUP from the result of the previous question, now grouping only by '**week_number**'. You can observe that consumption is falling in the last weeks of December.

4. Find out which buildings are the biggest energy consumers by calculating the average consumption per building and ordering the result.

5. DRILL DOWN the results of the previous question by day of the week (**week_day_name**) to determine on which day of the week do the biggest consumers consume the most.

6. DRILL DOWN the results of question 4 by time period (**day_period**) to understand on which periods of the day do the biggest consumers record the most energy consumption.

7. In order to analyse the distribution of the average consumption by building, by time period and by hour, present the results of average consumption by building by performing a 'ROLLUP' on the **day_period** and **hour_of_day** fields. You can use the **ROLLUP** clause or a combination of **UNION**s.

8. Create a table with the **results** of the previous question:

```
DROP TABLE IF EXISTS results;

CREATE TABLE results (
        building_name VARCHAR(20),
        day_period VARCHAR(20),
        hour_of_day INTEGER,
        avg_reading DOUBLE);

INSERT INTO results
SELECT building_name, day_period, hour_of_day, AVG(reading)
```

[(...) query of the previous question here]

9. Using the table ***results***, determine the periods whose average consumption is higher than the average consumption of all buildings. Tip: Since averages are already pre-calculated, you can use the **IS NULL** and the **IS NOT NULL** operators to get the results.

10. Using the **CUBE** operator, present a query having **building_name**, **day_period**, **hour_of_day** as vertices.