

Case Study - Data Engineer

Instructions

Jolanthon Inc is a software development company that provides technology solutions to hospitals. It has been operating for a year and would like to provide services to help hospitals better understand their operations. You are the data engineer for the company and have been tasked with setting up a process to take the data from multiple application databases from the client's cloud instances, process the data and then store it in a data warehouse.

You have been provided with three datasets which contain data for doctors, patients and hospital visits over a three month period.

Task 1

Your first task is to write a python script to process the provided files and create a single dataset that can be used to perform the following analysis

1. Number of hospital visits per day over the period
2. Number of patients attended to per doctor per month showing their names
3. The ratio of female to male patient visits per month

NB:-

- Make use of any python libraries of your choice
- There should be three(3) outputs from this task
 - The python script file(s)
 - The final dataset from processing the three files, in csv format
 - A text file containing 3 queries on your final dataset to obtain the outlined analyses above (you don't need to submit the actual results)

Task 2

Outline a high level architecture/design to deploy and schedule the above python script to run once a day. Your design should read the data from an RDBMS and write the resulting dataset to a Data Warehouse. The architecture can be based on any cloud platform of your choice

NB:-

- Your outline should not exceed one page
- It can be text, a diagram or a mixture of both

We would expect this to take a few days of wall-clock time (**ideally no more than a week**), and a few hours of CPU time. As we are also very busy there will no doubt be delays on both sides as we work on this together. Please set expectations as you would in your job!

Finally, once you have finished this, you'll present the project to us in a Zoom group call.

All the best!