‹epam›

EPAM Systems, RD Dep., RD Dep.

PROJECT REQUIREMENTS

SQL for Analysis

1   PROJECT SCOPE

In-scope:

1. Choose business area for your project.

2. Find **real** dataset (more preferably) or generate it in accordance with the selected business area.

3. Prepare data model, it should be Star, Snowflake or Fact Constellation Schema.

4. Create all necessary objects and load data into DB.

5. Analyze your data, prepare 3-5 reports about the most valuable business questions (from your point of view) which could be resolved with information from your schema.

2   GENERIC REQUIREMENTS

2.1   BUSINESS AREA

- There are no strict requirements for the subject of your project. It can be Sales or something related to Sport (various championships, Olympiads, etc.), Entertainment (cinema, games, etc.), Statistics, etc.

2.2   DATA

- The dataset must contain at least 2-3 thousand records. The larger your dataset is, the easier you will make an interesting analysis.

- The dataset must contain at least 1-2 numeric metrics.

2.3   DATA MODEL

- Your data model must include at least 1 fact table and 3 dimension tables.

2.4   REPORTING

- Each report is an SQL query that answers some business question. You can visualize the results of your analysis by adding diagrams, charts.

- Use window functions as much as possible during data analysis.

3 PROJECT RESULTS

Artifacts:

1. Text document (Microsoft Word) which describes all stages of the project:

   - your chosen business area
   - description of the initial dataset. Where did you find the data and what problems did you encounter while loading the data into DB
   - description of your data model. Briefly describe each table of your schema
   - any additional information

2. Text document which describes goals of your reports, SQL queries, result datasets and visualizations.

3. Entity Relation Diagram (ERD) of your schema (save it as PNG file).

4. DDL scripts to create all DB objects and scripts to load data into DB.

   To create backup of your schema run the following command in command line (change file path to **pg_dump.exe** if necessary):

   "**C:\PostgreSQL\12\bin\pg_dump.exe**" -h localhost -U postgres -F c -f **D:\sh.backup**

   where "**D:\sh.backup**" full path and name of **your** backup file (<u>can be changed</u>)

5. PPTX (Microsoft PowerPoint) or PDF file that you will use for project presentation. Your presentation must include the following chapters:

   - short description of business area
   - ERD of your schema
   - the results of data analysis
   - any additional information

Prepare all artifacts listed above, pack all files into zip-archive:
   - *Name this archive as **SQL_Analysis_Name_Surname_project.zip***
   - *Name of the PR should be: **SQL_Analysis_Project***