

DATA ENGINEER ASSESSMENT TASK

Table1 contains monthly export quantity figures of ProductX from various countries upto May 2021. Table2 is a list of countries and the Region those countries are grouped under.

ASSESSMENT TASKS

a) Create two sql db tables, 'Exports' and 'Countries', with data from Tables1 going into 'Exports' and data from Table2 going into 'Countries'. Link 'OriginCountry' and 'DestinationCountry' columns from Table1 (Exports) to 'Country' Column in Table 2 (Countries)

b) We now have June 2021 exports data (Table3). Load June 2021 data from Table3 into 'Exports' table using an ETL tool of your choice

c) Write sql queries to get the three reports mentioned below. You may store these queries also as Views if you like.

c.1 - Write query to produce yearly export figures from one country to the other. Output format as in Table4

c.2 - Write query to produce quarterly export figures from one country to the other. Output format as in Table5

c.3 - Write query to produce total yearly export figures by Region. Output format as in Table6

d) During the interview demonstrate task b described above. Do a demonstration of Table3 loaded as flat file and another demo separately of Table3 loaded as an Excel file.

e) Schedule (one of the two) packages in d as a job to run every 2 minutes. Demonstrate that the job is working. June 2021 data will be repeatedly added to 'Exports' but that is fine.

f) Describe method(s) for b and e if the data is on a shared Google drive or a similar cloud storage disk with the sql db being on-premises such as your laptop. Do a one slide presentation on the method(s). (Optional) If possible, please demo this during the interview.

g) Clean the data in Table 7 using python and store in db table named 'budget' with columns name 'Month', 'Year', 'Origin Country', 'Destination Country', 'Fiscal Year Period' and 'Budget Amount'. Perform the following steps:

g.1 - From 'Production date' column extract month and date and store in 'Month', 'Year' column respectively.

g.2 - From Origin and Destination country' column extract Origin and Destination country' separated by ':' and store in separate columns 'Origin country' and 'Destination country' respectively

g.3 - From 'Project Name' column extract month and fiscal year (e.g., May, FY2020). Concatenate these two (e.g., May 2021) and store in 'Fiscal Year Period' column

g.4 - From 'Amount' column store value as float in 'Budget Amount' column by removing whitespace and replace ',' with '.'.

EXPECTED DEMO/INTERVIEW SCHEDULE

Demo c - 5 mins

Demo b and d - 5 mins

Demo e - 5 mins

Discuss f - 5 mins

Discuss g - 5 mins

Q&A - 10 mins

If you need any clarifications on this assignment, please email your questions ahead of the interview. Send sql queries (c) and python script one day ahead of the interview. During the interview please be prepared to write some more basic SQL queries as per discussion.