Assignment 4

Assignment on Azure Cloud Platform

Due by Jul 23, 2023

1. **Note:**
   This assignment needs to be done by using the Azure Cloud Platform. In this assignment, you will be working with Azure Data Factory, Azure SQL DB, Blob storage account and ADLS Gen2.

   Submit a compressed archive (zip, tar, etc.) of your code, along with screenshots (output/input commands with results). Please include a pdf document with answers to the questions below.

   For Part A: Please submit all screenshots showing deployed resources in your Azure portal, Azure Blob Storage, Azure Data Factory, ADLS Gen 2 and Azure SQL DB including your account information at the top right corner of the webpage. Include the successful pipeline runs screenshots with triggers.

   For Part B: Please submit all screenshots showing deployed resources in your Azure portal, Azure SQL DB and Query Editor screenshots where you run your code with output.

   Contact your TA for any questions related to this assignment or post clarification questions to the Piazza platform.

Part A:

1. [Marks: 5] Create a resource group in your Azure portal and deploy three resources. Azure Data Factory, Azure SQL DB and Blob storage account.

2. [Marks: 15] Now create a pipeline in Azure Data Factory and copy *gender_jobs_data.csv* file from the Blob storage account to Azure SQL DB. (First copy this file from your local machine to Blob Storage). See this https://docs.microsoft.com/en-us/azure/data-factory/tutorial-copy-data-portal for reference.

3. [Marks: 10] Explain the different types of triggers available in ADF. Now create a schedule trigger and run your pipeline every 3 minutes. Show 5 successful runs.

4. [Marks: 20] A client needs to replicate objects from ADLS Gen 2 in Canada Central to ADLS Gen 2 in West Europe. Let's say they want to do this in a bi-directional way. How can you set this up?

[Hint: This probably can be done using Azure Data Factory and Event Triggers. For eg; every time there is a new Blob on one side, it needs to be replicated on the other one]

PART B:

In this part, you will use Query Editor in Azure SQL DB and use the *gender_jobs_data.csv* table to perform the below queries.

Data input

For part B implementation, use the same table that is provided to you.

- *Gender_jobs_data.csv*

Implementation
You need to use Azure SQL Database for this part.

1. [Marks:5] In the *gender_jobs_data* table - Filter all the OCCUPATIONS in MAJOR_CATEGORY of Computer, Engineering, and Science for the YEAR 2013

2. [Marks:5] In the *gender_jobs_data* table - How many OCCUPATIONS exist in the MINOR_CATEGORY of Business and Financial Operations overall?

3. [Marks:5] In the *gender_jobs_data* table - Get all relevant information for bus drivers across all years

4. [Marks:5] In the *gender_jobs_data* table - Summarize the total number of WORKERS_FEMALE in the MAJOR_CATEGORY of Management, Business, and Financial by each year.

5. [Marks:5] In the *gender_jobs_data* table - What were the total earnings of male (TOTAL_EARNINGS_MALE) employees in the Service MAJOR_CATEGORY for the year 2015?

6. [Marks:5] In the *gender_jobs_data* table - How many female workers were in management roles in the year 2015?

7. [Marks:5] In the *gender_jobs_data* table - Compare the TOTAL_EARNINGS_MALE and TOTAL_EARNINGS_FEMALE earnings irrespective of occupation by each year

8. [Marks:5] In the *gender_jobs_data* table - How much money (TOTAL_EARNINGS_FEMALE) did female workers make as engineers in 2016?

9. [Marks:10] What is the total number of full-time and part-time female workers versus male workers year over year?