**Lab 4 – Website Clickstream Ingestion into ADLS**

* Name : Xihai Ren
* Course: CST8921 010
* Professor : Ragini Madaan
* Date : 2025/06/02

**CST8921 – Cloud Industry Trends**

**Lab 4 – Website Clickstream Ingestion into ADLS**

## 

## Objective

## Simulate website click events (JSON), stream them via Event Hubs, process and store them as JSON files in ADLS Gen2 using Azure Functions.

## Prerequisites:

* Basic familiarity with azure event hub
* A computer with internet access
* Windows or mac machine
* Web browser
* Cloud portal access with any cloud service providers (AWS, Azure, GCP)

**Lab Activity Overview:**

## [User Click Events]

## ↓

## [Event Hubs] ---> [Azure Function (Event Hub trigger)]

## ↓

## [ADLS Gen2 JSON storage]

## Azure Services Used

| **Service** | **Purpose** |
| --- | --- |
| Event Hubs | Ingest real-time events (clicks) |
| Azure Function | Parses + writes to ADLS |
| ADLS Gen2 | Stores raw event JSON per user/date |

### Step 1: Event Hub Setup

 **Namespace**: clickstream-namespace

 **Event Hub**: click-events

 Set up **"Send"** policy or use **Managed Identity** for access

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

### **Step 2: Create ADLS Gen2 Container**

* Storage Account: **clickstreamstorage**
* Container: **click-events-raw**
* Folder structure (auto-created by Function):

**/click-events-raw/yyyy/mm/dd/hh/mm/uuid.json**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Step 3. Create Azure Function App**

**Runtime: Python or C#**

**Trigger: Event Hub**

**Function Role: Storage Blob Data Contributor on ADLS**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

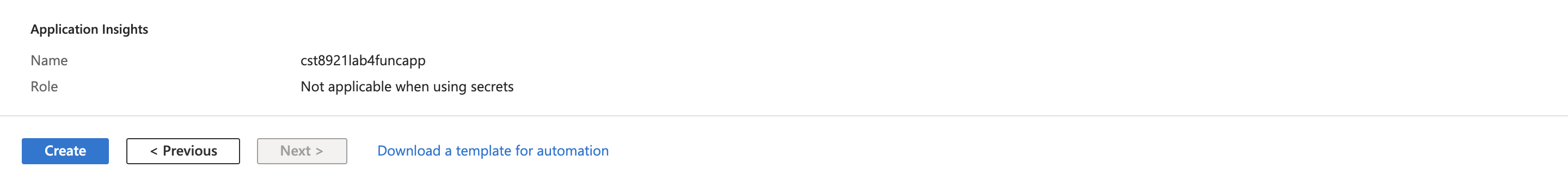
AI-generated content may be incorrect.

A white background with black dots

AI-generated content may be incorrect.

A close-up of a person

AI-generated content may be incorrect.



A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

### Step 4: Function Code (Python)

### import json, logging

### from datetime import datetime

### from azure.storage.blob import BlobServiceClient

### from uuid import uuid4

### # ADLS connection string (use Key Vault in production)

### BLOB\_CONN\_STR = "<your-storage-conn-string>"

### CONTAINER = "click-events-raw"

### def main(event: func.EventHubEvent):

### body = json.loads(event.get\_body().decode('utf-8'))

### ts = datetime.utcnow()

### blob\_path = f"{ts.year}/{ts.month:02d}/{ts.day:02d}/{ts.hour:02d}/{ts.minute:02d}/{uuid4()}.json"

### blob\_service = BlobServiceClient.from\_connection\_string(BLOB\_CONN\_STR)

### container\_client = blob\_service.get\_container\_client(CONTAINER)

### container\_client.upload\_blob(blob\_path, json.dumps(body), overwrite=False)

### logging.info(f"Stored event: {blob\_path}")

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Step 5: Clean the resources and delete all resources created in lab

## Important Notes:

For grading prepare a lab report with your findings and analysis and share that in an Assignments tab in Brightspace.