Assignment 5

Assignment on Azure Cloud Platform

Due by Aug 7, 2024

1. **Note:**

   Part B of this assignment can be done in groups of two students or individuals. Both students need to submit the assignment for both parts and provide both names, email and student ID at the top of the assignment.

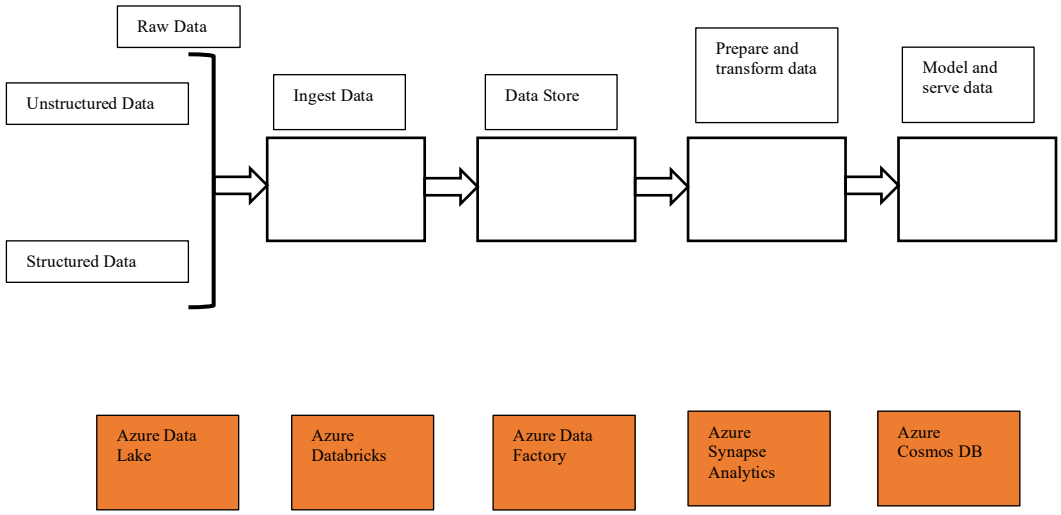
   Submit a compressed archive (zip, tar, etc.) of your code, along with the input and output files and screenshots (output/input commands with results). Please include your Azure Machine Learning Notebook with markdown.

   Also, include a pdf document with answers to the questions below. Please submit all screenshots showing deployed resources in your Azure portal provide an explanation for each step, also show final output screenshots.

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

PART A:

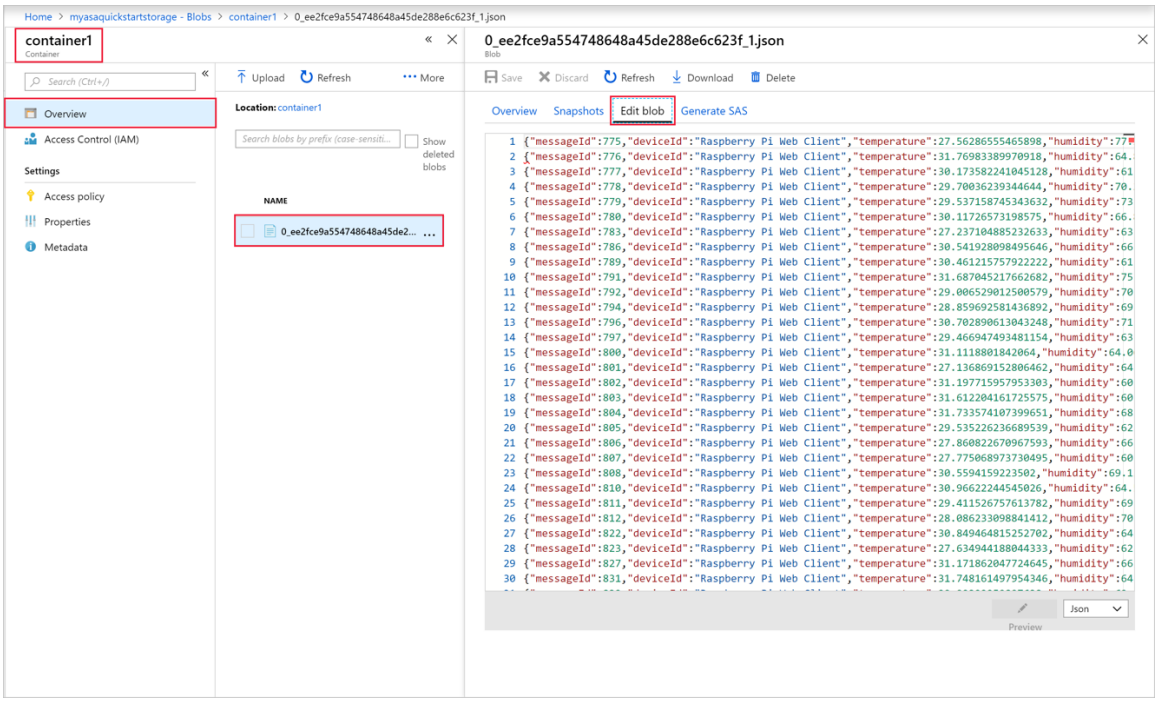
1. [Marks: 5] Explain below the 5 components shown in orange boxes. Explain which Azure components you will use in this big data architecture and why.



2. [Marks: 5] Explain how Stream Analytics works in Azure.

3. [Marks: 10] Deploy all the resources in Azure Portal. Implement a Stream Analytics job by using the Azure portal. See this for reference - https://learn.microsoft.com/en-us/azure/stream-analytics/stream-analytics-quick-create-portal

   For query use below:
   SELECT *
   INTO BlobOutput
   FROM IoTHubInput
   HAVING Temperature > 26
   See the below screenshot and show the top 50 results for your output. Make sure your screenshot shows your Azure account details.



Part B:

Data Input: Claim a dataset from Piazza - link. If the dataset is too large, you can take a subset of the data as well. No two groups can have the same dataset.

You need to solve a meaningful problem using this dataset.

Some problems to consider:

1. Fraud Detection System
2. Customer Churn Rate Prediction
3. Segmentation using Clustering
4. Recommendations with your Dataset
5. Sales Forecasting
6. Stock Price Predictions
7. Human Activity Recognition with Smartphones
8. Wine Quality Predictions
9. Breast Cancer Prediction

10. Sorting of Specific Tweets on Twitter etc.

Implement this part in Azure Machine Learning using Azure Notebook

1. [Marks: 10] Explain what problem you are going to solve using this dataset. Provide a brief overview of your problem statement. [Discuss your problem statement with your TAs if they approve then you can proceed with the next steps.]

2. [Marks: 15] Explain your dataset. Explore your dataset and provide at least 5 meaningful charts/graphs with an explanation.

3. [Marks: 15] Do data cleaning/pre-processing as required and explain what you have done for your dataset and why.

4. [Marks: 20] Implement 2 machine learning models, explain which algorithms you have selected and why. Compare them and show success metrics (Accuracy/RMSE/Confusion Matrix) as per your problem. Explain results.

5. [Marks: 20] Deploy a run-time pipeline for your dataset using Azure Designer Studio.
   Or
   Do hyperparameter tuning for your algorithms. Explain your results.
   Or
   Use Automated ML for your data set. Explain the best model results.