

Bases de Dados

1º semestre

Lab 12: Data Warehousing

Um campus universitário com 7 edifícios possui uma infra-estrutura de medição eletrónica dos consumos de energia. Para poder analisar a evolução dos consumos, os dados obtidos são guardados num Data Warehouse com um esquema em estrela de acordo como apresentado na Figura 1.

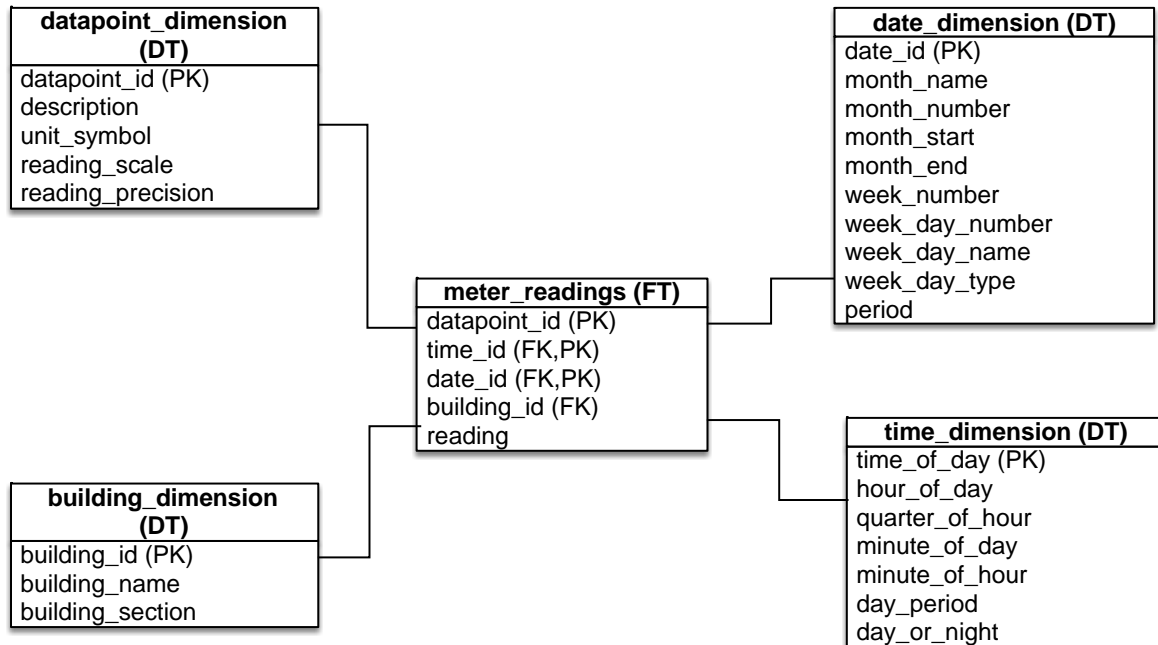


Figura 1. Esquema da base de dados de medições de consumos energéticos obtidos a partir de medidores de energia.

A tabela *meter_readings* é a tabela de factos, que contém os eventos de medição instantânea em watt-hora de cada edifício. A tabela *building_dimension* contém os detalhes de cada edifício. Cada medição refere-se a um medidor individualizado (datapoint) e está relacionada com um edifício num determinado instante. A tabela *time_dimension* caracteriza cada minuto de um dia. A tabela *date_dimension* caracteriza cada dia ao longo de um ano.

Neste exercício, devido a restrições de espaço apenas serão carregados os dados referentes ao mês de dezembro de 2013.

Carregamento de dados

O ficheiro *energy-data-backup.sql* contém um conjunto de instruções SQL para criar uma base de dados de leituras de medidores de consumos de energia elétrica do campus universitário.

Para carregar os dados para a Data Warehouse, deve dar os seguintes passos:

1. transferir o ficheiro *energy-data-backup.sql* para o Sigma e abrir um sessão no Postgres de acordo com as instruções de guias de laboratório anteriores (por exemplo Guia de Laboratório 1).
2. Executar as instruções no ficheiro *energy-sensor-data.sql* através do comando **source energy-data-backup.sql**

O sistema produz algumas mensagens de 'OK' à medida que executa as instruções do ficheiro.

Devido à dimensão do ficheiro é normal que o carregamento possa demorar alguns minutos.

Consultas OLAP

1. Apresente uma consulta para apurar o consumo médio de todo o campus por dia da semana. Quais os dias da semana em que se regista maior e/ou menor consumo?
2. Apresente uma consulta que permita apurar o consumo médio por edifício e por semana durante as três últimas semanas do ano.
3. Efetue o 'ROLLUP' a partir do resultado da alínea anterior, agrupando agora apenas por 'week_number'. Pode verificar que o consumo vai caindo nas últimas semanas de Dezembro.
4. Apure que edifícios são os maiores consumidores de energia, calculando o consumo médio por edifício e ordenando o resultado.
5. Efetue o 'DRILL DOWN' dos resultados da alínea anterior por dia da semana (**week_day_name**) para perceber em que dias da semana é que os maiores consumidores consomem mais energia.

6. Efetue o 'DRILL DOWN' dos resultados da alínea 4 por período horário (**day_period**) para perceber em que períodos horários é que os edifícios maiores consumidores consomem mais energia.
7. Para analisar a distribuição do consumo médio por edifício, por período horário e por hora, apresente agora os resultados do consumo médio por edifício efetuando um 'ROLLUP' sequencial sobre os campos **day_period** e **hour_of_day**. Utilize a cláusula **ROLLUP** do Posgres, caso esteja a usar uma versão superior ou igual ao Postgres 9.5, ou então recorra ao operador **UNION**.
8. Crie uma tabela **results** com os resultados da alínea anterior:

```
CREATE TABLE if not exists results(  
    building_name VARCHAR(20),  
    day_period VARCHAR(20),  
    hour_of_day INTEGER,  
    avg_reading DOUBLE PRECISION);  
  
INSERT INTO results  
Select building_name, day_period, hour_of_day, avg(reading) [...]  
[Query da alínea anterior]
```
9. Utilize a tabela **results** para determinar quais os períodos cuja média dos consumos médios é superior à média dos consumos de todos os edifícios. Sugestão: Uma vez que as médias se encontram já pré-calculadas, utilize **IS NULL** e **IS NOT NULL** para obter os resultados.
10. Tendo por base o operador **CUBE**, caso esteja a usar uma versão superior ou igual ao Postgres 9.5, apresente a consulta em posgres tendo como vértices **building_name**, **day_period**, **hour_of_day**.