Assignment 5

Assignment on Azure Cloud Platform

PartA

1.

1. Azure Data Lake

Explanation: Azure Data Lake is a scalable and secure data lake for high-performance analytics workloads. It is designed to store vast amounts of structured, semi-structured, and unstructured data. The data stored in Azure Data Lake can come from various sources such as IoT devices, social media, logs, or any other data-generating system.

2. Azure Databricks

Explanation: Azure Databricks is an analytics platform optimized for the Microsoft Azure cloud services platform. It provides a collaborative environment for data engineers, data scientists, and business analysts to perform advanced data analytics and machine learning tasks. It's built on Apache Spark, which allows for large-scale data processing.

3. Azure Data Factory

Explanation: Azure Data Factory is a cloud-based data integration service that allows you to create data-driven workflows for orchestrating and automating data movement and data transformation. It enables you to move data from various sources to a centralized data store and transform the data as required.

4. Azure Synapse Analytics

Explanation: Azure Synapse Analytics is an integrated analytics service that accelerates time to insight across data warehouses and big data systems. It provides a single service for end-to-end analytics solutions, allowing for complex data analysis, data transformation, and big data analytics.

5. Azure Cosmos DB
Explanation: Azure Cosmos DB is a globally distributed, multi-model database service. It offers high availability, low latency, and scalability, which is essential for applications requiring real-time data access and distributed data storage.

Matching Azure Components with the Big Data Architecture:

1. Raw Data (Unstructured and Structured Data):

Matched with Azure Data Lake:
Why: Azure Data Lake is designed to store large volumes of raw data in its native format (structured, semi-structured, or unstructured). This makes it the ideal component for handling the initial raw data phase in the architecture, where both structured and unstructured data need to be stored before processing.

2. Ingest Data:

Matched with Azure Data Factory:
Why: Azure Data Factory is specialized in data ingestion and orchestration. It can pull data from various sources (e.g., databases, SaaS applications, file systems) and ingest it into the data lake or other storage systems. This makes it the best fit for the "Ingest Data" phase, where the primary task is to collect and move data.

3. Data Store:

Matched with Azure Cosmos DB:

Why: After data has been ingested and possibly transformed, it needs to be stored in a way that allows for quick retrieval and scalable access. Azure Cosmos DB provides the necessary capabilities for storing data in a globally distributed manner, ensuring low-latency access, which is critical for applications requiring fast and reliable data retrieval.

4. Prepare and Transform Data:

Matched with Azure Databricks:
Why: Azure Databricks is optimized for preparing, cleaning, and transforming large datasets. It provides a robust environment for data engineers and scientists to process data using Apache Spark. This makes it the best choice for the "Prepare and Transform Data" stage, where heavy data processing and transformations occur.

5. Model and Serve Data:

Matched with Azure Synapse Analytics:
Why: Azure Synapse Analytics allows for running complex queries, building data models, and performing advanced analytics on large datasets. It integrates with various services for serving data insights and visualizations, making it the ideal choice for the final stage where the data is modeled and served to users or applications.

2.

Azure Stream Analytics is a real-time data processing service in Microsoft Azure. It is designed to analyze and process high volumes of fast streaming data from multiple sources simultaneously. The service is fully managed, which means users do not need to worry about the underlying infrastructure, and it can scale automatically to handle the load.

Input:

Stream Analytics can ingest data from various real-time data streams such as Azure Event Hubs, Azure IoT Hub, and Azure Blob Storage. These are the sources of the streaming data that the job will process.

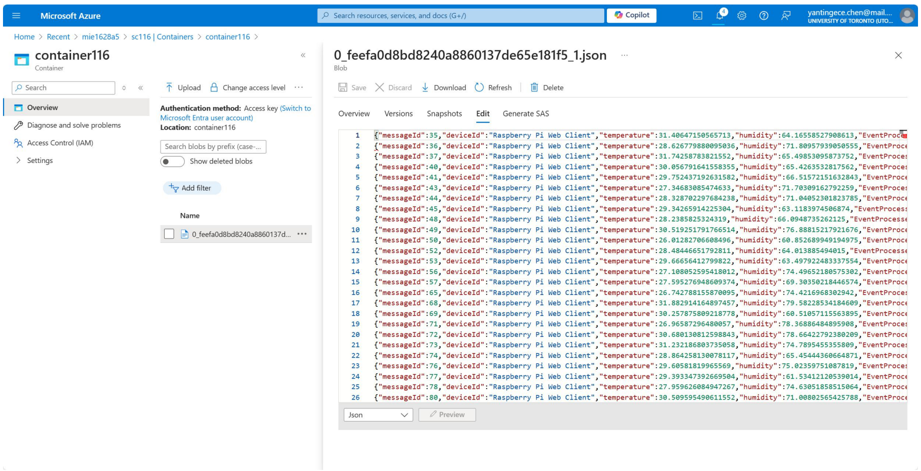
Query:

The core of Stream Analytics is the SQL-like query language that allows users to define the transformations, aggregations, filtering, and data manipulations they want to perform on the streaming data. This query can also perform temporal operations, like windowed aggregations, which are crucial for processing time-sensitive data. Example operations include joining streams, filtering data, computing aggregates, and detecting patterns in the data.

Output:

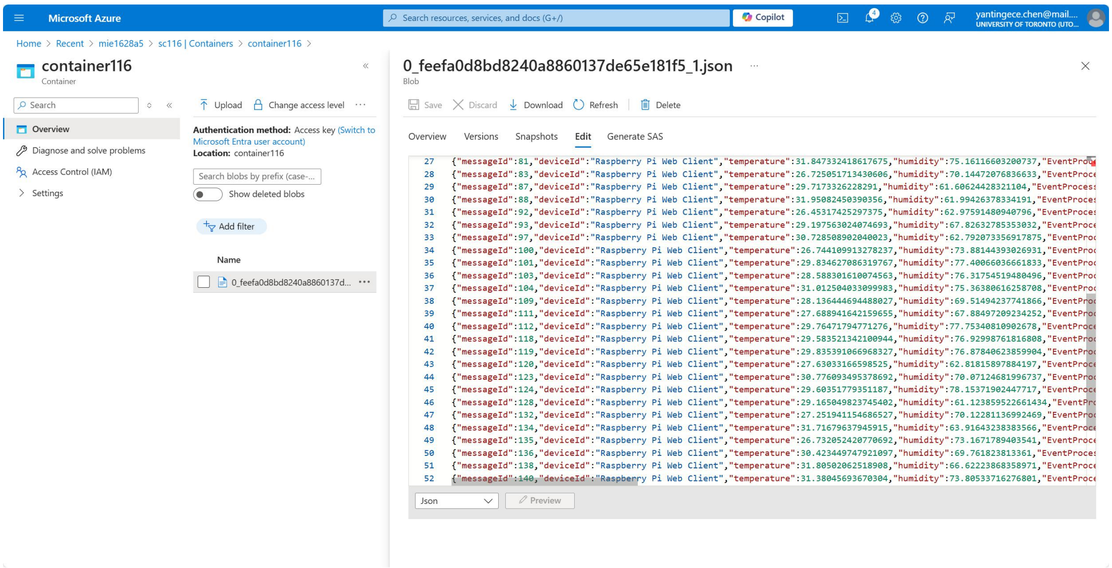
After processing, the transformed data can be sent to various destinations such as Azure SQL Database, Azure Cosmos DB, Azure Data Lake Storage, Azure Blob Storage, Power BI, or another Event Hub. This flexibility allows users to further analyze the processed data, visualize it, or trigger other processes based on it.

3.

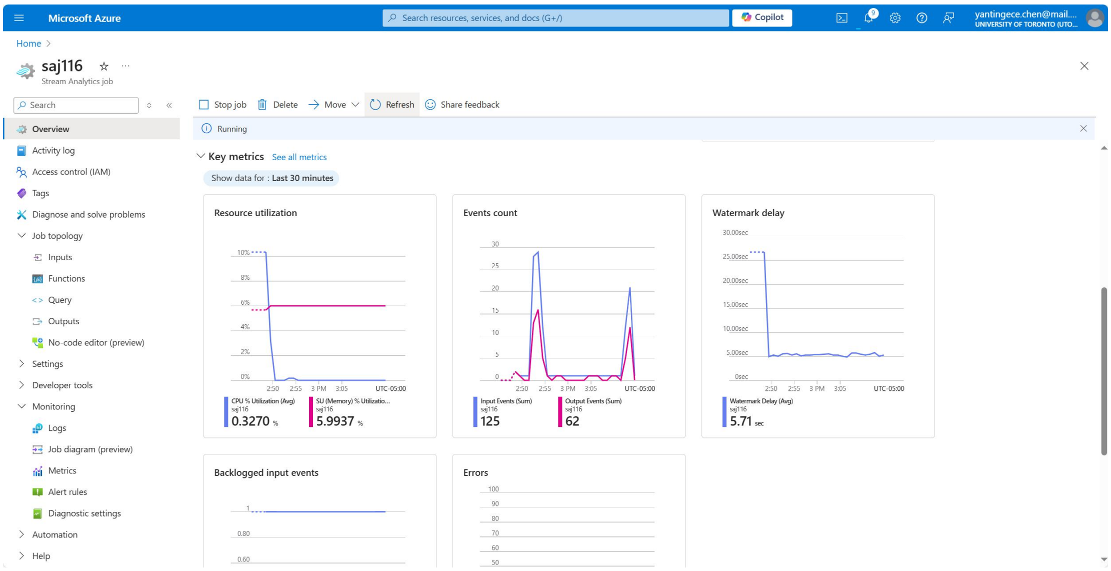

Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
yantingece.chen@mail....
UNIVERSITY OF TORONTO (UTO...
Home > Recent > mie1628a5 > sc116 | Containers > container116 >
container116
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Upload
Change access level
Search blobs by prefix (case-...
Show deleted blobs
Add filter
Name
0_feefa0d8bd8240a8860137d...
0_feefa0d8bd8240a8860137de65e181f5_1.json
Blob
Save
Discard
Download
Refresh
Delete
Overview
Versions
Snapshots
Edit
Generate SAS
Json
Preview

Authentication method: Access key (Switch to Microsoft Entra user account)
Location: container116

```
{"messageId":35,"deviceId":"Raspberry Pi Web Client","temperature":31.40647150565713,"humidity":64.16558527908613,"EventProce
{"messageId":36,"deviceId":"Raspberry Pi Web Client","temperature":28.626779880095036,"humidity":71.80957939050555,"EventProce
{"messageId":37,"deviceId":"Raspberry Pi Web Client","temperature":31.74258783821552,"humidity":65.49853095873752,"EventProces
{"messageId":40,"deviceId":"Raspberry Pi Web Client","temperature":30.056791641558355,"humidity":65.4263532817562,"EventProces
{"messageId":41,"deviceId":"Raspberry Pi Web Client","temperature":29.752437192631582,"humidity":66.51572151632843,"EventProce
{"messageId":43,"deviceId":"Raspberry Pi Web Client","temperature":27.34683085474633,"humidity":71.70309162792259,"EventProces
{"messageId":44,"deviceId":"Raspberry Pi Web Client","temperature":28.328702297684238,"humidity":71.04052301823785,"EventProce
{"messageId":45,"deviceId":"Raspberry Pi Web Client","temperature":29.34265914225304,"humidity":63.1183974506874,"EventProcess
{"messageId":48,"deviceId":"Raspberry Pi Web Client","temperature":28.2385825324319,"humidity":66.0948735262125,"EventProcesse
{"messageId":49,"deviceId":"Raspberry Pi Web Client","temperature":30.519251791766514,"humidity":76.88815217921676,"EventProce
{"messageId":50,"deviceId":"Raspberry Pi Web Client","temperature":26.01282706608496,"humidity":60.852689949194975,"EventProce
{"messageId":52,"deviceId":"Raspberry Pi Web Client","temperature":28.48446651792811,"humidity":64.013885494015,"EventProcesse
{"messageId":53,"deviceId":"Raspberry Pi Web Client","temperature":29.66656412799822,"humidity":63.497922483337554,"EventProce
{"messageId":56,"deviceId":"Raspberry Pi Web Client","temperature":27.108052595418012,"humidity":74.49652180575302,"EventProce
{"messageId":57,"deviceId":"Raspberry Pi Web Client","temperature":27.595276948609374,"humidity":69.30350218446574,"EventProce
{"messageId":65,"deviceId":"Raspberry Pi Web Client","temperature":26.742788155870095,"humidity":74.4216968302942,"EventProces
{"messageId":68,"deviceId":"Raspberry Pi Web Client","temperature":31.882914164897457,"humidity":79.58228534184609,"EventProce
{"messageId":69,"deviceId":"Raspberry Pi Web Client","temperature":30.257875809218778,"humidity":60.51057115563895,"EventProce
{"messageId":71,"deviceId":"Raspberry Pi Web Client","temperature":26.96587296480057,"humidity":78.36886484895908,"EventProces
{"messageId":72,"deviceId":"Raspberry Pi Web Client","temperature":30.680130812598843,"humidity":78.66422792380209,"EventProce
{"messageId":73,"deviceId":"Raspberry Pi Web Client","temperature":31.232186803735058,"humidity":74.7895455355809,"EventProces
{"messageId":74,"deviceId":"Raspberry Pi Web Client","temperature":28.864258130078117,"humidity":65.45444360664871,"EventProce
{"messageId":76,"deviceId":"Raspberry Pi Web Client","temperature":29.60581819965569,"humidity":75.02359751087819,"EventProces
{"messageId":77,"deviceId":"Raspberry Pi Web Client","temperature":29.393347392669504,"humidity":61.53412120539014,"EventProce
{"messageId":78,"deviceId":"Raspberry Pi Web Client","temperature":27.959626084947267,"humidity":74.63051858515064,"EventProce
{"messageId":80,"deviceId":"Raspberry Pi Web Client","temperature":30.509595490611552,"humidity":71.00802565425788,"EventProce
```


Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
yantingece.chen@mail....
UNIVERSITY OF TORONTO (UTO...
Home > Recent > mie1628a5 > sc116 | Containers > container116 >
container116
Container
Search
Overview
Diagnose and solve problems
Access Control (IAM)
Settings
Upload
Change access level
Authentication method: Access key (Switch to Microsoft Entra user account)
Location: container116
Search blobs by prefix (case-...
Show deleted blobs
Add filter
Name
0_feefa0d8bd8240a8860137d...
0_feefa0d8bd8240a8860137de65e181f5_1.json
Blob
Save
Discard
Download
Refresh
Delete
Overview
Versions
Snapshots
Edit
Generate SAS
Json
Preview

```
{"messageId":81,"deviceId":"Raspberry Pi Web Client","temperature":31.847332418617675,"humidity":75.16116603200737,"EventProc
{"messageId":83,"deviceId":"Raspberry Pi Web Client","temperature":26.725051713430606,"humidity":70.14472076836633,"EventProce
{"messageId":87,"deviceId":"Raspberry Pi Web Client","temperature":29.7173326228291,"humidity":61.60624428321104,"EventProcess
{"messageId":88,"deviceId":"Raspberry Pi Web Client","temperature":31.95082450390356,"humidity":61.99426378334191,"EventProces
{"messageId":92,"deviceId":"Raspberry Pi Web Client","temperature":26.45317425297375,"humidity":62.97591480940796,"EventProces
{"messageId":93,"deviceId":"Raspberry Pi Web Client","temperature":29.197563024074693,"humidity":67.82632785353032,"EventProce
{"messageId":97,"deviceId":"Raspberry Pi Web Client","temperature":30.728508902040023,"humidity":62.792073356917875,"EventProc
{"messageId":100,"deviceId":"Raspberry Pi Web Client","temperature":26.744109913278237,"humidity":73.88144393026931,"EventProc
{"messageId":101,"deviceId":"Raspberry Pi Web Client","temperature":29.834627086319767,"humidity":77.40066036661833,"EventProc
{"messageId":103,"deviceId":"Raspberry Pi Web Client","temperature":28.588301610074563,"humidity":76.31754519480496,"EventProc
{"messageId":104,"deviceId":"Raspberry Pi Web Client","temperature":31.012504033099983,"humidity":75.36380616258708,"EventProc
{"messageId":109,"deviceId":"Raspberry Pi Web Client","temperature":28.136444694488027,"humidity":69.51494237741866,"EventProc
{"messageId":111,"deviceId":"Raspberry Pi Web Client","temperature":27.688941642159655,"humidity":67.88497209234252,"EventProc
{"messageId":112,"deviceId":"Raspberry Pi Web Client","temperature":29.76471794771276,"humidity":77.75340810902678,"EventProce
{"messageId":118,"deviceId":"Raspberry Pi Web Client","temperature":29.583521342100944,"humidity":76.92998761816808,"EventProc
{"messageId":119,"deviceId":"Raspberry Pi Web Client","temperature":29.835391066968327,"humidity":76.87840623859904,"EventProc
{"messageId":120,"deviceId":"Raspberry Pi Web Client","temperature":27.63033166598525,"humidity":62.81815897884197,"EventProce
{"messageId":123,"deviceId":"Raspberry Pi Web Client","temperature":30.776093495378692,"humidity":70.07124681996737,"EventProc
{"messageId":124,"deviceId":"Raspberry Pi Web Client","temperature":29.60351779351187,"humidity":78.15371902447717,"EventProce
{"messageId":128,"deviceId":"Raspberry Pi Web Client","temperature":29.165049823745402,"humidity":61.123859522661434,"EventPr
{"messageId":132,"deviceId":"Raspberry Pi Web Client","temperature":27.251941154686527,"humidity":70.12281136992469,"EventProc
{"messageId":134,"deviceId":"Raspberry Pi Web Client","temperature":31.71679637945915,"humidity":63.91643238383566,"EventProce
{"messageId":135,"deviceId":"Raspberry Pi Web Client","temperature":26.732052420770692,"humidity":73.1671789403541,"EventProce
{"messageId":136,"deviceId":"Raspberry Pi Web Client","temperature":30.423449747921097,"humidity":69.761823813361,"EventProces
{"messageId":138,"deviceId":"Raspberry Pi Web Client","temperature":31.80502062518908,"humidity":66.62223868358971,"EventProce
{"messageId":140,"deviceId":"Raspberry Pi Web Client","temperature":31.38045693670304,"humidity":73.80533716276801,"EventProce
```


Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
yantingece.chen@mail....
UNIVERSITY OF TORONTO (UTO...
Home >
saj116
Stream Analytics job
Search
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Job topology
Inputs
Functions
Query
Outputs
No-code editor (preview)
Settings
Developer tools
Monitoring
Logs
Job diagram (preview)
Metrics
Alert rules
Diagnostic settings
Automation
Help
Stop job
Delete
Move
Refresh
Share feedback
Running
Key metrics
See all metrics
Show data for : Last 30 minutes
Resource utilization
10%
8%
6%
4%
2%
0%
2:50
2:55
3 PM
3:05
UTC-05:00
CPU % Utilization (Avg)
saj116
0.3270 %
SU (Memory) % Utilizatio...
saj116
5.9937 %
Events count
30
25
20
15
10
5
0
2:50
2:55
3 PM
3:05
UTC-05:00
Input Events (Sum)
saj116
125
Output Events (Sum)
saj116
62
Watermark delay
30.00sec
25.00sec
20.00sec
15.00sec
10.00sec
5.00sec
0sec
2:50
2:55
3 PM
3:05
UTC-05:00
Watermark Delay (Avg)
saj116
5.71 sec
Backlogged input events
1
0.80
0.60
Errors
100
90
80
70
60
50