

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**  
**KHOA HỆ THÔNG THÔNG TIN**



**BÁO CÁO BÀI TẬP ĐỒ ÁN**  
**ĐỀ TÀI: PHÂN TÍCH DỮ LIỆU LUỒNG VÀ TRỰC**  
**QUAN KẾT QUẢ LÊN POWER BI**

**LỚP: IS402.P11.HTCL**

**Giáo viên hướng dẫn:**

Ths. Hà Lê Hoài Trung

**Sinh viên thực hiện:**

Nguyễn Bình Nguyễn - 21522391

Ché Duy Khang – 21522187

Nguyễn Thế Vinh – 21522794

**THÀNH PHỐ HỒ CHÍ MINH, 2024**

# MỤC LỤC

1. GIỚI THIỆU BÀI TOÁN .....	6
2. GIỚI THIỆU ĐỀ TÀI .....	6
3. CƠ SỞ LÝ THUYẾT.....	7
3.1.    Giới thiệu chung:.....	7
3.2.    Cơ chế xử lý: .....	7
3.2.1.    Đầu vào (Inputs) .....	8
3.2.2.    Đầu ra (Outputs) .....	9
3.2.3.    Truy vấn (Queries) .....	9
3.3.    Có hai cách sử dụng Azure Stream Analytics:.....	10
3.4.    Một số tính năng của Azure Stream Analytics:.....	10
3.5.    Ứng dụng dịch vụ Cloud .....	11
3.6.    Định dạng lưu trữ: .....	12
4. MÔ HÌNH DỮ LIỆU .....	12
4.1.    Quản lý hiệu suất.....	12
4.2.    Select into.....	13
4.3.    Lọc dữ liệu (WHERE) .....	13
4.4.    Tổng hợp qua các cửa sổ thời gian.....	14
5. DEMO.....	14
5.1.    Đăng nhập vào Azure.....	14
5.2.    Tạo Stream Analytics Job .....	19
5.3.    Cấu hình Stream Analytics Job để truy vấn dữ liệu.....	20
5.4.    Chạy Stream Analytics Job và tiến hành phân tích bằng Power BI.....	22
5.5.    Đo tốc độ đọc ghi .....	26

5.6.	Share report cho các thành viên trong nhóm.....	27
5.7.	Chỉnh sửa layout của Dashboard cho Mobile .....	28
5.8.	Tạo luồng xử lý dữ liệu tự động ETL .....	29
6.	THAM KHẢO .....	37

# DANH MỤC HÌNH

Hình 1: Sơ đồ xử lý sự kiện của Azure Stream Analytics.....	8
Hình 2: Trang chủ Microsoft Azure portal .....	15
Hình 3: Tạo mới một Event Hubs .....	15
Hình 4: Cấu hình Event Hubs .....	15
Hình 5: Giao diện khi tạo xong và mở Event Hubs .....	16
Hình 6: Nhấn tạo Event Hubs .....	16
Hình 7: Chọn Shared access polices và tạo key .....	17
Hình 8: Copy connection string .....	17
Hình 9: Dán connection string vào value của file telcodagen.exe.config .....	17
Hình 10: Tạo Storage Account.....	18
Hình 11: Chọn cấu hình Storage Account .....	18
Hình 12: Chạy lên .\telcodatagen.exe 1000 0.2 2 .....	19
Hình 13: Tạo Stream Analytics jobs .....	19
Hình 14: Cấu hình Stream Analytics Job.....	20
Hình 15: Tạo input cho Stream Analytics Job .....	20
Hình 16: Tạo output thứ nhất Blob .....	21
Hình 17: Tạo output thứ hai Power BI .....	21
Hình 18: Tạo câu truy vấn và save .....	22
Hình 19: Chọn start job ở Stream Analytics Job .....	22
Hình 20: Giao diện sau khi start thành công.....	23
Hình 21: Vào Storage Account để lấy data.....	23
Hình 22: Import dữ liệu .....	24
Hình 23: Tạo các report.....	24
Hình 24: Tạo các report.....	25
Hình 25: Import dữ liệu .....	25
Hình 26: Dashboard thứ hai để phân tích dữ liệu Hình 26:.....	26
Hình 27: Vào Stream Analytics Job để chọn biểu đồ Event Count .....	26
Hình 28: Chọn thời gian khoảng một phút và kết quả .....	27

<i>Hình 29: Share report .....</i>	27
<i>Hình 30: Chọn nút layout cho mobile .....</i>	28
<i>Hình 31: Dashboard khi xem trên mobile .....</i>	29
<i>Hình 32: Tạo mới 1 pipeline .....</i>	29
<i>Hình 33: Chọn activity là data flow .....</i>	30
<i>Hình 34: Chọn source là Blob Storage .....</i>	30
<i>Hình 35: Format là json .....</i>	31
<i>Hình 36: Thiết lập đường dẫn cho source .....</i>	32
<i>Hình 37: Sử dụng Derived Column .....</i>	33
<i>Hình 38: Tạo 2 cột mới là Date và Time .....</i>	33
<i>Hình 39: Thiết lập đích đến là Sink .....</i>	34
<i>Hình 40: Chọn dataset là Azure Blob Storage .....</i>	34
<i>Hình 41: Chọn format là CSV .....</i>	35
<i>Hình 42: Chọn output ra 1 file tên là CallsCSV.csv .....</i>	35
<i>Hình 43: Mapping các cột .....</i>	35
<i>Hình 44: Tạo trigger .....</i>	36
<i>Hình 45: Publish tất cả: .....</i>	36
<i>Hình 46: File CallsCSV.csv .....</i>	36

## 1. GIỚI THIỆU BÀI TOÁN

- Bài toán này thuộc nhóm các bài toán liên quan đến xử lý, lưu trữ và trực quan hóa dữ liệu luồng (streaming data).
- Việc xử lý, phân tích và trực quan hóa dữ liệu luồng mang lại nhiều lợi ích như: ra quyết định nhanh chóng, phát hiện xu hướng, cảnh báo sự cố kịp thời, v.v.
- Loại dữ liệu: Dữ liệu trong bài toán này là dữ liệu luồng, bán cấu trúc (JSON), được lấy từ việc tự động tạo ra các cuộc gọi ngẫu nhiên.
- Các thành phần trong mô hình:
  - PaaS (Platform as a Service)
    - Azure Event Hubs: Để thu thập và xử lý dữ liệu luồng.
    - Azure Stream Analytics: Để phân tích và xử lý dữ liệu luồng
    - Azure Data Lake Storage: Để lưu trữ dữ liệu.
  - SaaS (Software as a Service)
    - Power BI: Để trực quan hóa dữ liệu sau khi xử lý.

## 2. GIỚI THIỆU ĐỀ TÀI

Ngày nay, lượng dữ liệu thời gian thực không lồ được tạo ra bởi các ứng dụng kết nối, thiết bị và cảm biến Internet of Things (IoT), cùng với nhiều nguồn khác. Sự phát triển của các nguồn dữ liệu trực tuyến đã khiến khả năng tiêu thụ và đưa ra quyết định thông tin từ những dữ liệu này gần như ngay lập tức trở thành một yêu cầu hoạt động đối với nhiều tổ chức.

Một số ví dụ điển hình về khái lượng công việc xử lý dữ liệu thời gian thực:

- Các nhà máy sản xuất sử dụng dữ liệu từ các cảm biến IoT để theo dõi các tài sản có giá trị cao từ xa.
- Giao dịch thẻ tín dụng được kiểm tra thời gian thực để phát hiện và ngăn chặn các hoạt động có khả năng gian lận.

- Cửa hàng trực tuyến phân tích dữ liệu clickstream thời gian thực để đưa ra gợi ý sản phẩm cho người tiêu dùng khi họ duyệt trang web.

Azure Stream Analytics cung cấp một bộ xử lý dữ liệu thời gian thực dựa trên đám mây, có thể sử dụng để lọc, tổng hợp và xử lý luồng dữ liệu thời gian thực từ các nguồn khác nhau.

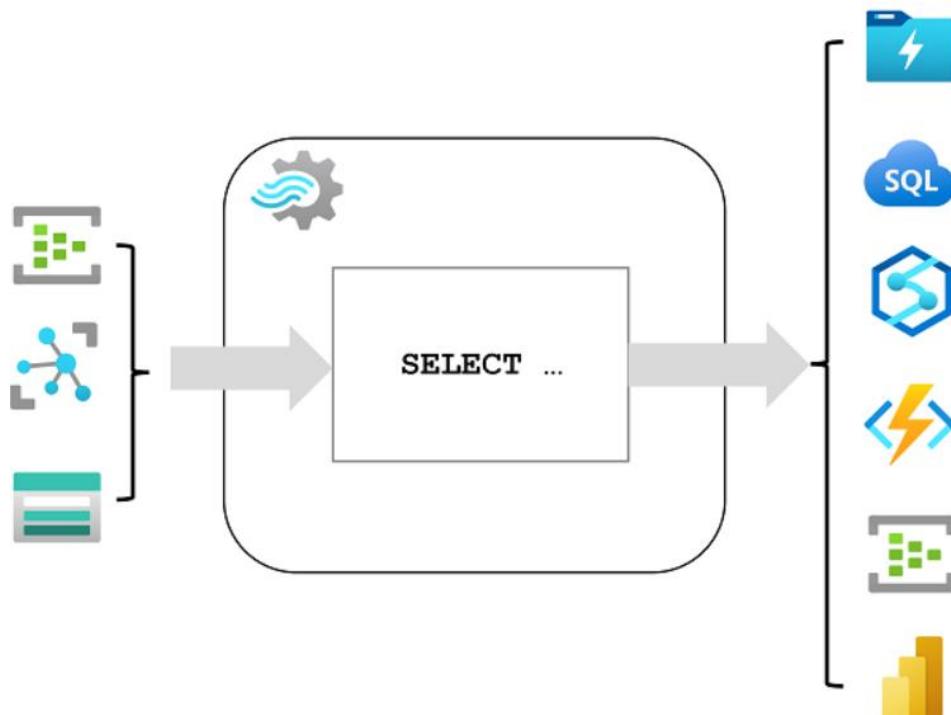
## 3. CƠ SỞ LÝ THUYẾT

### 3.1. Giới thiệu chung:

Azure Stream Analytics là một dịch vụ trong nền tảng đám mây của Microsoft Azure, được thiết kế để xử lý và phân tích dữ liệu dòng (streaming data) từ nhiều nguồn khác nhau như cảm biến, máy chủ, thiết bị IoT và ứng dụng trực tuyến. Dịch vụ này cung cấp khả năng xử lý dữ liệu thời gian thực, đồng thời và có khả năng mở rộng linh hoạt để xử lý các luồng dữ liệu lớn.

### 3.2. Cơ chế xử lý:

Azure Stream Analytics hoạt động theo cơ chế xử lý luồng dữ liệu (stream processing) theo thời gian thực, với mô hình truyền tải và xử lý sự kiện qua ba bước chính: Input (Nhập dữ liệu), Query (Truy vấn), và Output (Xuất dữ liệu)



Hình 1: Sơ đồ xử lý sự kiện của Azure Stream Analytics

- Dữ liệu đầu vào được nhập từ: Azure event hub, Azure IoT Hub, hay Azure Blob Storage.
- Xử lý dữ liệu bằng cách sử dụng các câu truy vấn để chọn, chiếu, hoặc tổng hợp dữ liệu.
- Kết quả dữ liệu đầu ra được ghi chặng hạn như là Azure Data Lake Gen 2, Azure SQL Database, Azure Blob Storage ,Azure Synapse Analytics, Azure Functions, Azure event hub, Microsoft Power BI,...

Các câu truy vấn luồng (Stream Analytics query) sẽ chạy liên tục, dữ liệu đầu vào mới sẽ được xử lý và lưu trữ kết quả đầu ra.

### 3.2.1. Đầu vào (Inputs)

Azure Stream Analytics có thể nhận dữ liệu đầu vào từ nhiều nguồn khác nhau như:

- Azure Event Hubs.
- Azure Blob Storage.
- Azure Data Lake Storage Gen2.

- Azure IoT Hub.

Ngoài ra thì cũng có thể định nghĩa các đầu vào tham chiếu(reference inputs) dùng để nhập dữ liệu tĩnh bổ sung dữ liệu theo thời gian thực.

### 3.2.2. Đầu ra (Outputs)

Là đích đến mà kết quả của quá trình truyền tải dữ liệu được gửi tới. Azure Stream Analytics hỗ trợ một loạt các kết quả đầu ra:

- Duy trì kết quả của quá trình Stream để có thể thực hiện phân tích thêm, bằng cách lưu chúng vào trong Data Lake, hoặc kho dữ liệu.
- Hiển thị trực quan hóa dữ liệu luồng dữ liệu theo thời gian thực, bằng cách thêm dữ liệu vào tập dữ liệu trong Power BI.
- Tạo ra các bộ lọc hay tóm tắt để có thể xử lý tiếp theo. Có thể viết kết quả vào trong Event Hub.

### 3.2.3. Truy vấn (Queries)

Logic xử lý dòng dữ liệu được đóng gói trong một truy vấn. Các truy vấn được xác định bằng các câu lệnh SQL SELECT các trường dữ liệu từ (FROM) một hoặc nhiều nguồn dữ liệu, lọc hoặc tổng hợp dữ liệu và ghi kết quả vào (INTO) một đầu ra.

- Xử lý dữ liệu:

- Stream Analytics Query Language (SAQL) là một ngôn ngữ truy vấn được sử dụng trong Azure Stream Analytics để phân tích dữ liệu theo thời gian thực. SAQL là một phiên bản thu gọn của ngôn ngữ truy vấn SQL (Structured Query Language), được sử dụng phổ biến trong các cơ sở dữ liệu.
- Cú pháp của SAQL rất giống với T-SQL (Transact-SQL), là một phiên bản mở rộng của SQL được sử dụng trong các cơ sở dữ liệu Microsoft SQL Server. Các nhà phát triển quen thuộc với T-SQL sẽ dễ dàng làm quen và sử dụng SAQL.

### **3.3. Có hai cách sử dụng Azure Stream Analytics:**

- Sử dụng Azure Stream Analytics job trong Azure subscription, cài đặt đầu vào và đầu ra, định nghĩa các câu truy vấn cần thiết trong quá trình thực hiện. Các câu truy vấn sử dụng cấu trúc truy vấn SQL, có thể kết hợp thêm với tham chiếu tĩnh từ nhiều nguồn dữ liệu để cung cấp các giá trị tra cứu.
- Nếu dữ liệu phức tạp hoặc chuyên sâu, thì có thể sử dụng Stream Analytics cluster.

### **3.4. Một số tính năng của Azure Stream Analytics:**

- Exactly Once Event Processing: Azure Stream Analytics đảm bảo rằng mỗi sự kiện chỉ được xử lý đúng một lần duy nhất, tránh việc dữ liệu bị mất hoặc bị xử lý nhiều lần. Điều này đảm bảo tính nhất quán và tin cậy của dữ liệu.
- At-Least-Once Event Delivery: trong quá trình gửi dữ liệu từ nguồn đến đích, mỗi sự kiện sẽ được gửi đi ít nhất một lần là cần thiết. Azure Stream Analytics đảm bảo khi có sự cố xảy ra trong quá trình gửi dữ liệu, sự kiện sẽ được gửi lại để đảm bảo không có sự mất mát dữ liệu.
- Recovery Capabilities: Azure Stream Analytics có khả năng tự động khôi phục và tiếp tục quá trình xử lý dữ liệu mà không mất dữ liệu khi có sự cố xảy ra.
- Checkpointing: Azure Stream Analytics sử dụng tính năng checkpointing để lưu trữ trạng thái của công việc và tiếp tục từ checkpoint gần nhất.
- Azure Stream Analytics là một platform-as-a-service (PaaS) nên nó cung cấp một mô hình lập trình linh hoạt và có độ tin cậy cao, và có hiệu suất cao, bởi vì cho phép tính toán trong bộ nhớ. Sử dụng SQL-like-query-language

### **3.5. Ứng dụng dịch vụ Cloud**

- Azure Event Hubs.
  - Là một nền tảng phân phối sự kiện phân tán, có khả năng thu thập và xử lý hàng triệu sự kiện mỗi giây. Nó được thiết kế để xử lý luồng dữ liệu thời gian thực từ nhiều nguồn khác nhau, như các thiết bị IoT, ứng dụng web/mobile, logs, và nhiều nguồn khác.
- Azure Blob Storage.
  - Azure Blob Storage là một dịch vụ lưu trữ đối tượng (object storage) trong đám mây của Microsoft.
  - Nó cho phép lưu trữ và truy xuất các loại dữ liệu khác nhau như văn bản, hình ảnh, video và dữ liệu nhị phân.
  - Blob Storage có thể lưu trữ dữ liệu không cấu trúc với khối lượng lớn, phù hợp cho các ứng dụng lưu trữ dữ liệu IoT, phân tích dữ liệu và sao lưu.
- Power BI
  - Power BI là một bộ công cụ phân tích dữ liệu và BI (Business Intelligence) của Microsoft.
  - Nó cho phép kết nối, xử lý, phân tích và trực quan hóa dữ liệu từ nhiều nguồn khác nhau, bao gồm cả dịch vụ Azure.
  - Với Power BI, người dùng có thể tạo các báo cáo, bảng điều khiển (dashboard) và visualizations để hiểu rõ hơn về dữ liệu của mình.
  - Power BI tích hợp chặt chẽ với các dịch vụ Azure như Storage, SQL Database, Data Lake, giúp người dùng có thể kết nối và phân tích dữ liệu một cách dễ dàng.

### **3.6. Định dạng lưu trữ:**

JSON (JavaScript Object Notation) là 1 định dạng hoán vị dữ liệu nhanh. Chúng dễ dàng cho chúng ta đọc và viết. Dễ dàng cho thiết bị phân tích và phát sinh. Là tập hợp của các cặp tên và giá trị

## **4. MÔ HÌNH DỮ LIỆU**

### **4.1. Quản lý hiệu suất**

Giải pháp tối ưu hiệu suất:

- Streaming Units (SU):

- Streaming Units xác định khả năng xử lý dữ liệu của một job Azure Stream Analytics. Tăng số lượng SU cho phép xử lý nhiều dữ liệu hơn cùng một lúc, điều này trực tiếp giúp cải thiện tốc độ ghi và đọc dữ liệu. Việc chọn đúng số lượng SU là rất quan trọng để tối ưu hóa hiệu suất mà không gây lãng phí tài nguyên.
- Chi phí và tối ưu: Quản lý số lượng SU một cách hợp lý giúp tối ưu chi phí trong khi vẫn duy trì hiệu suất cần thiết cho các ứng dụng. Nếu bạn biết được khối lượng dữ liệu sẽ xử lý, bạn có thể điều chỉnh SU cho phù hợp, từ đó giảm chi phí không cần thiết.
- Stream Analytics sử dụng phân vùng trong các đầu vào và đầu ra để chia dữ liệu dòng. Sự chia này cho phép xử lý song song của các bước đầu vào, đầu ra và biến đổi. Mỗi bước biến đổi trong truy vấn công việc phải sử dụng PARTITION BY để tận dụng việc phân vùng. Sau đó, việc xử lý được phân phối trên các SUs được cung cấp.

- Event ordering:

- ASA hỗ trợ xử lý micro-batch trên dữ liệu phụ thuộc vào thời gian. Dữ liệu phụ thuộc vào thời gian chứa một hoặc nhiều dấu thời gian. Ở mức cơ bản, một dấu thời gian đơn lẻ chứa ngày và giờ. Các ứng

dụng thông thường của dấu thời gian bao gồm thời gian tạo thông báo và thời gian nộp. Mặc dù thời gian tạo và thời gian nộp nên gần như giống nhau, đôi khi thời gian nộp trễ so với thời gian tạo. Trong trường hợp này, cách tốt nhất là lặp lại việc gửi thông báo cho đến khi việc nộp thành công. Trong tình huống này, các dấu thời gian tạo và nộp có thể khác biệt đáng kể.

## 4.2. Select into

Cách đơn giản nhất để nhập dữ liệu đang phát vào Azure Synapse Analytics là ghi lại các tên cột cần thiết cho mỗi sự kiện bằng một truy vấn SELECT...INTO:

```
SELECT
    EventEnqueuedUtcTime AS ReadingTime,
    SensorID,
    ReadingValue
INTO
    [synapse-output]
FROM
    [streaming-input] TIMESTAMP BY EventEnqueuedUtcTime
```

## 4.3. Lọc dữ liệu (WHERE)

Trong một số trường hợp, bạn có thể muốn lọc dữ liệu để chỉ bao gồm các sự kiện cụ thể bằng cách thêm một mệnh đề WHERE. Ví dụ, truy vấn sau đây ghi dữ liệu chỉ cho các sự kiện có giá trị trường ReadingValue là âm.

```
SELECT
    EventEnqueuedUtcTime AS ReadingTime,
    SensorID,
    ReadingValue
INTO
    [synapse-output]
```

```
FROM  
[streaming-input] TIMESTAMP BY EventEnqueuedUtcTime  
WHERE ReadingValue < 0
```

#### 4.4. Tổng hợp qua các cửa sổ thời gian

Thông thường các truy vấn dữ liệu đang phát tổng hợp dữ liệu qua các khoảng thời gian, hoặc cửa sổ thời gian. Để thực hiện điều này, chúng ta có thể sử dụng một mệnh đề GROUP BY kèm theo một hàm Window định nghĩa loại cửa sổ mà mình muốn xác định (ví dụ, tumbling, hopping, hoặc sliding).

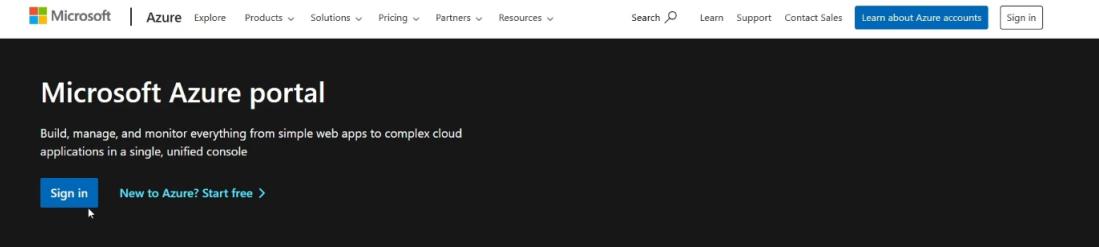
Ví dụ: nhóm các cảm biến đang phát vào cửa sổ tumbling (tuần tự, không chồng chéo) 1 phút, ghi lại thời gian bắt đầu và kết thúc của mỗi cửa sổ và giá trị tối đa cho mỗi cảm biến. Mệnh đề HAVING lọc kết quả để chỉ bao gồm các cửa sổ nơi ít nhất một sự kiện đã xảy ra.

```
SELECT  
DateAdd(second, -60, System.TimeStamp) AS StartTime,  
System.TimeStamp AS EndTime,  
SensorID,  
MAX(ReadingValue) AS MaxReading  
INTO  
[synapse-output]  
FROM  
[streaming-input] TIMESTAMP BY EventEnqueuedUtcTime  
GROUP BY SensorID, TumblingWindow(second, 60)  
HAVING COUNT(*) >= 1
```

### 5. DEMO

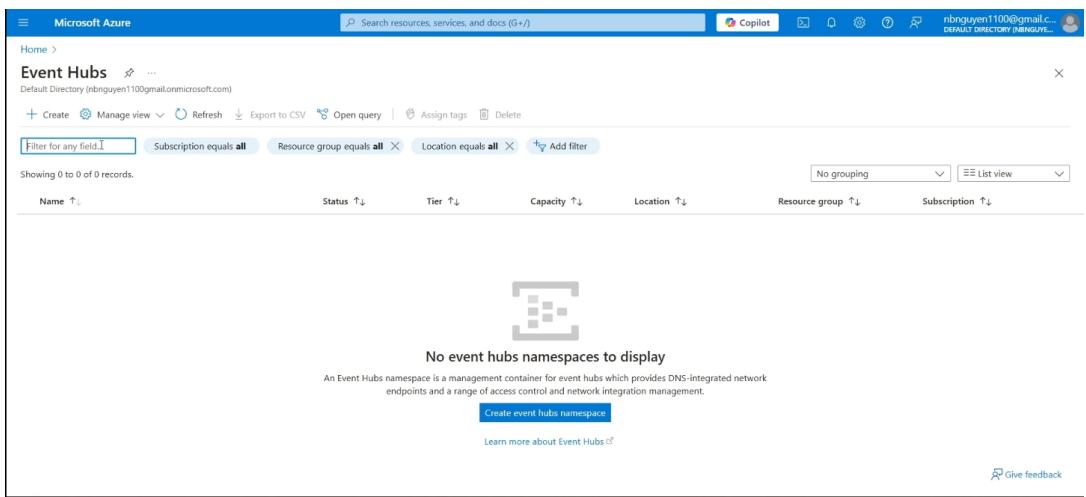
#### 5.1. Đăng nhập vào Azure

Bước 1: Đăng nhập vào Azure



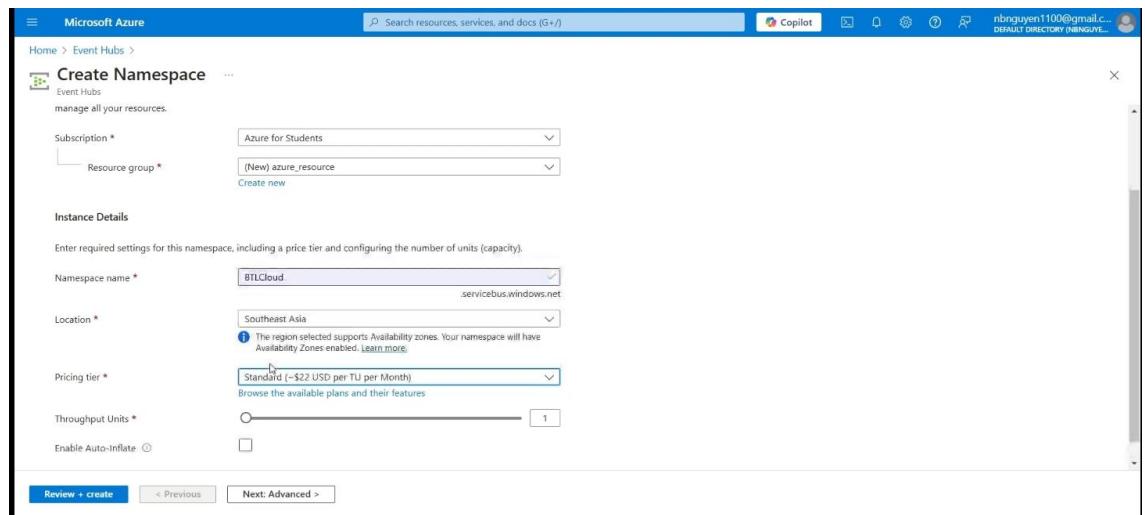
Hình 2: Trang chủ Microsoft Azure portal

Bước 2: Chọn Create để tạo mới 1 Event Hubs



Hình 3: Tạo mới một Event Hubs

Bước 3: Lựa chọn cấu hình cho Event Hubs, dưới đây là lựa chọn cơ bản



Hình 4: Cấu hình Event Hubs

Bước 4: Đây là giao diện khi tạo xong và mở Event Hubs

The screenshot shows the Microsoft Azure portal interface for the 'BTLCloud' Event Hubs Namespace. The top navigation bar includes 'Microsoft Azure', a search bar, and various account and service icons. The main content area is titled 'Event Hubs Namespace' with a sub-section 'Overview'. On the left, there's a sidebar with navigation links: 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Data Explorer (preview)', 'Events', 'Settings' (which is expanded to show 'Shared access policies', 'Scale', 'Geo-Recovery', 'Networking', 'Identity', 'Configuration', and 'Properties'), and a 'Search' bar at the top.

**Essential**

Resource group (move)	: azure_resource	Created	: Friday, November 1, 2024 at 16:17:49 GMT+7
Status	: Active	Updated	: Sunday, November 3, 2024 at 11:43:32 GMT+7
Location	: Southeast Asia	Zone Redundancy	: Enabled
Subscription (move)	: Azure for Students	Pricing tier	: Standard
Subscription ID	: b5e7e7ec-7e78-4817-a8cc-917954e949b7	Throughput Units	: 1 unit
Host name	: BTLCloud.servicebus.windows.net	Auto-inflate throughput ...	: Disabled
Tags (edit)	: Add tags	Local Authentication	: Enabled

**Metrics**

Namespace Contents: 0 EVENT HUBS | KAFKA SURFACE: ENABLED | ZONE REDUNDANCY: ENABLED

Show data for the last: 1 hour | 6 hours | 12 hours | 1 day | 7 days | 30 days

**Metrics Data:**

Category	Value
Requests	100
Messages	100
Throughput	1000

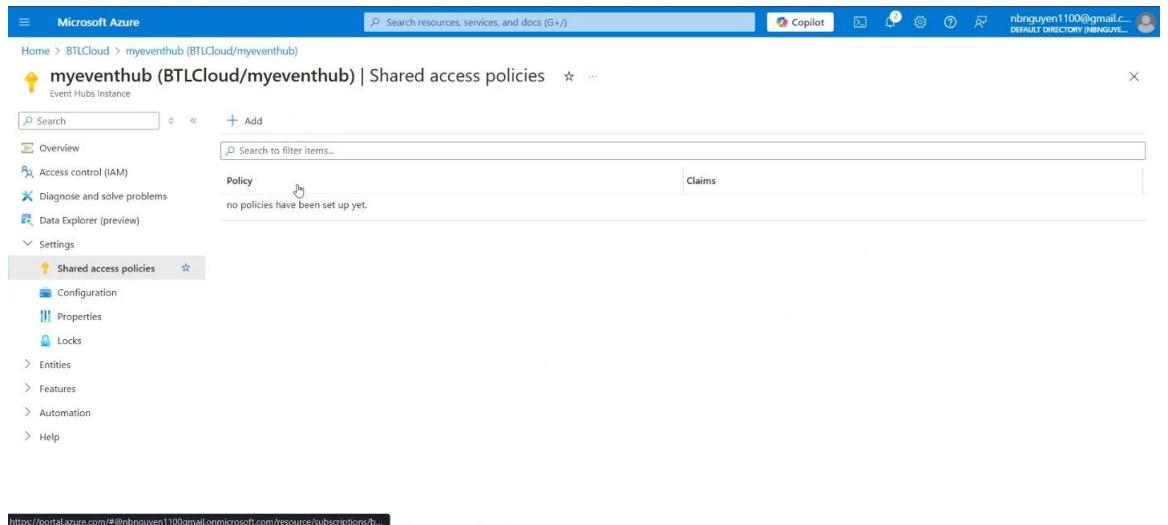
Hình 5: Giao diện khi tạo xong và mở Event Hubs

## Bước 5: Nhấn vào tạo Event Hubs

The screenshot shows the 'Create Event Hub' wizard in the Microsoft Azure portal. The current step is 'Basics'. The 'Event Hub Details' section requires entering settings for the event hub, including partition count and message retention. The 'Name' field is set to 'MyEventHub'. The 'Partition count' slider is set to 1. The 'Retention' section allows configuring cleanup policy (set to 'Delete') and retention time (set to 1 hour). At the bottom, there are buttons for 'Review + create', '< Previous', and 'Next: Capture >'.

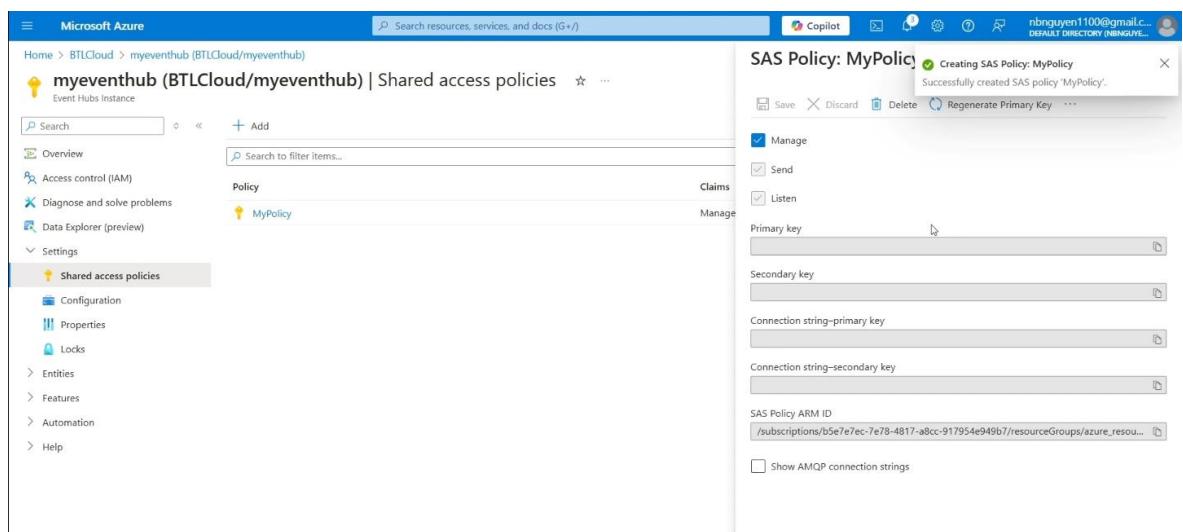
Hình 6: Nhấn tạo Event Hubs

## Bước 6: Chọn Shared access policies và tạo key



Hình 7: Chọn Shared access polices và tạo key

### Bước 7: Copy connection string



Hình 8: Copy connection string

### Bước 8: Vào telcodagen.exe.config dán chuỗi trên vào chỗ value của connection string và lưu ý value của EventHubName là giá trị của EntityPath

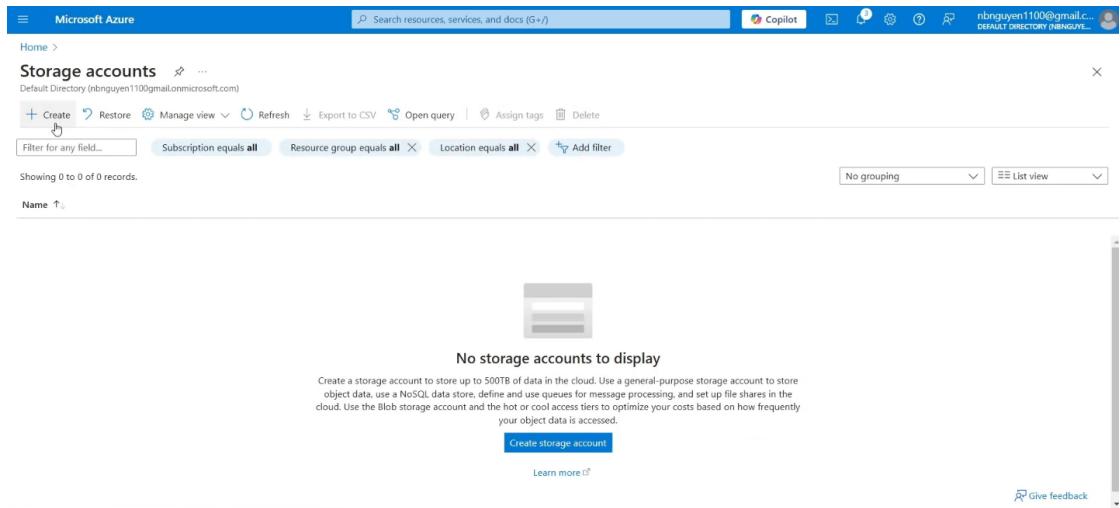
```

<?xml version="1.0" encoding="utf-8"?>
<configuration>
  <appSettings>
    <!-- Service Bus specific app settings for messaging connections -->
    <add key="EventHubName" value="myeventhub"/>
    <add key="Microsoft.ServiceBus.ConnectionString" value="Endpoint=sb://bt1cloud.servicebus.windows.net" />
  </appSettings>
</configuration>

```

Hình 9: Dán connection string vào value của file telcodagen.exe.config

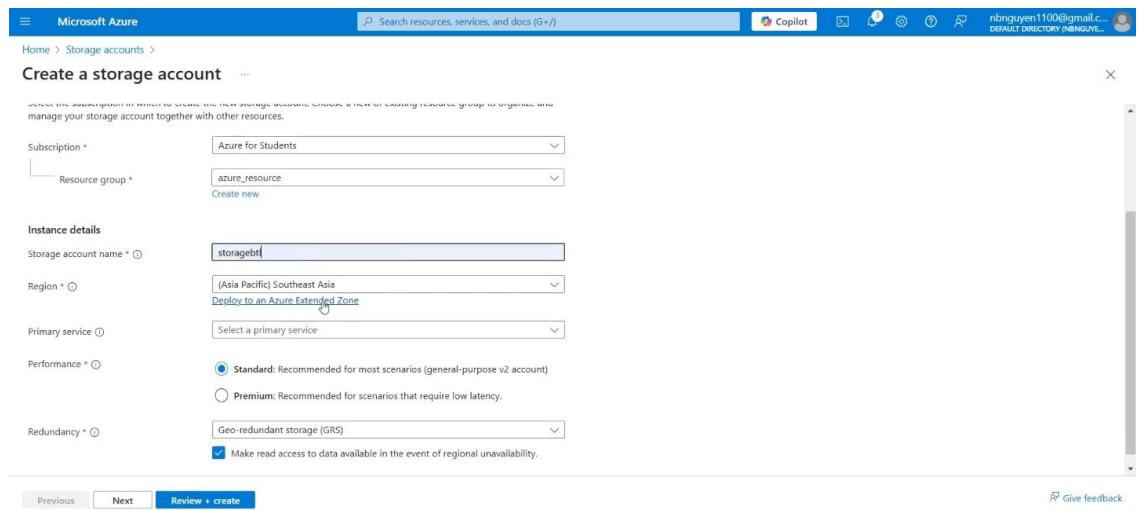
## Bước 9: Tạo Storage Account



The screenshot shows the Microsoft Azure Storage accounts page. At the top, there is a search bar and various navigation icons. Below the header, the page title is "Storage accounts". A sub-header indicates the default directory is "nbnguyen1100@gmail.com". There are buttons for "+ Create", "Restore", "Manage view", "Refresh", "Export to CSV", "Open query", "Assign tags", and "Delete". Below these are filters for "Subscription equals all", "Resource group equals all", and "Location equals all". A message states "Showing 0 to 0 of 0 records." and "Name ↑". In the center, there is a large gray icon representing storage. Below the icon, the text "No storage accounts to display" is shown. A descriptive paragraph explains what a storage account is and how to create one. It includes a "Create storage account" button and a "Learn more" link. At the bottom right, there is a "Give feedback" link.

Hình 10: Tạo Storage Account

## Bước 10: Lựa chọn cấu hình

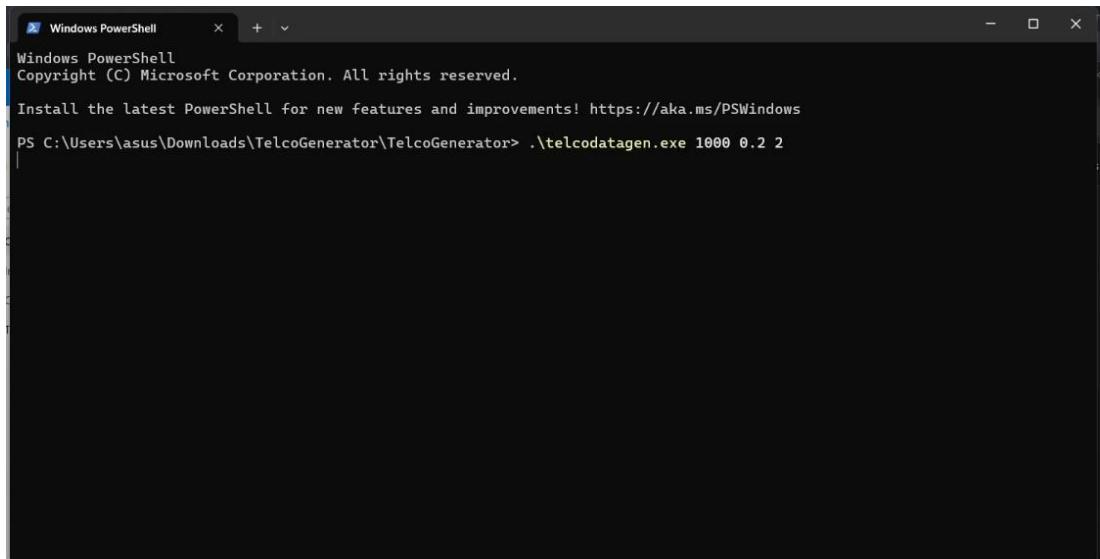


The screenshot shows the "Create a storage account" configuration page. At the top, it says "Create a storage account". Below that, there is a note about managing storage accounts with other resources. The "Subscription" dropdown is set to "Azure for Students" and the "Resource group" dropdown is set to "azure\_resource". Under "Instance details", the "Storage account name" is "storageb1", "Region" is "(Asia Pacific) Southeast Asia", and "Primary service" is "Select a primary service". The "Performance" section has "Standard" selected. Under "Redundancy", "Geo-redundant storage (GRS)" is chosen and a checkbox for "Make read access to data available in the event of regional unavailability" is checked. At the bottom, there are "Previous" and "Next" buttons, and a "Review + create" button.

Hình 11: Chọn cấu hình Storage Account

## Bước 11: Mở thư mục TelcoGenerator và chạy lệnh .\telcodatagen.exe 1000 0.2

2



```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

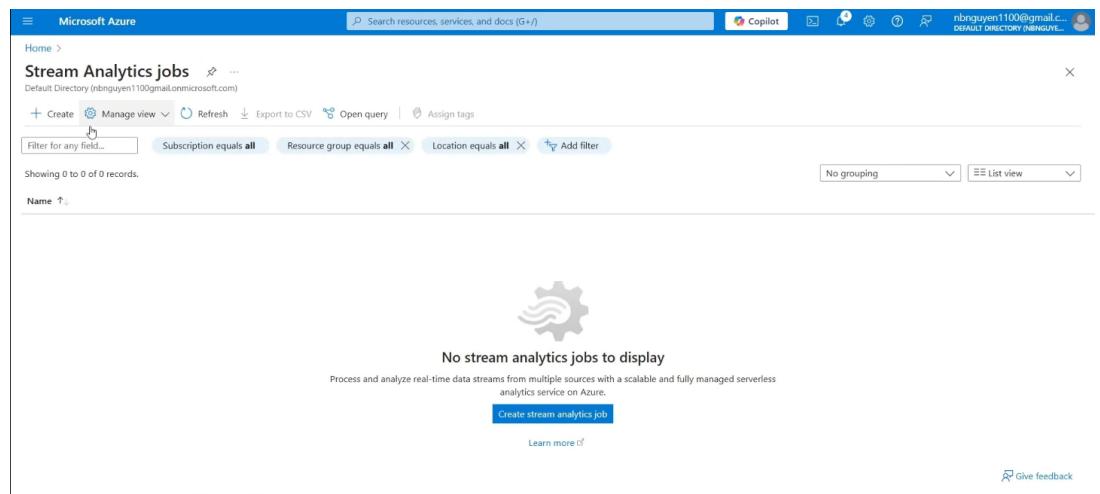
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\asus\Downloads\TelcoGenerator> .\telcodatagen.exe 1000 0.2 2
```

Hình 12: Chạy lên .\telcodatagen.exe 1000 0.2 2

## 5.2. Tạo Stream Analytics Job

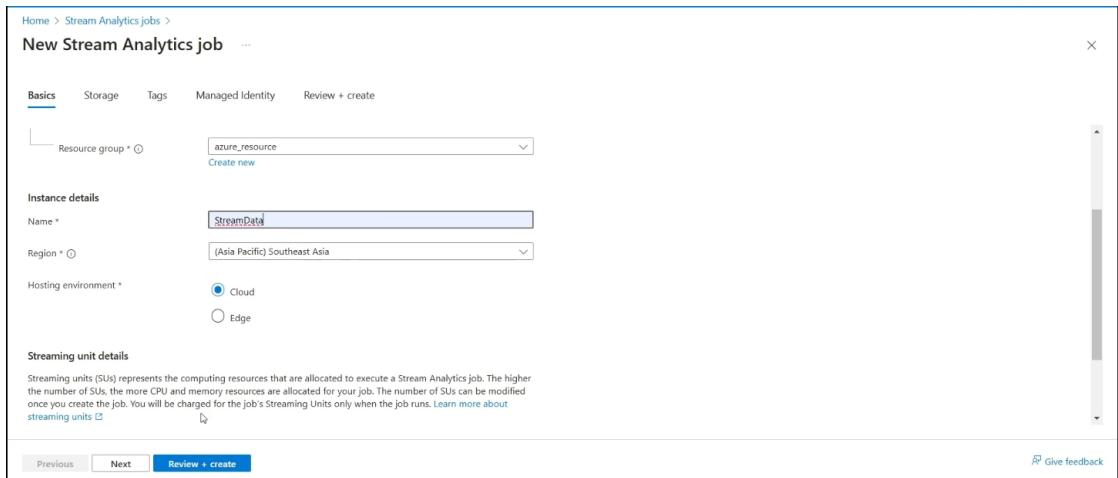
### Buộc 1: Tạo Stream Analytics Job



The screenshot shows the Microsoft Azure Stream Analytics jobs page. At the top, there's a search bar and a Copilot button. Below the search bar, there are filters for Subscription, Resource group, and Location, all set to 'all'. A 'Create' button is visible. The main area displays a message: 'No stream analytics jobs to display' with a sub-message: 'Process and analyze real-time data streams from multiple sources with a scalable and fully managed serverless analytics service on Azure.' A 'Create stream analytics job' button is prominently displayed.

Hình 13: Tạo Stream Analytics jobs

### Buộc 2: Cấu hình cho Stream Analytics job



Hình 14: Cấu hình Stream Analytics Job

## 5.3. Cấu hình Stream Analytics Job để truy vấn dữ liệu

Bước 1: Tạo Input cho Stream Analytics Job

Hình 15: Tạo input cho Stream Analytics Job

Bước 2: Tạo output thứ nhất là Blob

Hình 16: Tạo output thứ nhất Blob

### Bước 3: Tạo output thứ hai là Power BI

Hình 17: Tạo output thứ hai Power BI

### Bước 4: Tạo câu truy vấn và nhấn save

```

SELECT
    WindowEnd,
    TotalCalls,
    FraudulentCalls,
    SwitchNum
INTO "myBOutput"
FROM (
    SELECT
        System. Timestamp AS WindowEnd,
        cs1.TotalCalls,
        SUM(CASE WHEN CS1.SwitchNum != CS2.SwitchNum THEN 1 ELSE 0 END) AS FraudulentCalls
    FROM "CallStream" CS1 TIMESTAMP BY CallRecTime
    JOIN "CallStream" CS2 ON CS1.WindowEnd = CS2.WindowEnd AND CS1.SwitchNum = CS2.SwitchNum
)

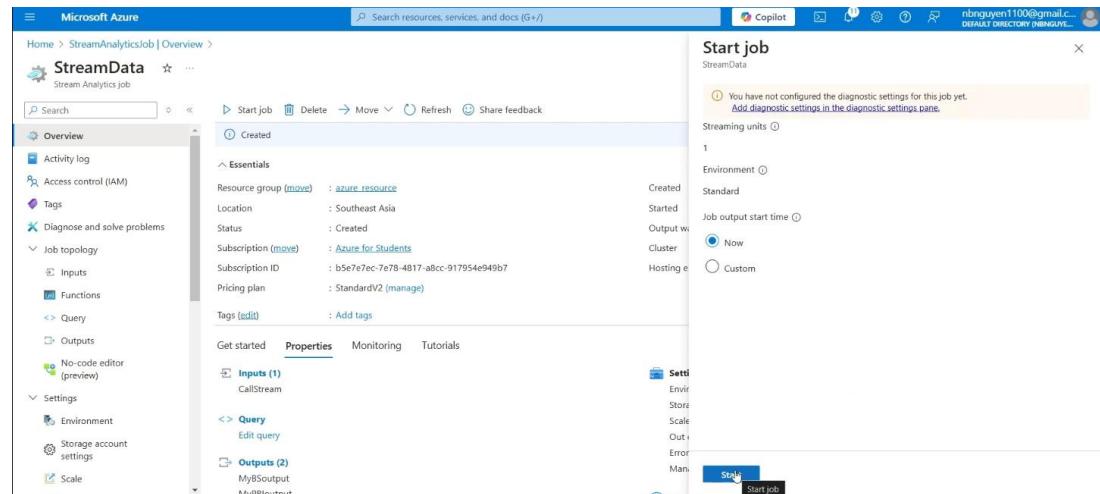
```

WindowEnd	FraudulentCalls	SwitchNum
"2024-11-03T04:45:43.000000Z"	1	"Germany"
"2024-11-03T04:45:44.000000Z"	1	"China"

Hình 18: Tạo câu truy vấn và save

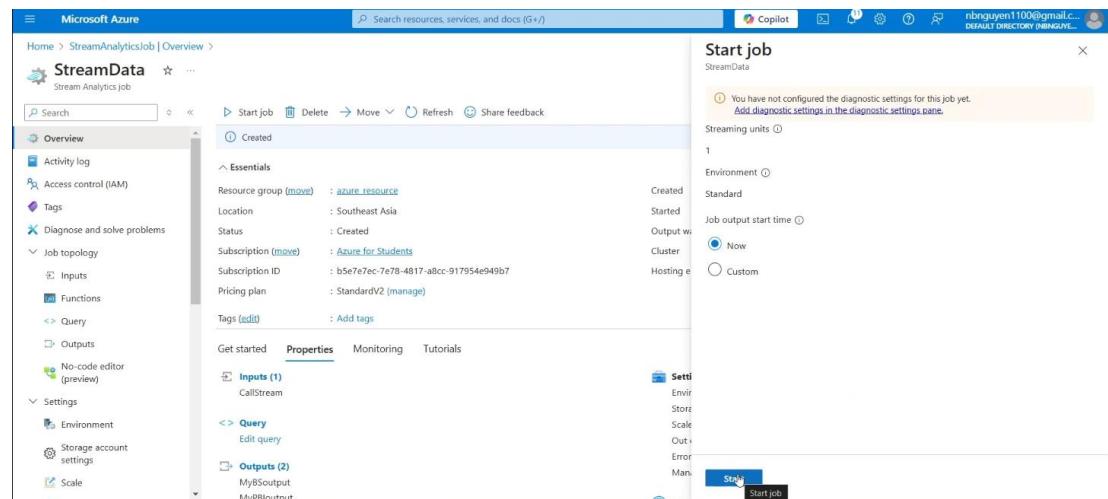
## 5.4. Chạy Stream Analytics Job và tiến hành phân tích bằng Power BI

Bước 1: Chọn start job



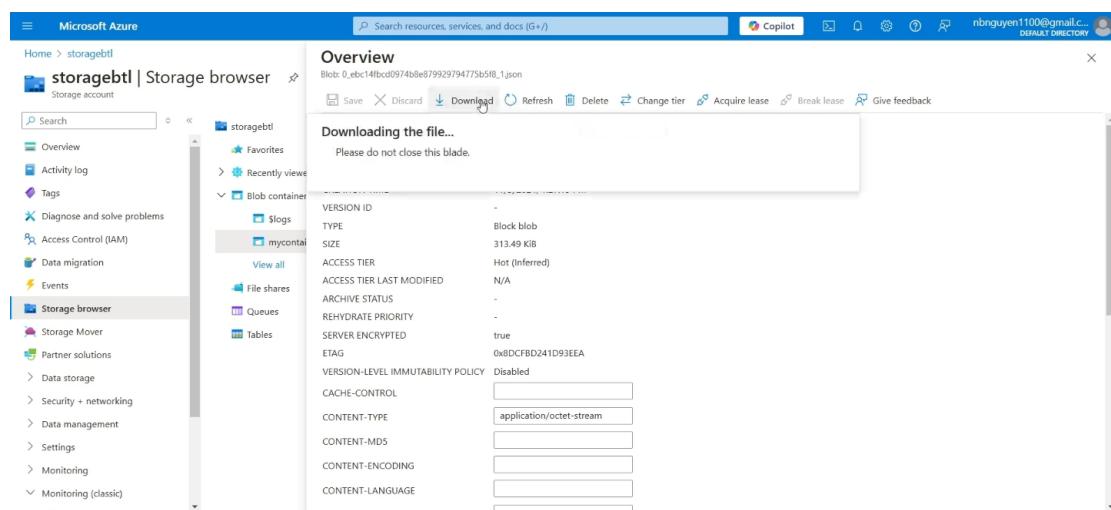
Hình 19: Chọn start job ở Stream Analytics Job

Bước 2: Giao diện sau khi start thành công



Hình 20: Giao diện sau khi start thành công

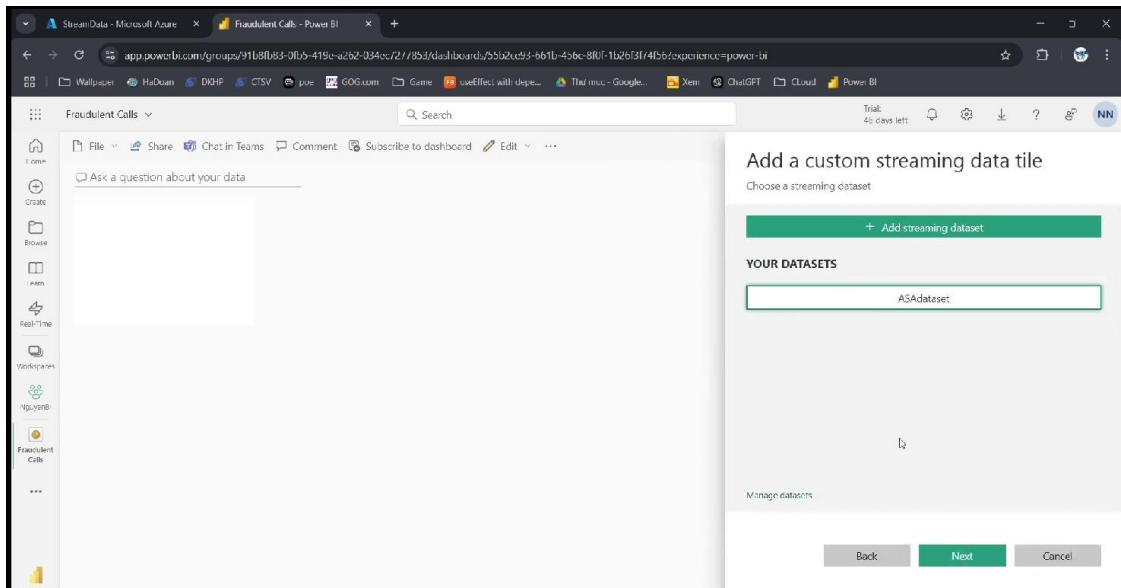
### Bước 3: Vào storage account để lấy data



Hình 21: Vào Storage Account để lấy data

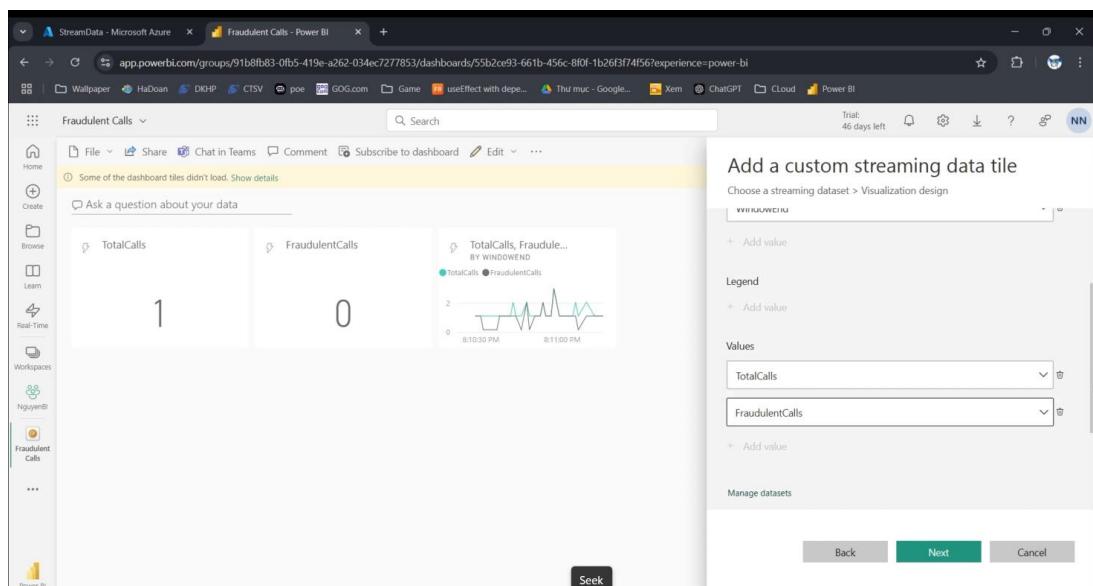
### Tạo Dashboard với dữ liệu được cập nhật theo thời gian thực

#### Bước 4: Import dữ liệu



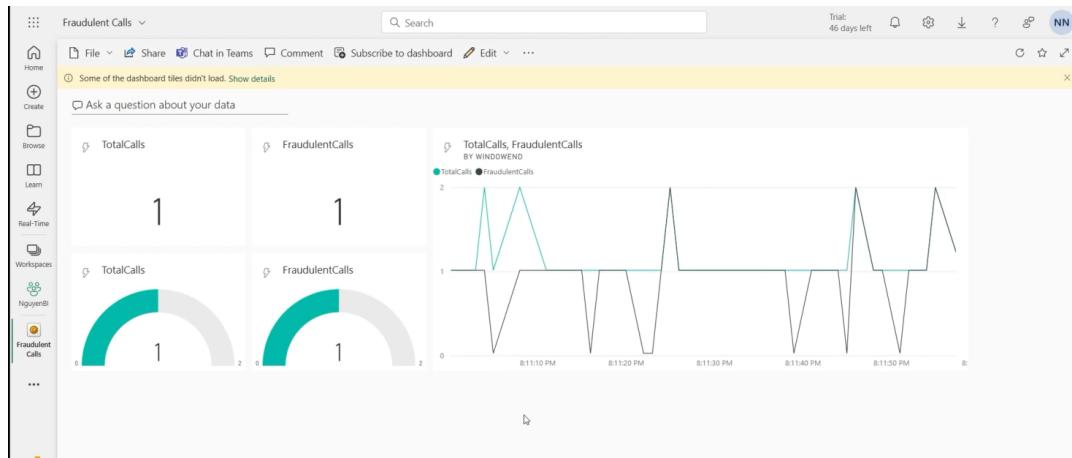
Hình 22: Import dữ liệu

### Bước 5: Tạo các report cho dữ liệu



Hình 23: Tạo các report

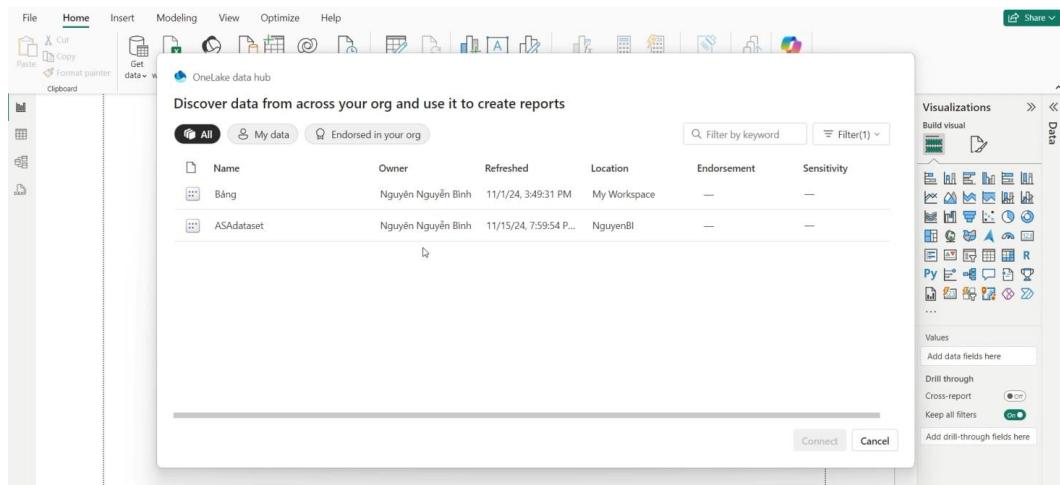
Kết quả:



Hình 24: Tạo các report

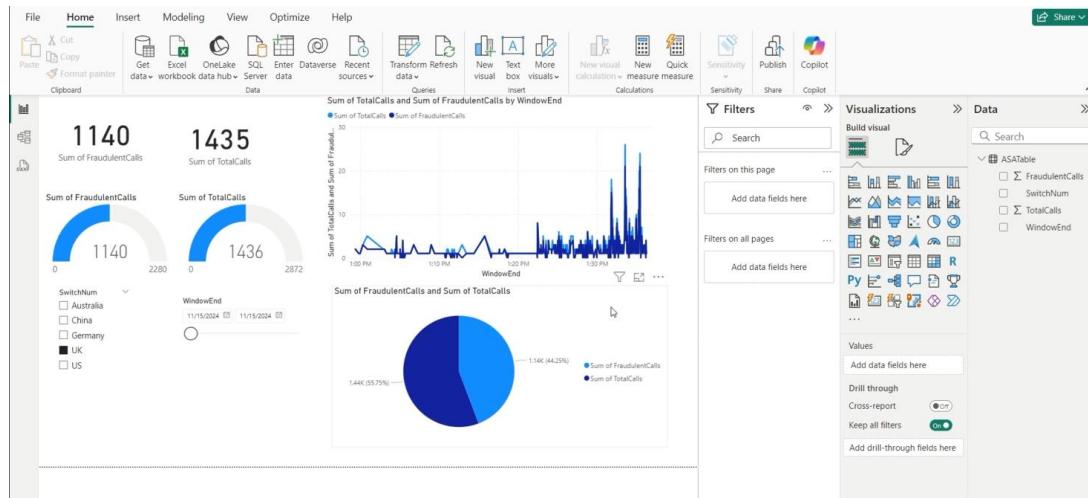
## Tạo Dashboard phân tích dữ liệu với filter và một số report

### Bước 6: Import dữ liệu



Hình 25: Import dữ liệu

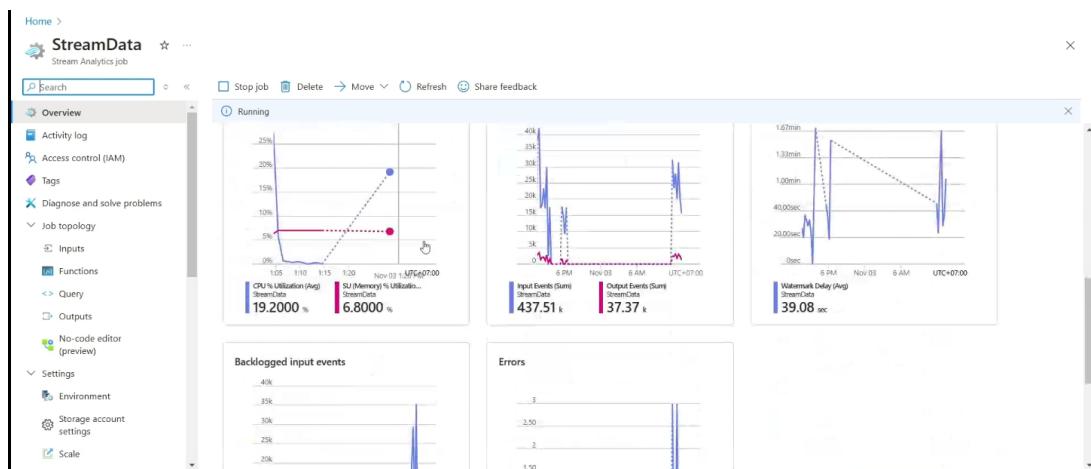
### Bước 7: Tạo các report và sau đây là kết quả



Hình 26: Dashboard thứ hai để phân tích dữ liệu Hình 26:

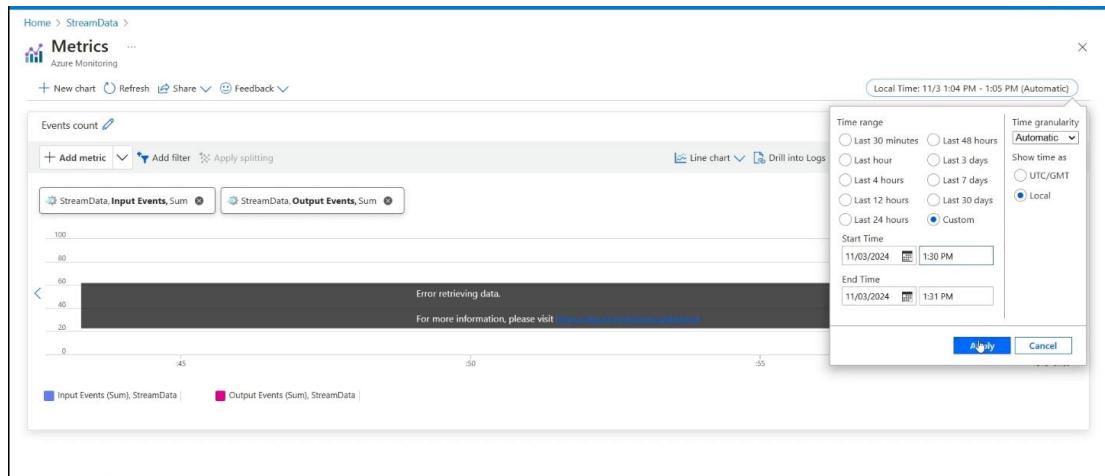
## 5.5. Đo tốc độ đọc ghi

Bước 1: Vào Stream Analytics Job và chọn biểu đồ Event Count



Hình 27: Vào Stream Analytics Job để chọn biểu đồ Event Count

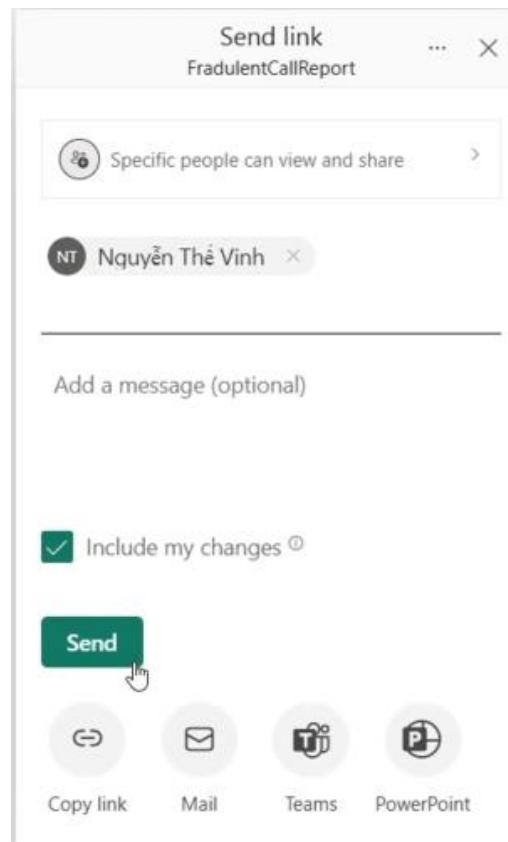
Bước 2: Chọn khoảng thời gian một phút và dưới đây là kết quả



Hình 28: Chọn thời gian khoảng một phút và kết quả

## 5.6. Share report cho các thành viên trong nhóm

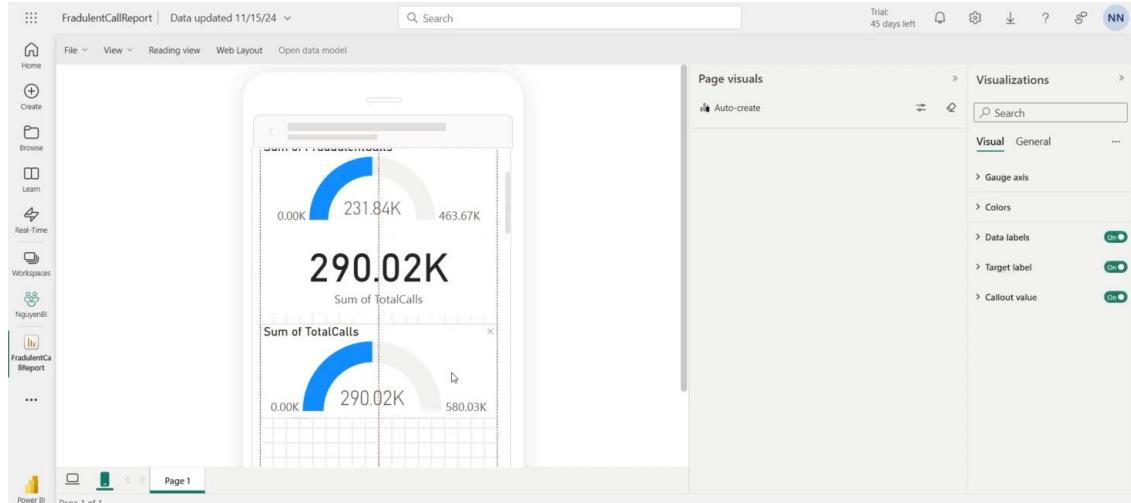
Chọn nút share và chọn người cần share



Hình 29: Share report

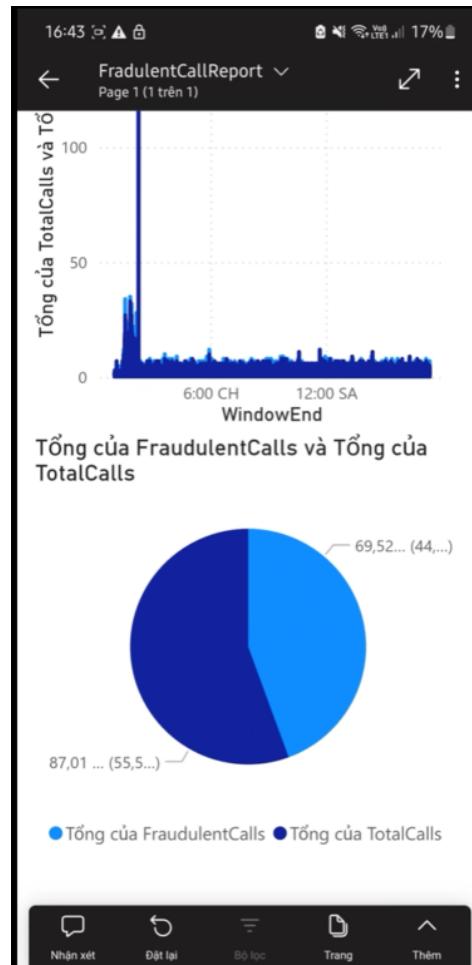
## 5.7. Chính sửa layout của Dashboard cho Mobile

Bước 1: Chọn nút layout cho mobile (có hình điện thoại ở góc trái màn hình)



Hình 30: Chọn nút layout cho mobile

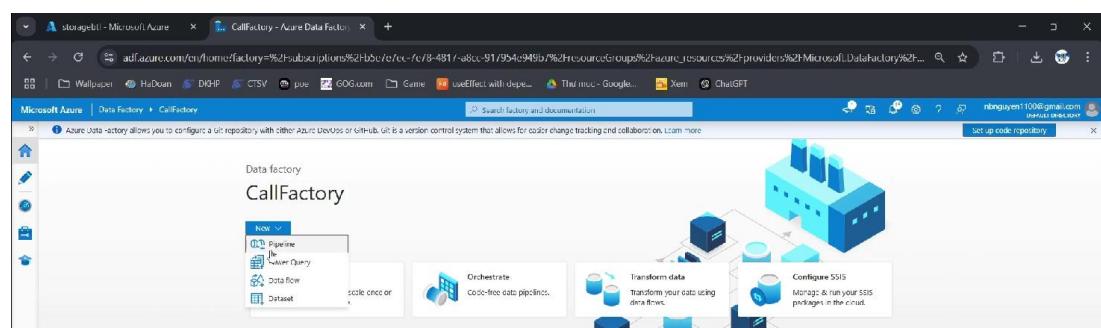
Kết quả:



Hình 31: Dashboard khi xem trên mobile

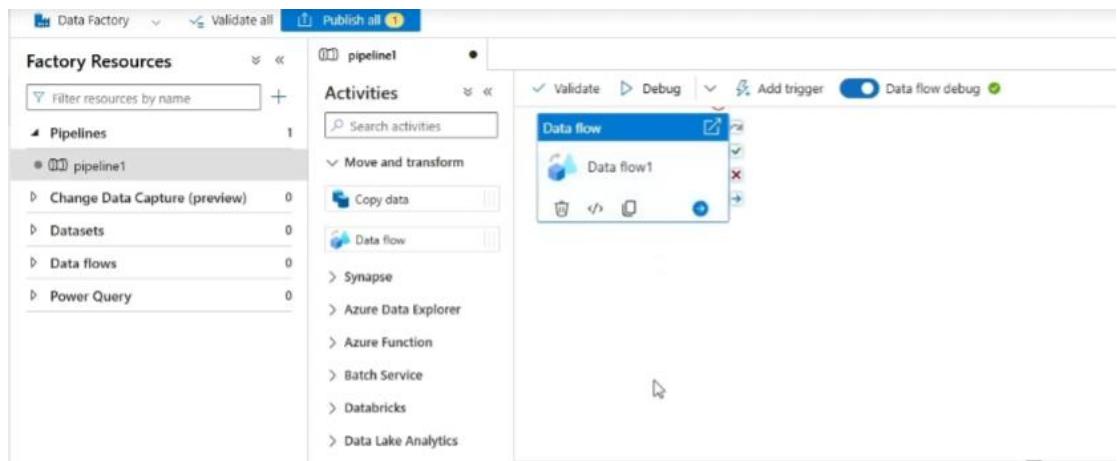
## 5.8. Tạo luồng xử lý dữ liệu tự động ETL

Bước 1: Tạo mới 1 pipeline



Hình 32: Tạo mới 1 pipeline

Bước 2: Chọn activity là data flow

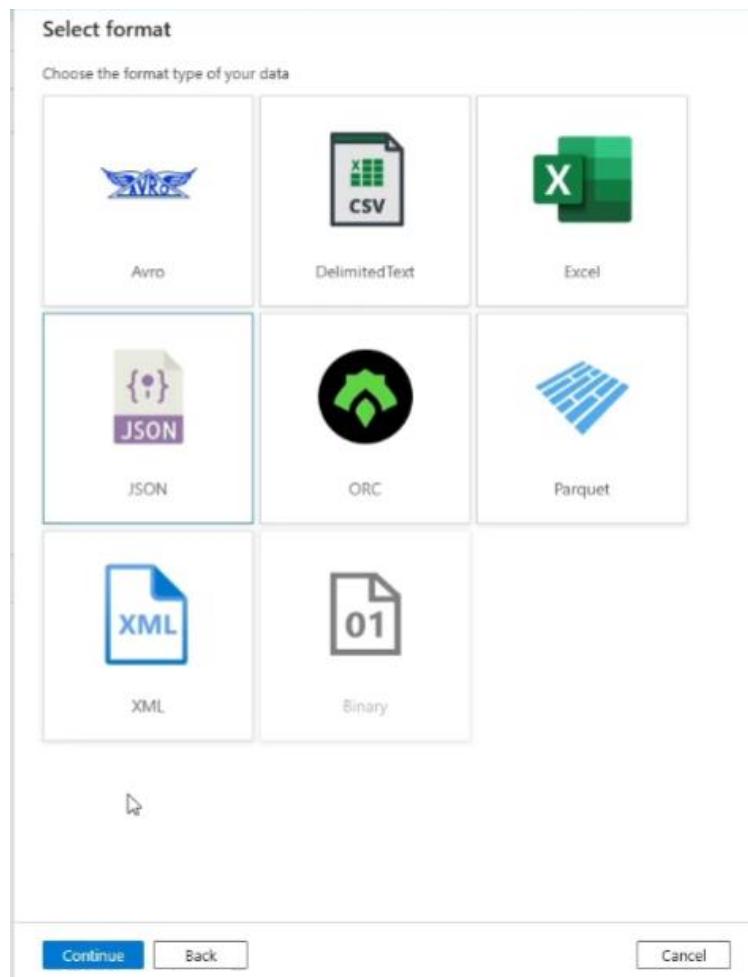


Hình 33: Chọn activity là data flow

### Bước 3: Chọn source là Blob Storage

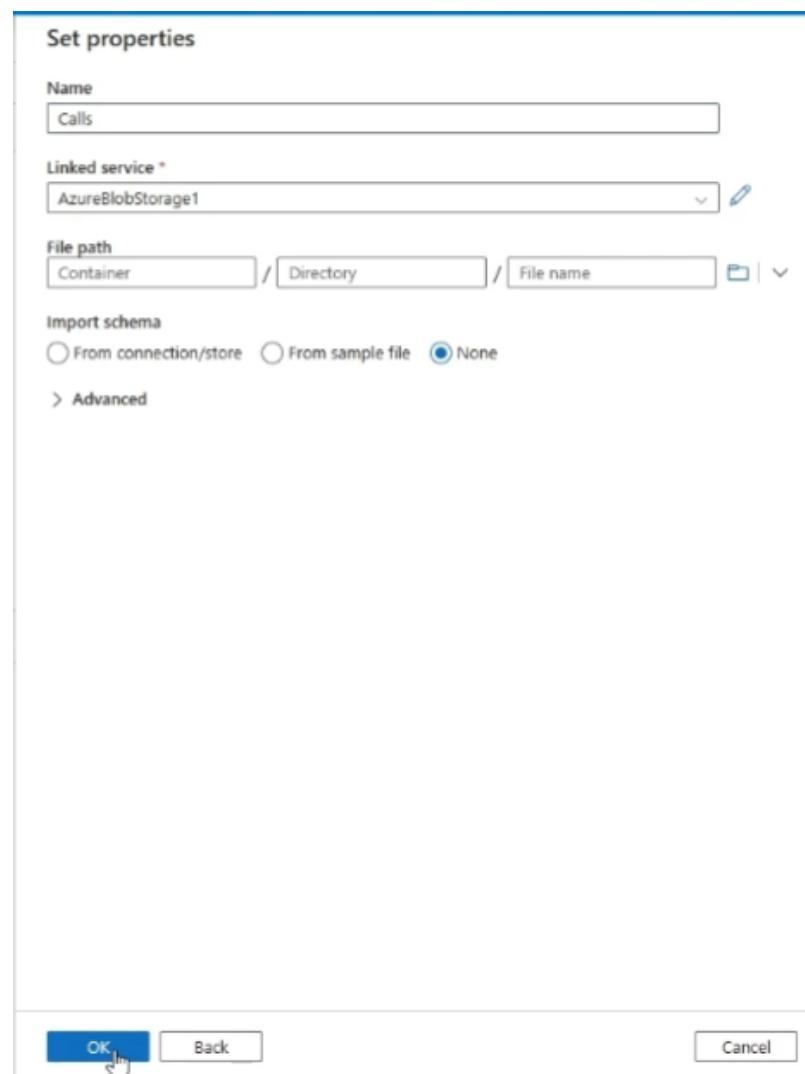
Hình 34: Chọn source là Blob Storage

### Bước 4: Format là json



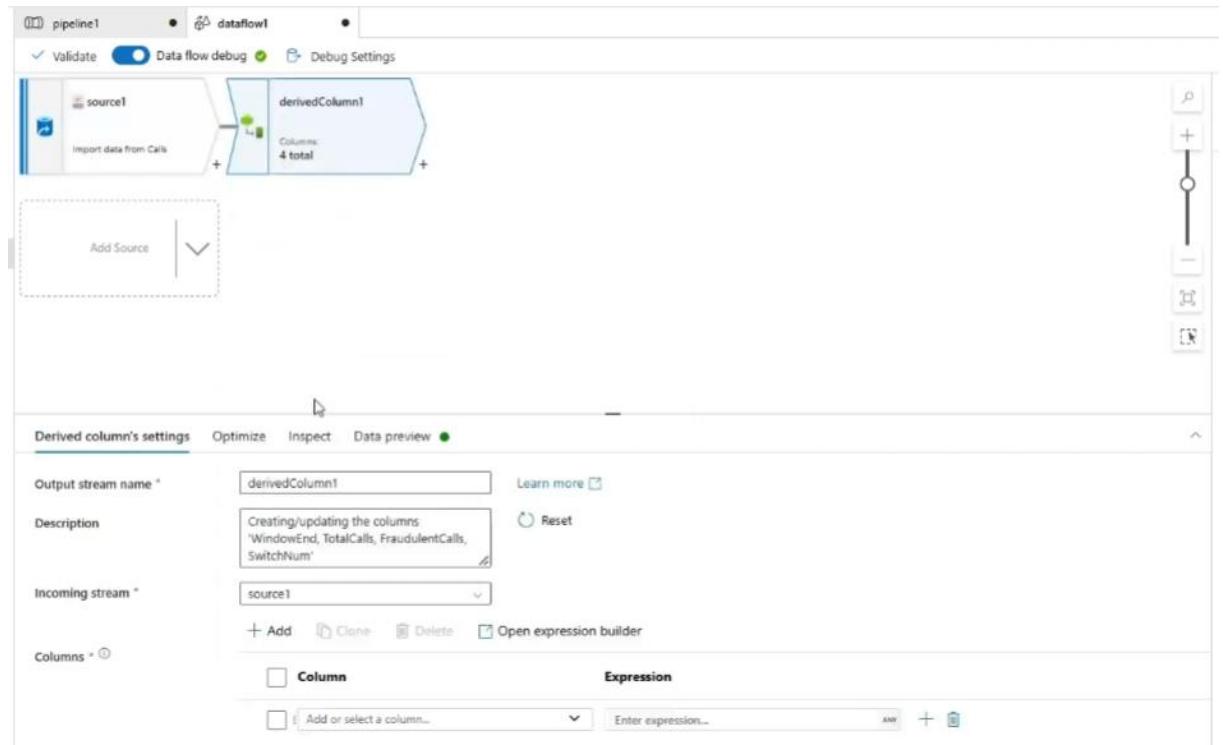
Hình 35: Format là json

Bước 5: Thiết lập đường dẫn cho source

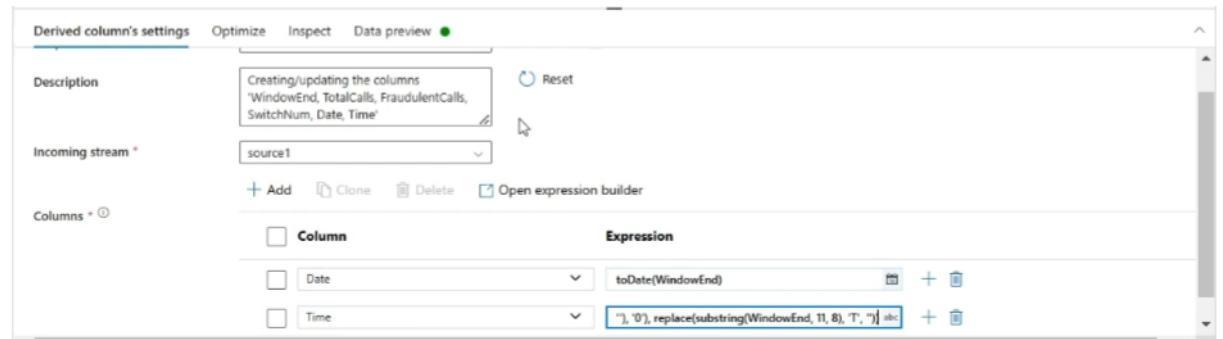


Hình 36: Thiết lập đường dẫn cho source

Bước 6: Sử dụng Derived Column để tách cột DateTime chứa nội dung theo kiểu Timestamp thành 2 cột là Date và Time

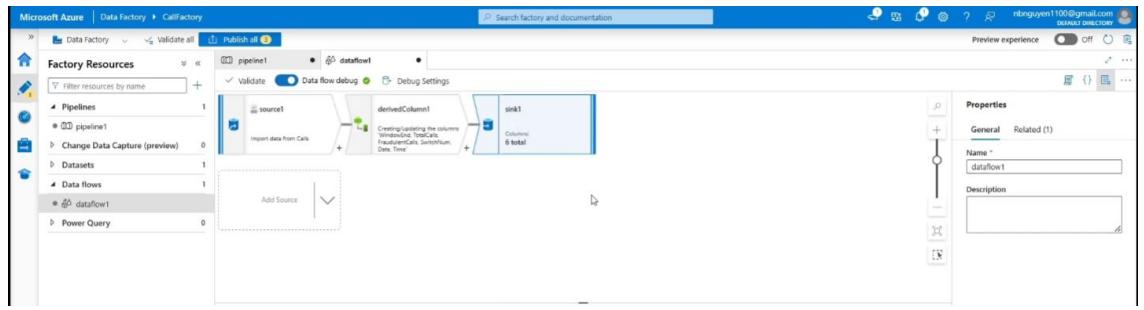


Hình 37: Sử dụng Derived Column



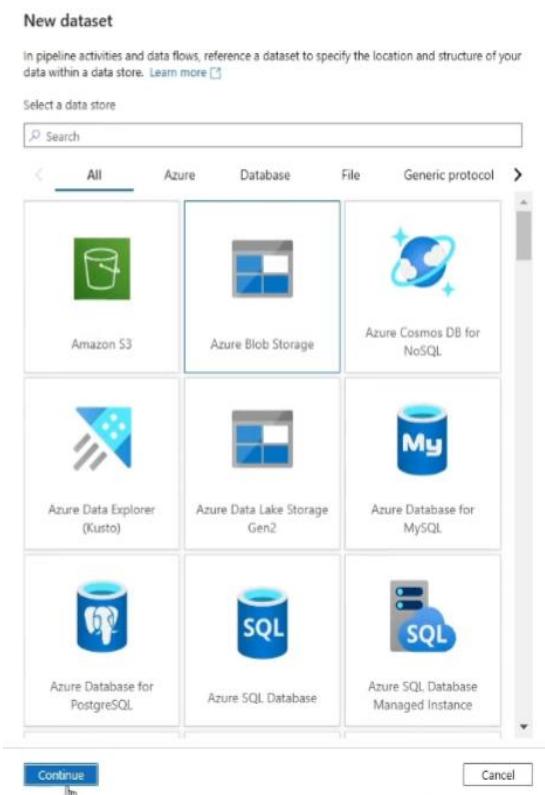
Hình 38: Tạo 2 cột mới là Date và Time

Bước 7: Thiết lập đích đến là Sink

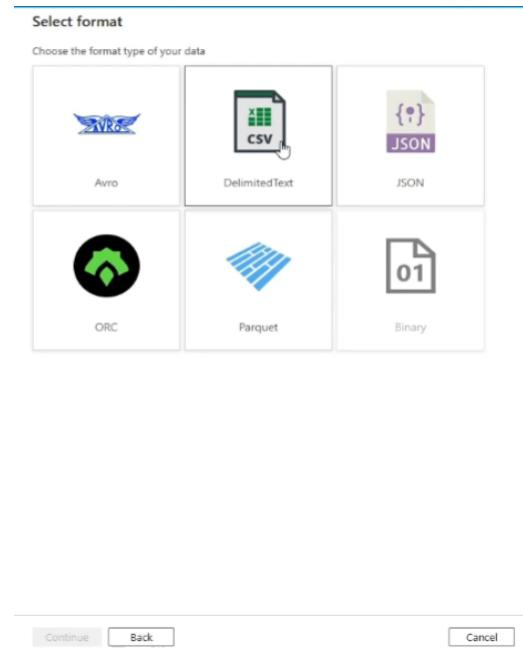


Hình 39: Thiết lập đích đến là Sink

Bước 8: Chọn dataset là Azure Blob Storage và format là CSV



Hình 40: Chọn dataset là Azure Blob Storage



Hình 41: Chọn format là CSV

#### Bước 9: Chọn output ra 1 file tên là CallsCSV.csv

Hình 42: Chọn output ra 1 file tên là CallsCSV.csv

#### Bước 10: Mapping các cột

Hình 43: Mapping các cột

## Bước 11: Tạo trigger

New trigger

Name \*  
trigger2

Description

Type \*  
Schedule

Start date \*  
11/17/2024, 11:42:20 AM

Time zone \*  
Bangkok, Hanoi, Jakarta (UTC+7)

Recurrence \*  
Every 1d Hour(s)

Specify an end date

Annotations

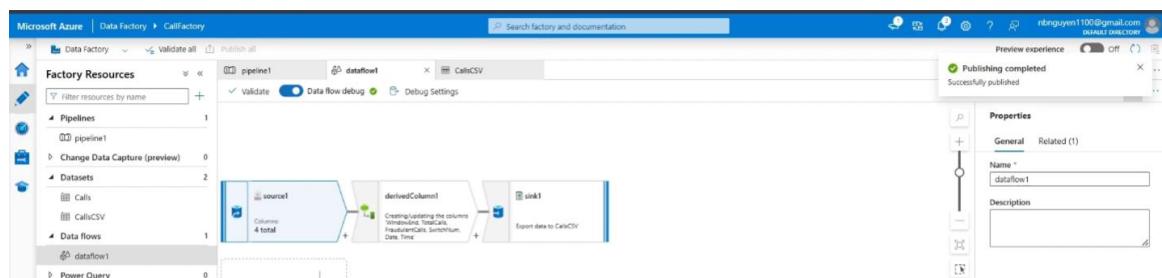
+ New

Start trigger  
 Start trigger on creation

OK Cancel

Hình 44: Tạo trigger

## Bước 12: Publish tất cả sau khi đã cấu hình xong



Hình 45: Publish tất cả:

Kết quả: file CSV sau khi xử lý sẽ có dạng như sau

A1	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	TotalCalls	Fraudulent	SwitchNum	Date	Time															
2	1	1	China	11/17/2024	10:12:03															
3	1	1	US	11/17/2024	10:12:03															
4	3	3	Germany	11/17/2024	10:12:03															
5	1	1	UK	11/17/2024	10:12:03															
6	3	1	US	11/17/2024	10:12:03															
7	4	4	Germany	11/17/2024	10:12:03															

Hình 46: File CallsCSV.csv

## 6. THAM KHẢO

- [1] “*Welcome to Azure Stream Analytics*”. Available at  
<https://learn.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction> [Accessed 29 October 2024]
- [2] “*Stream Analytics Query Language Reference*”. Available at  
<https://learn.microsoft.com/en-us/stream-analytics-query/stream-analytics-query-language-reference> [Accessed 29 October 2024]
- [3] “*Tutorial: Analyze fraudulent call data with Stream Analytics and visualize results in Power BI dashboard*”. Available at  
<https://learn.microsoft.com/en-us/azure/stream-analytics/stream-analytics-real-time-fraud-detection> [Accessed 30 October 2024]