Test for DWH/BI Analyst

Instructions

Questions are in English. We will appreciate if you answer in English. Please do not use external information sources (books, internet), use only your current knowledge. Questions are organized into sections and ordered by difficulty (roughly); we start with easy questions and proceed to more complex. You can choose to answer questions in any order. If you do not know the answer, please indicate this by writing N/A.

Section A. Describe Yourself

Evaluate yourself on scale 1 (basic knowledge) to 5 (excellent knowledge). You can provide additional information. Fill other areas into empty lines, for example, other languages.

Area	Knowledge
SQL	4
ETL tools – list at most three 1. Talend 2. Xplenty 3. AWS Glue	4 3 3
Reporting / Analytical Tools – list at most three 1. Tableau 2. QlikView 3. Google Analytics	5 4 4

What database design principles do you know?

 I am not sure what you expecting, but I would say that while designing the database you have to make sure that the six objectives must be fulfilled:

1. Usability
2. Extensibility
3. Data Integrity
4. Performance

5. Availability
6. Security

I design and construct the database and I make sure these objectives are fulfilled.

The Basic Principles I know are:

1. Primary keys
2. Index
3. Foreign Keys
4. Normalization
5. Understanding the business to create tables
6. Parent and Child Tables
7. ER Diagrams etc

How large was the largest database that you worked with as developer/analyst (on the IT/delivery side), or user (on the business / client-side)? Specify the number of records/dimensions.

I work on Weather Data, Satellite Data, RTK Data and most of the databases are in Terra Bytes. I have recently done the RTK (Real-Time Kinematic) analysis with the MongoDB and PostgreSQL using JavaScript. The database was around 25 TB.

What are the most valuable books/courses have you completed in the last 3 years related to data warehouses and business intelligence?

I have completed my Master of Science in Data Engineering – 2017
AWS Machine Learning Course from Coursera – 2020
Marketing Analysis Course – 2020 from Coursera

Section B. Practical assignment

You received the following information in several .csv files:

- List of customers with their personal data
- List of accounts that belong to the customers
- List of transactions from/to these accounts

You need to:
- Model the data, so that all information can be stored in a relational database. The choice of data types, indices and relations is upon your decision

- Provide an SQL script to create such database schema · Populate the schema with the data from the .csv files.
- Provide a query that returns transactions for the users 345 and 1234, aggregated monthly, sorted by month, for the period from 15.02.2020 till 06.06.2020:

id_person	month	sum_of_transactions
1234	02.2020	####
1234	03.2020	####
1234	04.2020	####
1234	05.2020	####
1234	06.2020	####
345	02.2020	####
345	03.2020	####
345	04.2020	####
345	05.2020	####
345	06.2020	####

Please note, that these .csv files are extracted manually from the company's data warehouse and they might be malformed. Data inconsistencies might occur.

The resulting report MUST look exactly like shown in the table above in terms of column names and order.

Deliverables:
- Schema creation script (SQL)
- Database population script or ETL project (language or tool of your choice), including all data preparation steps
- SQL query to generate the report