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 the 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 PostgreSQL, DB2, MySQL	5
ETL tools – list at most three 1. Tivoli Directory Integrator 2. Pentaho 3.	 4 1
Reporting / Analytical Tools – list at most three 1. Cognos Analytics 2. Tableau 3. Microsoft Excel	 4 3 4

Java	2
Python	2
JavaScript	2
UML	4
GIT	3

What database design principles do you know?

1. Define the purpose of the database;
2. Depending on the purpose of the database it should be build as normalized or denormalized (performance factor is one of the key metrics that should be processed there) -> normalization | denormalization;
3. Check out the system entities and relationships -> create the ER diagram (define names (naming convention), data types, constraints, keys, indexes);
4. Separate the data into the logical blocks
5. Define the process to check the integrity of the data and whether all the data attributes apply to the ACID rule (atomicity, consistency, isolation and durability).
6. Test Users Concurrency & Security of the Data

*The database must be not overcomplicated, so the database desinged should be as easy as possible.

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.

more than 1TB , ~10^10 rows per table

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

Courses:
Udemy BI Couse (with statistics, database and python lessons , build on a real business example)
Udemy Automate the Boring Staff with Python (Udemy Course with the advanced python coding with real examples)
Tableau Internal courses (Desktop I, II, III)

Books:
J.K. Date "Introduction to the databases"
Wiegers "Software Requirements"
McConnel "Code Complete"

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