**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO BÀI TẬP LỚN**

**ĐIỆN TOÁN ĐÁM MÂY**

**ĐỀ TÀI: XÂY DỰNG HỆ THỐNG TRỰC QUAN DỮ LIỆU CỔ PHIẾU DỰA TRÊN CÔNG NGHỆ XỬ LÝ LUỒNG VÀ PHÂN TÍCH THỜI GIAN THỰC TRÊN AZURE**

Lớp: IS402.O21.HTCL

Giảng viên hướng dẫn: ThS. Hà Lê Hoài Trung

Sinh viên thực hiện:

Nguyễn Xuân Tuấn Kiệt 20521502

Trần Duy Khánh 20521457

Trần Thạnh Phong 20521750

**TP. Hồ Chí Minh, tháng 05 năm 2024**

**LỜI CẢM ƠN**

Đầu tiên, nhóm chúng em xin gửi lời cảm ơn chân thành đến quý Thầy cô giảng viên Trường Đại học Công nghệ thông tin – Đại học Quốc gia TP.HCM và quý thầy cô khoa Hệ thống Thông tin đã giúp cho nhóm chúng em có những kiến thức cơ bản làm nền tảng để thực hiện đề tài này.

Đặc biệt, nhóm chúng em xin gửi lời cảm ơn và lòng biết ơn sâu sắc nhất tới thầy giáo – ThS. Hà Lê Hoài Trung, người đã hướng dẫn cho nhóm chúng em trong suốt thời gian làm đề tài. Thầy đã trực tiếp hướng dẫn tận tình và đóng góp nhiều ý kiến quý giá giúp nhóm chúng em hoàn thành tốt báo cáo môn học của mình. Một lần nữa em chân thành cảm ơn thầy và chúc thầy dồi dào sức khoẻ.

Trong thời gian một học kỳ thực hiện đề tài, nhóm chúng em đã vận dụng những kiến thức nền tảng đã tích lũy đồng thời kết hợp với việc học hỏi và nghiên cứu những kiến thức mới từ thầy cô, bạn bè cũng như nhiều nguồn tài liệu tham khảo. Từ đó, nhóm chúng em vận dụng tối đa những gì đã thu thập được để hoàn thành một báo cáo đồ án tốt nhất. Tuy nhiên, vì kiến thức chuyên môn còn hạn chế và nhóm còn thiếu nhiều kinh nghiệm thực tiễn nên nội dung của báo cáo không tránh khỏi những thiếu xót, nhóm chúng em rất mong nhận được sự góp ý, chỉ bảo thêm của quý thầy cô nhằm hoàn thiện những kiến thức của mình để nhóm chúng em có thể dùng làm hành trang thực hiện tiếp các đề tài khác trong tương lai cũng như là trong việc học tập và làm việc sau này.

Một lần nữa xin gửi đến thầy cô, bạn bè lời cảm ơn chân thành và tốt đẹp nhất!

Thành phố Hồ Chí Minh, tháng 05 năm 2024 Nhóm sinh viên thực hiện

NHẬN XÉT CỦA GIẢNG VIÊN

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

# **MỤC LỤC**

[**MỤC LỤC** 5](#_Toc167970538)

[**MỤC LỤC ẢNH** 6](#_Toc167970539)

[**MỤC LỤC BẢNG** 8](#_Toc167970540)

[**NỘI DUNG** 9](#_Toc167970541)

[**CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI** 9](#_Toc167970542)

[**1.1.** **Mô tả đề tài** 9](#_Toc167970543)

[**1.2.** **Mục tiêu đề tài** 10](#_Toc167970544)

[**CHƯƠNG 2. LÝ THUYẾT TỔNG QUAN** 11](#_Toc167970545)

[**2.1. Giới thiệu** 11](#_Toc167970546)

[**2.2. Kiến trúc hệ thống** 11](#_Toc167970547)

[**2.2.1. Azure Event Hubs** 11](#_Toc167970548)

[**2.2.2. Azure Stream Analytics** 12](#_Toc167970549)

[**2.2.3. Azure Blob Storage** 12](#_Toc167970550)

[**2.2.4. Azure SQL Database** 12](#_Toc167970551)

[**2.2.5. Azure Machine Learning** 12](#_Toc167970552)

[**2.2.6. Azure Data Factory** 13](#_Toc167970553)

[**2.2.7. Microsoft Power BI** 13](#_Toc167970554)

[**CHƯƠNG 3. Dữ liệu và cấu hình** 14](#_Toc167970555)

[**3.1. Mô tả dữ liệu** 14](#_Toc167970556)

[**3.2. Cài đặt và cấu hình trên Azure** 14](#_Toc167970557)

[**CHƯƠNG 4. THỰC THI VÀ KẾT QUẢ** 41](#_Toc167970558)

[**4.1. Thực thi trên Azure** 41](#_Toc167970559)

[**4.2. Kết quả và đánh giá.** 44](#_Toc167970560)

[**CHƯƠNG 5. TỔNG KẾT** 47](#_Toc167970561)

[**5.1. Kết luận** 47](#_Toc167970562)

[**5.2. Hướng phát triển** 47](#_Toc167970563)

[**BẢNG PHÂN CÔNG CÔNG VIỆC** 48](#_Toc167970564)

[**TÀI LIỆU THAM KHẢO** 49](#_Toc167970565)

# **MỤC LỤC ẢNH**

[Hình 1: Sơ đồ hoạt động của hệ thống 9](#_Toc167970567)

[Hình 2: Biểu tượng Azure Event Hubs 11](#_Toc167970568)

[Hình 3: Biểu tượng Azure Stream Analytics 12](#_Toc167970569)

[Hình 4: Biểu tượng Azure Blob Storage 12](#_Toc167970570)

[Hình 5: Biểu tượng Azure SQL Database 12](#_Toc167970571)

[Hình 6: Biểu tượng Azure Machine Learning 13](#_Toc167970572)

[Hình 7: Biểu tượng Azure Data Factory 13](#_Toc167970573)

[Hình 8: Biểu tượng Microsoft Power BI 13](#_Toc167970574)

[Hình 9: Minh họa tạo Event Hubs 15](#_Toc167970575)

[Hình 10: Minh họa cấu hình Event Hubs. 15](#_Toc167970576)

[Hình 11: Minh họa tạo sự kiện trong Event Hubs 16](#_Toc167970577)

[Hình 12: Minh họa tạo Shared Access Policies. 16](#_Toc167970578)

[Hình 13: Minh họa cấu hình Stream Analytics job mới 17](#_Toc167970579)

[Hình 14: Minh họa cấu hình input 17](#_Toc167970580)

[Hình 15: Minh họa cấu hình storage account 18](#_Toc167970581)

[Hình 16: Minh họa tạo output bằng Blob storage 18](#_Toc167970582)

[Hình 17: Minh họa tạo output bằng Power BI 19](#_Toc167970583)

[Hình 18: Minh họa cấu hình output Power BI 19](#_Toc167970584)

[Hình 19: Màn hình file Python gọi dữ liệu từ API đến EventHubs 20](#_Toc167970585)

[Hình 20: Minh họa cấu hình Data Factory 21](#_Toc167970586)

[Hình 21: Minh họa trang trang khởi động Data Factory 21](#_Toc167970587)

[Hình 22: Minh họa tạo Dataset trên Data Factory 22](#_Toc167970588)

[Hình 23: Minh họa chọn nguồn dữ liệu 22](#_Toc167970589)

[Hình 24: Minh họa chọn kiểu dữ liệu 23](#_Toc167970590)

[Hình 25: Minh họa tạo linked service 23](#_Toc167970591)

[Hình 26: Minh họa cấu hình linked service (1) 24](#_Toc167970592)

[Hình 27: Minh họa cấu hình linked service (2) 24](#_Toc167970593)

[Hình 28: Minh họa kết quả preview data 25](#_Toc167970594)

[Hình 29: Minh họa cấu hình SQL Database 25](#_Toc167970595)

[Hình 30: Minh họa liên kết dataset đến SQL Database 26](#_Toc167970596)

[Hình 31: Minh họa tạo Linked service cho SQL Database. 26](#_Toc167970597)

[Hình 32: Minh họa tạo linked service mới 27](#_Toc167970598)

[Hình 33: Minh họa tạo pipeline 27](#_Toc167970599)

[Hình 34: Minh họa chọn activities 27](#_Toc167970600)

[Hình 35: Minh họa copy data Activities 28](#_Toc167970601)

[Hình 36: Minh họa cấu hình Source 28](#_Toc167970602)

[Hình 37: Minh họa cấu hình Sink 29](#_Toc167970603)

[Hình 38: Minh họa cấu hình Query editor 29](#_Toc167970604)

[Hình 39: Minh họa Preview Data 30](#_Toc167970605)

[Hình 40: Minh họa cấu hình Machine Learning 30](#_Toc167970606)

[Hình 41: Màng hình dịch vụ Azure Machine Learning 31](#_Toc167970607)

[Hình 42: Minh họa cấu hình datastore 31](#_Toc167970608)

[Hình 43: Minh họa tạo data asset 32](#_Toc167970609)

[Hình 44: Minh họa cấu hình data asset 32](#_Toc167970610)

[Hình 45: Minh họa cấu hình Source storage type 33](#_Toc167970611)

[Hình 46: Minh họa cấu hình Query 33](#_Toc167970612)

[Hình 47: Minh họa cấu hình Schema 34](#_Toc167970613)

[Hình 48: Minh họa Review Data Asset 34](#_Toc167970614)

[Hình 49: Minh họa chọn Designer 35](#_Toc167970615)

[Hình 50: Minh họa tạo Select columns 35](#_Toc167970616)

[Hình 51: Minh họa cấu hình Select columns 36](#_Toc167970617)

[Hình 52: Minh họa tạo và cấu hình Split Data 36](#_Toc167970618)

[Hình 53: Minh họa module Machine Learning Model và Train Model 37](#_Toc167970619)

[Hình 54: Minh họa thêm Score Model và Evaluate Model 37](#_Toc167970620)

[Hình 55: Minh họa tạo compute cluster 38](#_Toc167970621)

[Hình 56: Minh họa cấu hình pipeline job 38](#_Toc167970622)

[Hình 57: Minh họa Submit Model 39](#_Toc167970623)

[Hình 58: Minh họa ML chạy thành công 40](#_Toc167970624)

[Hình 59: Màn hình hiển thị chuyển dữ liệu lên EventHubs thành công 41](#_Toc167970625)

[Hình 60: Minh họa cấu hình Start Job cho Stream Analytics 42](#_Toc167970626)

[Hình 61: Minh họa dừng lấy dữ liệu thành công (Close) 42](#_Toc167970627)

[Hình 62: Minh họa chuyển dữ liệu sang Azure SQL Database thành công 43](#_Toc167970628)

[Hình 63: Minh họa dữ liệu ở Azure SQL database 43](#_Toc167970629)

[Hình 64: Minh họa chạy model 44](#_Toc167970630)

[Hình 65: Minh họa kiểm tra dữ liệu thành công 44](#_Toc167970631)

[Hình 66: Minh họa dữ liệu tải lên Blob Storage thành công 45](#_Toc167970632)

[Hình 67: Minh họa dữ liệu tải lên PowerBI thành công 45](#_Toc167970633)

[Hình 68: Minh họa dữ liệu trực quan của ML và Visualizations trên Power BI 46](#_Toc167970634)

# **MỤC LỤC BẢNG**

[Bảng 1: Bảng mô tả dữ liệu 14](#_Toc167970507)

[Bảng 2: Bản phân công công việc 48](#_Toc167970508)

# **NỘI DUNG**

# **CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI**

## **Mô tả đề tài**

Đề tài này yêu cầu xây dựng một hệ thống trực quan dữ liệu cổ phiếu dựa trên công nghệ xử lý luồng và phân tích thời gian thực trên Azure.

Hệ thống sẽ nhận dữ liệu mẫu để xử lý và trực quan:

* Dữ liệu chứng khoán được tạo ra và gửi đến Azure Event Hubs.
* Dữ liệu từ Event Hubs được xử lý bởi Azure Stream Analytics.
* Kết quả xử lý bởi Stream Analytics được lưu trữ vào Azure SQL Database.
* Dữ liệu từ Azure SQL Database được sử dụng để huấn luyện mô hình học máy bằng Azure Machine Learning.
* Kết quả từ Machine Learning được lưu trữ vào Azure SQL Database.
* Dữ liệu từ Azure SQL Database và kết quả từ Machine Learning được trực quan hóa bằng Power BI.

A diagram of a data flow

Description automatically generated

Hình 1: Sơ đồ hoạt động của hệ thống

Công nghệ sử dụng: Microsoft Azure, Visual Studio Code, ngôn ngữ lập trình Python.

## **Mục tiêu đề tài**

Xây dựng một hệ thống có thể thu thập dữ liệu chứng khoán thời gian thực cũng như xử lý dữ liệu với độ trễ tối đa 10 giây.

Huấn luyện các mô hình học máy có độ chính xác cao.

Cung cấp giao diện người dùng thân thiện để người dùng có thể dễ dàng tương tác với hệ thống.

Đồng thời, đề tài cũng nhằm mục đích tìm hiểu sâu hơn về các dịch vụ của Azure như:

* Event Hubs
* Stream Analytics
* Azure SQL Database
* Machine Learning và Power BI

Không chỉ thế dự án còn:

* Cung cấp một giải pháp tổng hợp và phân tích dữ liệu dạng luồng hiệu quả trên Azure.
* Giúp các tổ chức khai thác tối đa giá trị từ dữ liệu của họ.
* Góp phần thúc đẩy đổi mới trong lĩnh vực phân tích dữ liệu.

Lợi ích của đề tài này bao gồm:

* Tăng hiệu quả hoạt động của doanh nghiệp.
* Cải thiện việc ra quyết định.
* Giảm chi phí.
* Tạo ra các sản phẩm và dịch vụ mới.
* Nâng cao khả năng cạnh tranh

# **CHƯƠNG 2. LÝ THUYẾT TỔNG QUAN**

## **2.1. Giới thiệu**

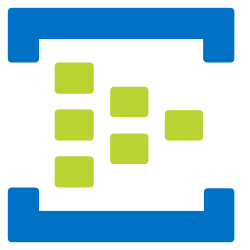
Hệ thống trực quan hóa dữ liệu cổ phiếu là một công cụ giúp nhà đầu tư theo dõi và phân tích biến động giá cổ phiếu theo thời gian thực. Hệ thống này sử dụng công nghệ xử lý luồng và phân tích dữ liệu tiên tiến để thu thập, xử lý và hiển thị dữ liệu giá cổ phiếu thời gian thực một cách trực quan bằng Azure. Nhờ đó, nhà đầu tư có thể dễ dàng nắm bắt được xu hướng thị trường, đưa ra quyết định mua bán cổ phiếu kịp thời và hiệu quả.

* **Hệ thống trực quan hóa (Data Visualization System)** là một công cụ giúp chuyển đổi dữ liệu thô thành các hình ảnh, biểu đồ, đồ thị dễ hiểu để người dùng có thể dễ dàng nắm bắt thông tin. Hệ thống này sử dụng các kỹ thuật trực quan như màu sắc, kích thước, hình dạng,... để thể hiện dữ liệu một cách trực quan và sinh động.
* **Dữ liệu cổ phiếu** là thông tin về giá cả, khối lượng giao dịch, và các yếu tố khác liên quan đến việc mua bán cổ phiếu của một công ty. Dữ liệu này được cung cấp bởi các sàn giao dịch chứng khoán và các nhà cung cấp dữ liệu tài chính.
* **Dữ liệu thời gian thực (real-time data)** là dữ liệu được thu thập, xử lý và truyền tải ngay khi nó được tạo ra, không cần lưu trữ trước. Dữ liệu này thường đến từ các nguồn liên tục thay đổi như cảm biến, mạng xã hội, giao dịch tài chính,...
* **Công nghệ xử lý luồng (Stream Processing Technology)** là một tập hợp các kỹ thuật và công cụ được sử dụng để thu thập, xử lý và phân tích dữ liệu theo thời gian thực. Dữ liệu dạng luồng có thể bao gồm dữ liệu từ các nguồn khác nhau như cảm biến, mạng xã hội, giao dịch tài chính,...
* **Azure** là nền tảng điện toán đám mây của Microsoft, cung cấp nhiều dịch vụ và công cụ giúp các doanh nghiệp và tổ chức xây dựng, triển khai và quản lý các ứng dụng, dịch vụ và cơ sở hạ tầng trên nền tảng đám mây

## **2.2. Kiến trúc hệ thống**

### **2.2.1. Azure Event Hubs**

Thu thập dữ liệu bán hàng từ các nguồn khác nhau như POS, CRM, website bán hàng,..



Hình 2: Biểu tượng Azure Event Hubs

### **2.2.2. Azure Stream Analytics**

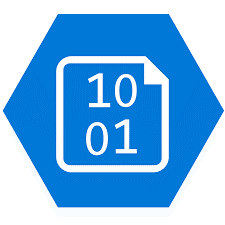
Xử lý dữ liệu bán hàng theo thời gian thực, trích xuất thông tin hữu ích và tạo ra các chỉ số kinh doanh.



Hình 3: Biểu tượng Azure Stream Analytics

### **2.2.3. Azure Blob Storage**

Azure Blob Storage là một dịch vụ lưu trữ đám mây mạnh mẽ và linh hoạt, cung cấp nhiều tính năng giúp người dùng lưu trữ, truy cập, quản lý, phân tích và bảo vệ dữ liệu phi cấu trúc một cách dễ dàng và hiệu quả



Hình 4: Biểu tượng Azure Blob Storage

### **2.2.4. Azure SQL Database**

Lưu trữ dữ liệu giá cổ phiếu đã được xử lý .



Hình 5: Biểu tượng Azure SQL Database

### **2.2.5. Azure Machine Learning**

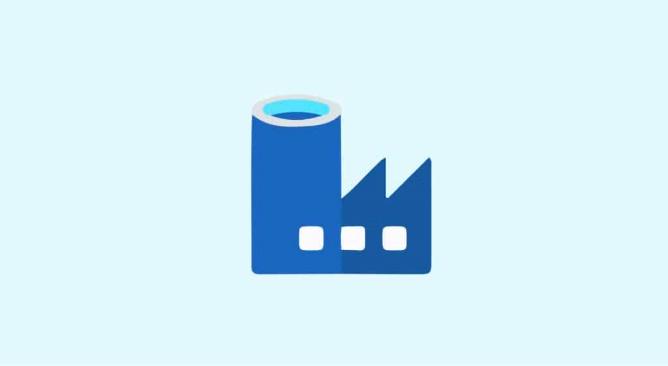
Azure Machine Learning là một dịch vụ cho phép bạn xây dựng, triển khai và quản lý mô hình học máy.



Hình 6: Biểu tượng Azure Machine Learning

### **2.2.6. Azure Data Factory**

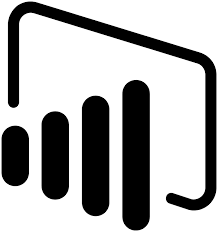
Azure Data Factory: Azure Data Factory là một dịch vụ tích hợp dữ liệu dựa trên đám mây cho phép chúng ta tạo các quy trình làm việc dựa trên dữ liệu trong đám mây để điều phối và tự động hóa việc di chuyển dữ liệu và biến đổi dữ liệu.



Hình 7: Biểu tượng Azure Data Factory

### **2.2.7. Microsoft Power BI**

Trực quan hóa dữ liệu giá cổ phiếu dưới dạng biểu đồ, đồ thị, báo cáo, ...



Hình 8: Biểu tượng Microsoft Power BI

# **CHƯƠNG 3. Dữ liệu và cấu hình**

## **3.1. Mô tả dữ liệu**

Dữ liệu: Giá cổ phiểu NVIDIA real time được lấy từ API của website google Finance bằng

yfinance là một thư viện mã nguồn mở phổ biến được viết bằng Python để truy cập dữ liệu tài chính từ Yahoo Finance. Nó cung cấp các hàm đơn giản để lấy dữ liệu lịch sử và thời gian thực về giá cổ phiếu, khối lượng giao dịch, thông tin công ty và nhiều hơn nữa. Thư viện này được sử dụng rộng rãi bởi các nhà đầu tư, nhà phân tích tài chính và các nhà khoa học dữ liệu để nghiên cứu thị trường chứng khoán, xây dựng mô hình tài chính và phát triển các ứng dụng tài chính

Định dạng: JSON

Yahoo Finance là một nền tảng trực tuyến uy tín cung cấp nhiều thông tin và công cụ tài chính miễn phí cho người dùng cá nhân và nhà đầu tư. Nổi tiếng với giao diện trực quan và dễ sử dụng, Yahoo Finance mang đến cho bạn trải nghiệm toàn diện về thị trường tài chính toàn cầu.

Nguồn dữ liệu: https://finance.yahoo.com/quote/NVDA

Bảng mô tả dữ liệu như sau:

|  |  |  |
| --- | --- | --- |
| Tên thuộc tính | Mô tả | Kiểu |
| currentTime  (yfinance tạo ) | Thời gian hiện tại | datetime |
| open | Giá mở cửa | float |
| symbol | Ký hiệu của cổ phiếu | string |
| currentPrice | Giá hiện tại | float |
| marketCap | Giá trị vốn hóa thị trường của công ty, là tổng giá trị của tất cả các cổ phiếu trong lưu thông. | big int |
| volume | Khối lượng giao dịch | big int |

Bảng 1: Bảng mô tả dữ liệu

## **3.2. Cài đặt và cấu hình trên Azure**

Thao tác trên nhánh 1:

* Bước 1: Tạo Event Hubs

A computer screen with a blue and white screen

Description automatically generated

Hình 9: Minh họa tạo Event Hubs

* Bước 2: Cấu hình Event Hubs

A screenshot of a computer

Description automatically generated

Hình 10: Minh họa cấu hình Event Hubs.

* Bước 3: Tạo 1 sự kiện trong Event Hubs

A screenshot of a computer

Description automatically generated

Hình 11: Minh họa tạo sự kiện trong Event Hubs

* Bước 4: Tạo 1 Shared access policies. Tạo Shared Access policies, nhấn Add, tiếp theo tích chọn vào Manage, Listen, Send và cuối cùng là nhấn Save.

A screenshot of a computer

Description automatically generated

Hình 12: Minh họa tạo Shared Access Policies.

* Bước 5: Tạo và cấu hình phân tích luồng Stream Analytics Job

A screenshot of a computer

Description automatically generated

Hình 13: Minh họa cấu hình Stream Analytics job mới

* Bước 6: Tạo input cho Stream Analytics Job. Với Input là Event hub vừa tạo ở bước trước đó.

A screenshot of a computer

Description automatically generated

Hình 14: Minh họa cấu hình input

* Bước 7: Tạo Storage Account

A screenshot of a computer

Description automatically generated

Hình 15: Minh họa cấu hình storage account

* Bước 8: Tạo output cho Stream Analytics. Chọn Query trong Job topology rồi bấm dấu “+” ở outputs sau đó chọn Blob storage/ADLS Gen2

A screenshot of a computer

Description automatically generated

Hình 16: Minh họa tạo output bằng Blob storage

* Bước 9: Tạo output Power BI cho Stream Analytics

A screenshot of a computer

Description automatically generated

Hình 17: Minh họa tạo output bằng Power BI

* Bước 10: Cấu hình output Power BI. Đăng nhập bằng tài khoản Miscrosoft sử dụng trên Power BI

A screenshot of a computer

Description automatically generated

Hình 18: Minh họa cấu hình output Power BI

* Bước 11: Tạo file python trên Visual Studio để gọi dữ liệu từ API. Code kết nối từ API lên Event Hubs:
* connection\_str = Endpoint=sb://cloudproject.servicebus.windows.net/
* SharedAccessKeyName=PstockPrice
* SharedAccessKey=” Lấy ở bước tạo Shared AccessPolicy”
* EntityPath=projectstockprice
* eventhubs\_name = 'projectstockprice'

A screenshot of a computer

Description automatically generated

Hình 19: Màn hình file Python gọi dữ liệu từ API đến EventHubs

Thao tác trên nhánh 2:

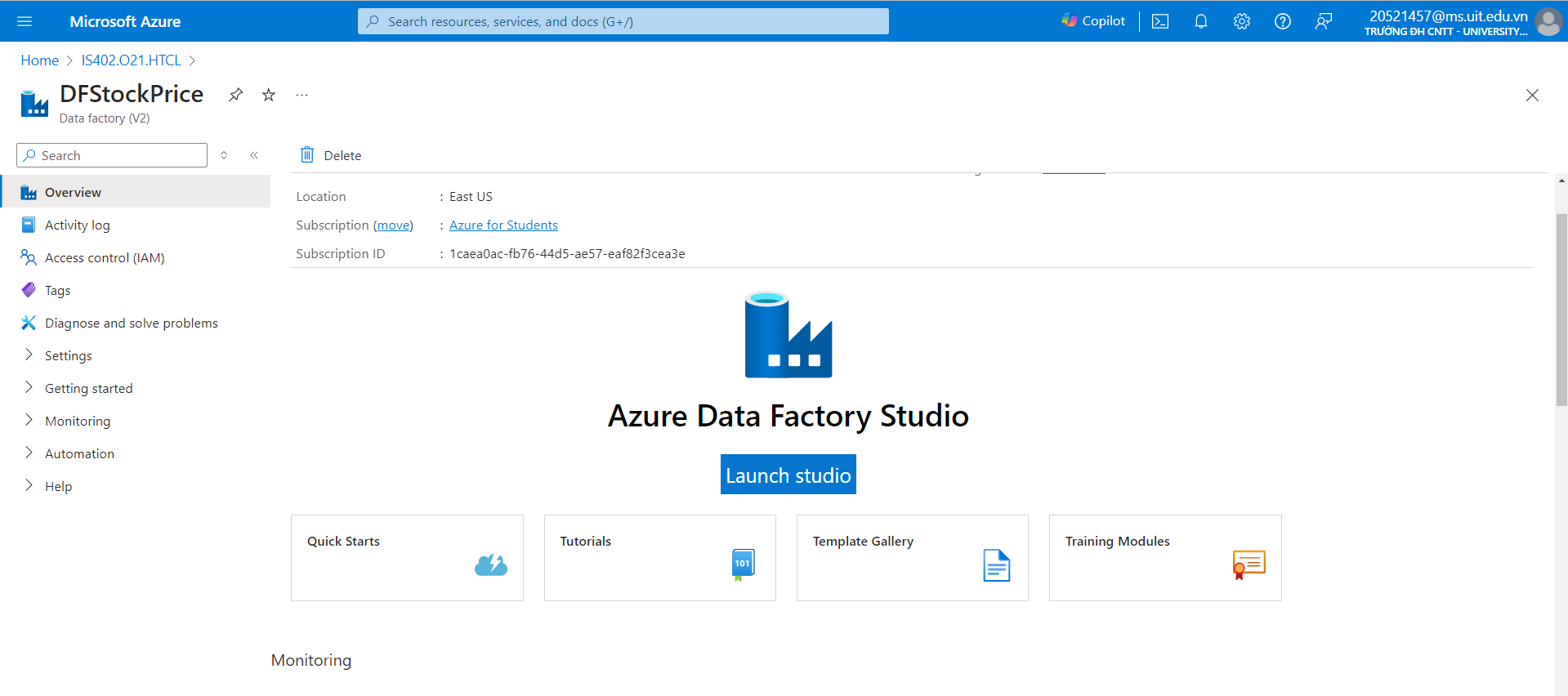
* Bước 1: Tạo Data Factory để vận chuyển dữ liệu từ blob storage qua Azure SQL Database

A screenshot of a computer

Description automatically generated

Hình 20: Minh họa cấu hình Data Factory

* Bước 2: Chọn Launch Studio để mở giao diện của Data Factory



Hình 21: Minh họa trang trang khởi động Data Factory

* Bước 3: Tạo dataset mới trong Data Factory

A screenshot of a computer

Description automatically generated

Hình 22: Minh họa tạo Dataset trên Data Factory

* Bước 4: Chọn Blob Storage để lấy dữ liệu từ Blob storage qua

A screenshot of a computer

Description automatically generated

Hình 23: Minh họa chọn nguồn dữ liệu

* Bước 5: Chọn định dạng kiểu dữ liệu kiểu JSON

A screenshot of a computer

Description automatically generated

Hình 24: Minh họa chọn kiểu dữ liệu

* Bước 6: Tạo 1 Linked service mới để lấy dữ liệu từ Blob Storage Gen 2

A screenshot of a computer

Description automatically generated

Hình 25: Minh họa tạo linked service

* Bước 7: Cấu hình linked service tới Azure Blob Storage và chọn Storage Account phù hợp đã tạo trên nhánh 1.

A screenshot of a computer

Description automatically generated

Hình 26: Minh họa cấu hình linked service (1)

* Bước 8: Chọn File path cho đúng đường dẫn dữ liệu bấm Preview data để xem thử data.

A screenshot of a computer

Description automatically generated

Hình 27: Minh họa cấu hình linked service (2)

A screenshot of a computer

Description automatically generated

Hình 28: Minh họa kết quả preview data

* Bước 9: Tạo SQL database.

A screenshot of a computer

Description automatically generated

Hình 29: Minh họa cấu hình SQL Database

* Bước 10: Tạo liên kết từ dataset để truyền dữ liệu đến SQL Database vừa tạo.

A screenshot of a computer

Description automatically generated

Hình 30: Minh họa liên kết dataset đến SQL Database

* Bước 11: Tạo Linked service mới cho SQL Database vừa tạo.

A screenshot of a computer

Description automatically generated

Hình 31: Minh họa tạo Linked service cho SQL Database.

* Bước 12: Tạo linked service và chọn Server name và Database name phù hợp.

A screenshot of a computer

Description automatically generated

Hình 32: Minh họa tạo linked service mới

* Bước 13: Tạo pipeline mới.

A screenshot of a computer

Description automatically generated

Hình 33: Minh họa tạo pipeline

* Bước 14: Chọn activities

A screenshot of a computer

Description automatically generated

Hình 34: Minh họa chọn activities

* Bước 15: Cấu hình copy data Activities.



Hình 35: Minh họa copy data Activities

* Bước 16: Cấu hình Source

A screenshot of a computer

Description automatically generated

Hình 36: Minh họa cấu hình Source

* Bước 17: Cấu hình Sink

A screenshot of a computer

Description automatically generated

Hình 37: Minh họa cấu hình Sink

* Bước 18: Đăng nhập để kiểm tra dữ liệu.

A computer screen shot of a computer

Description automatically generated

Hình 38: Minh họa cấu hình Query editor

* Bước 19: Preview Data

A screenshot of a computer

Description automatically generated

Hình 39: Minh họa Preview Data

* Bước 20: Tạo Azure Machine Learning

A screenshot of a computer

Description automatically generated

Hình 40: Minh họa cấu hình Machine Learning

* Bước 21: Vào workspace của ML launch studio và tạo datastore

A screenshot of a computer

Description automatically generated

Hình 41: Màng hình dịch vụ Azure Machine Learning

* Bước 22: Tạo datastore cho Azure ML

A screenshot of a computer

Description automatically generated

Hình 42: Minh họa cấu hình datastore

* Bước 23: Tạo data asset cho Azure ML

A screenshot of a computer

Description automatically generated

Hình 43: Minh họa tạo data asset

* Bước 24: Cấu hình data asset Data source chọn SQL Database

A screenshot of a computer

Description automatically generated

Hình 44: Minh họa cấu hình data asset

* Bước 25: Cấu hình Source storage type

A screenshot of a computer

Description automatically generated

Hình 45: Minh họa cấu hình Source storage type

* Bước 26: Cấu hình Query

A screenshot of a computer

Description automatically generated

Hình 46: Minh họa cấu hình Query

* Bước 27: Cấu hình Schema và bấm create

A screenshot of a computer

Description automatically generated

Hình 47: Minh họa cấu hình Schema

A screenshot of a computer

Description automatically generated

Hình 48: Minh họa Review Data Asset

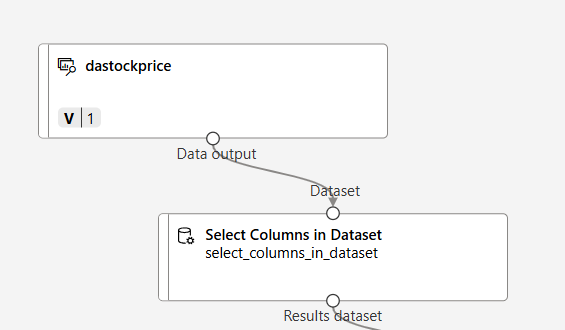
* Bước 28: Vào phần Designer trong Azure ML để thiết kế luồng cho ML

A screenshot of a computer

Description automatically generated

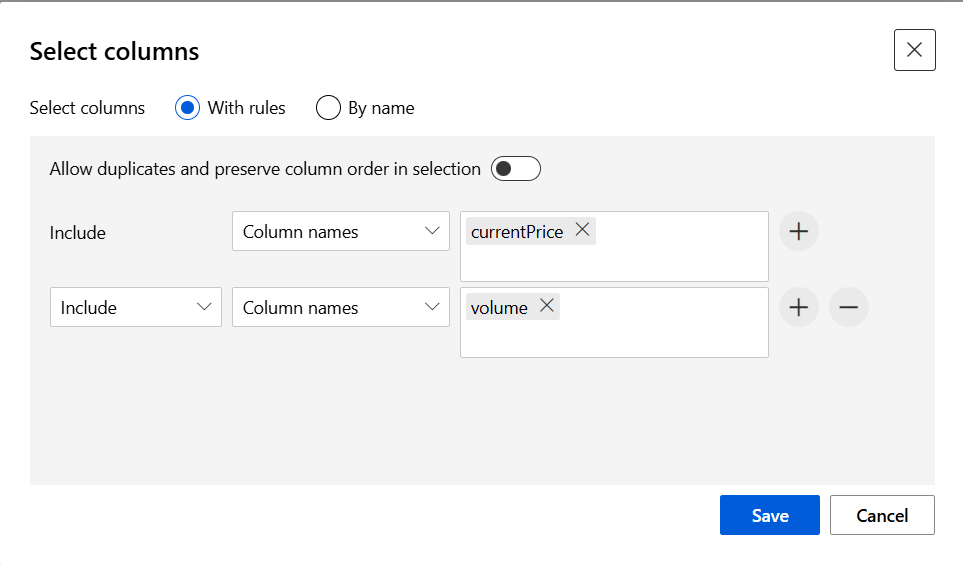
Hình 49: Minh họa chọn Designer

* Bước 29: Kéo thả phần Data đã liên kết và tạo Select Columns is Dataset



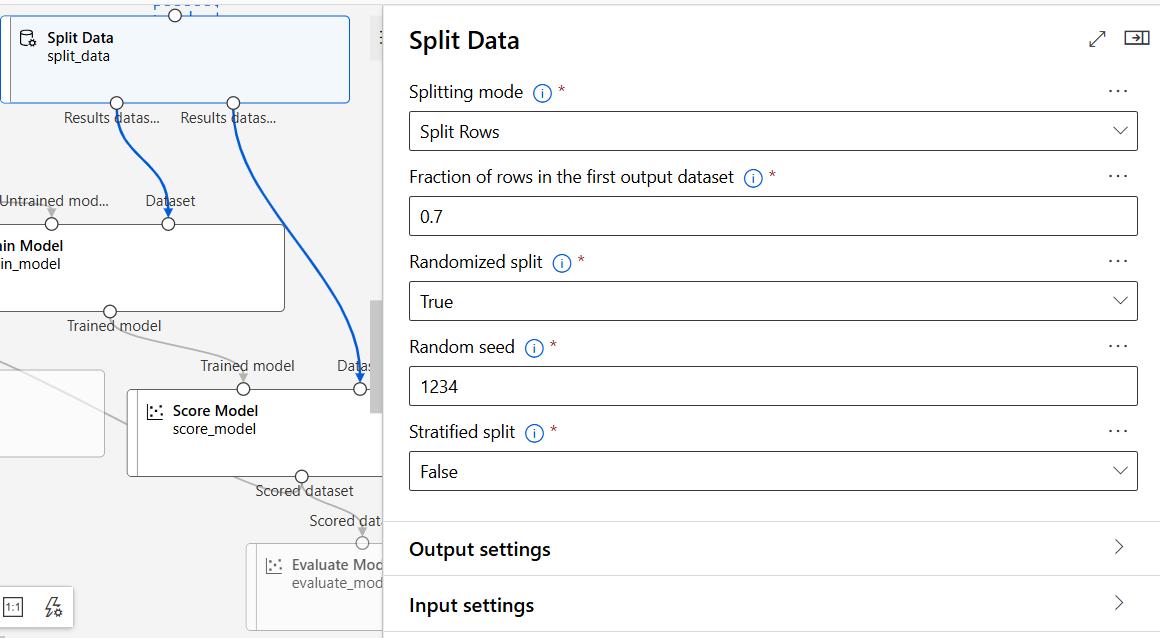
Hình 50: Minh họa tạo Select columns

* Bước 30: Cấu hình Select columns và chọn 2 thuộc tính cần thiết cho việc dự đoán.



Hình 51: Minh họa cấu hình Select columns

* Bước 31: Tạo và cấu hình Split Data chia thành 2 tập Train và Test



Hình 52: Minh họa tạo và cấu hình Split Data

* Bước 32: Chọn Machine Learning Model và Train Model với ML được chọn là Linear Regression, Neural network regression và Decision Forest để dự đoán giá cổ phiếu (currentPrice)

A screenshot of a computer

Description automatically generated

Hình 53: Minh họa module Machine Learning Model và Train Model

* Bước 33: Sau đó tạo thêm Score Model và Evaluate Model để đánh giá Model. Và bấm Save để lưu mô hình.

A screenshot of a computer

Description automatically generated

Hình 54: Minh họa thêm Score Model và Evaluate Model

* Bước 34: Chọn compute cluster để chạy Model Machine Learning.

A screenshot of a computer

Description automatically generated

Hình 55: Minh họa tạo compute cluster

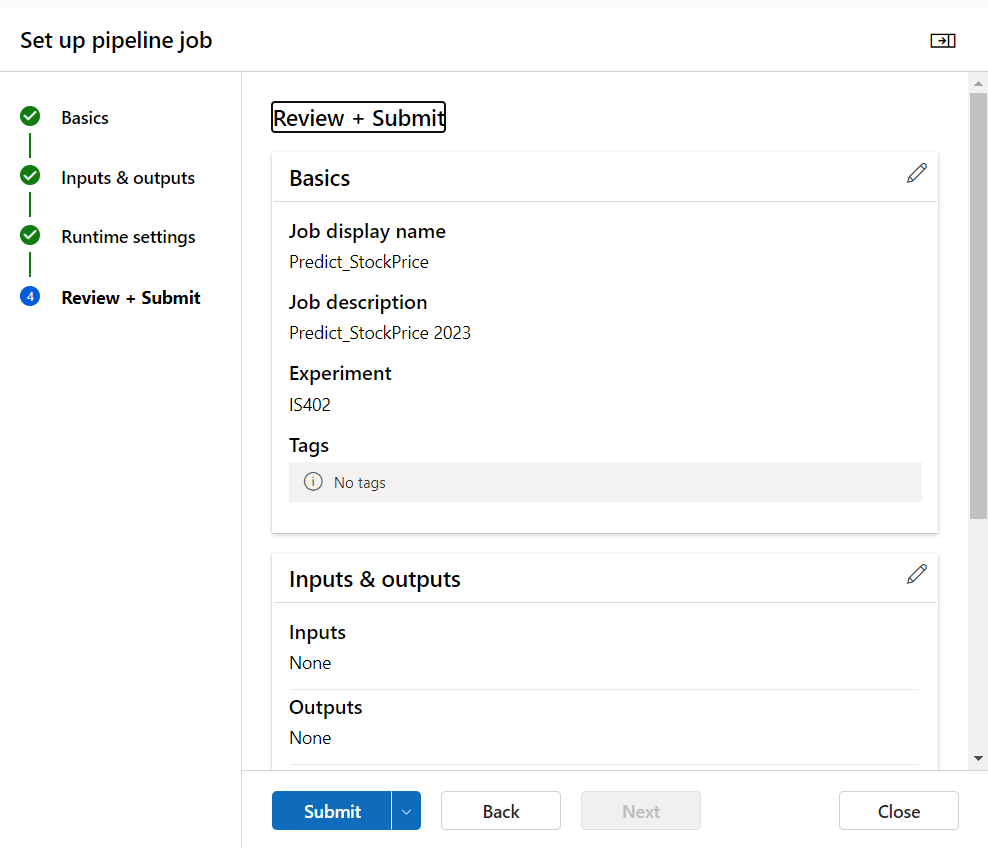
* Bước 35: Cấu hình pipeline job

A screenshot of a computer job

Description automatically generated

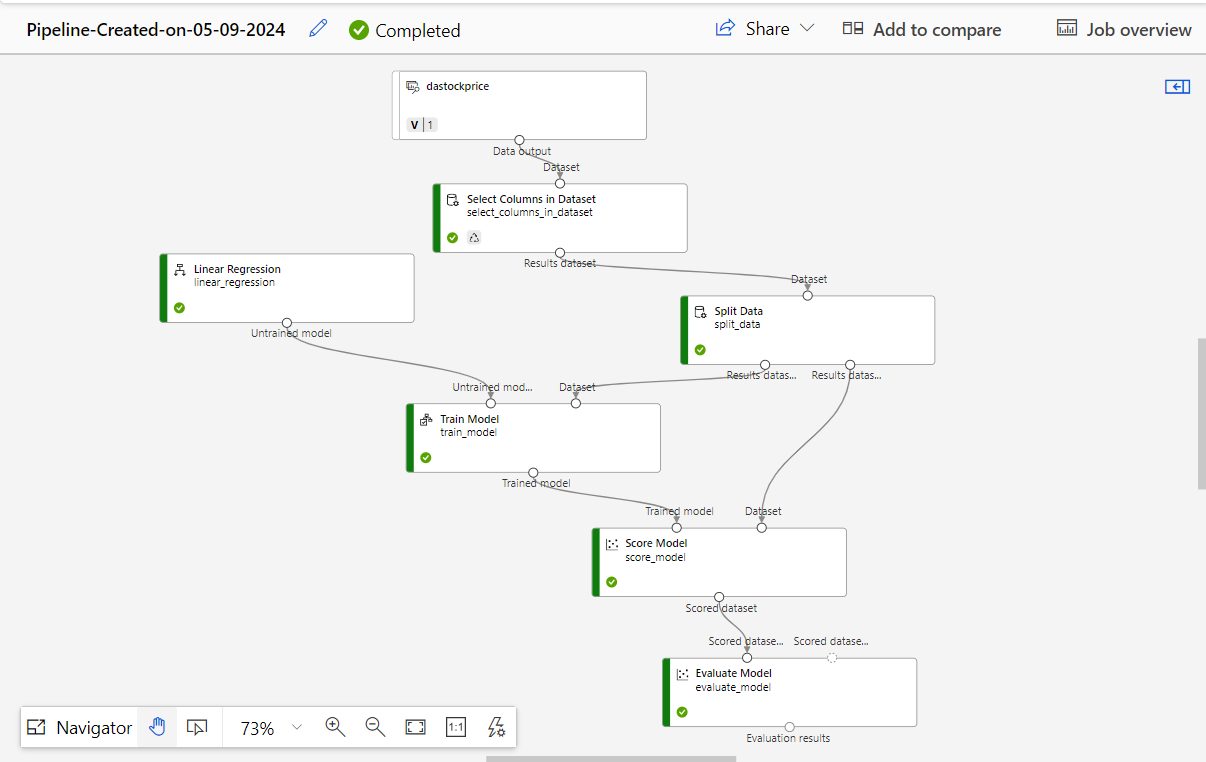
Hình 56: Minh họa cấu hình pipeline job

* Bước 36: Tại mục Review + Submit bấm Submit để chạy pipeline



Hình 57: Minh họa Submit Model

* Bước 37: Sơ đồ hoạt động đã chạy thành công của Machine Learning



Hình 58: Minh họa ML chạy thành công

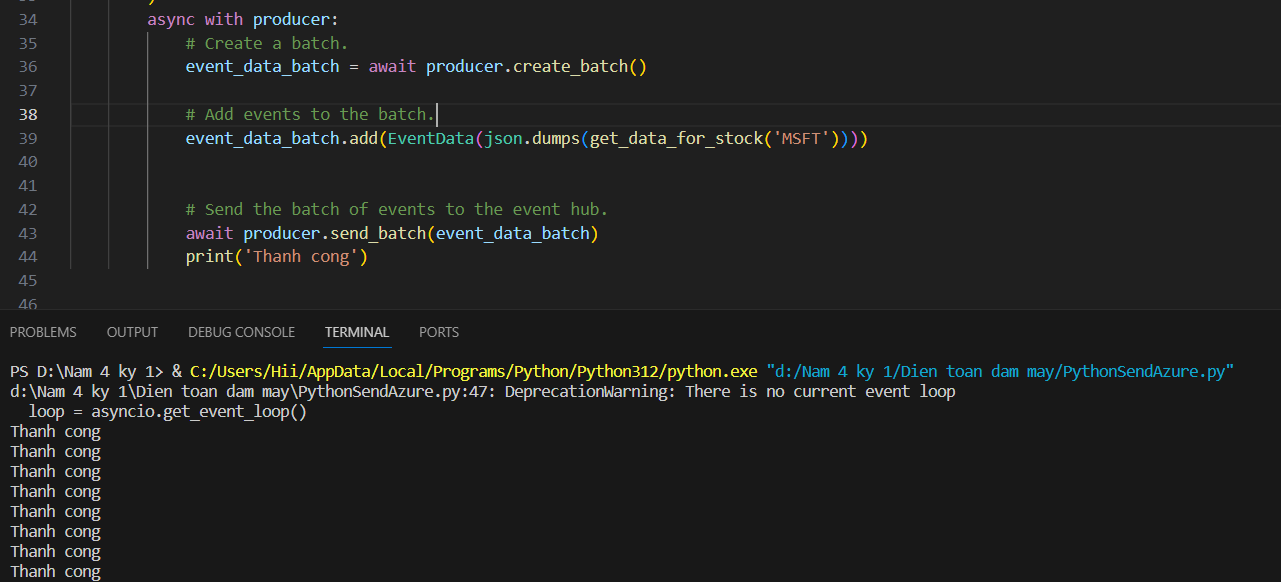
# **CHƯƠNG 4. THỰC THI VÀ KẾT QUẢ**

## **4.1. Thực thi trên Azure**

Luồng chạy hoạt động trên hệ thống:

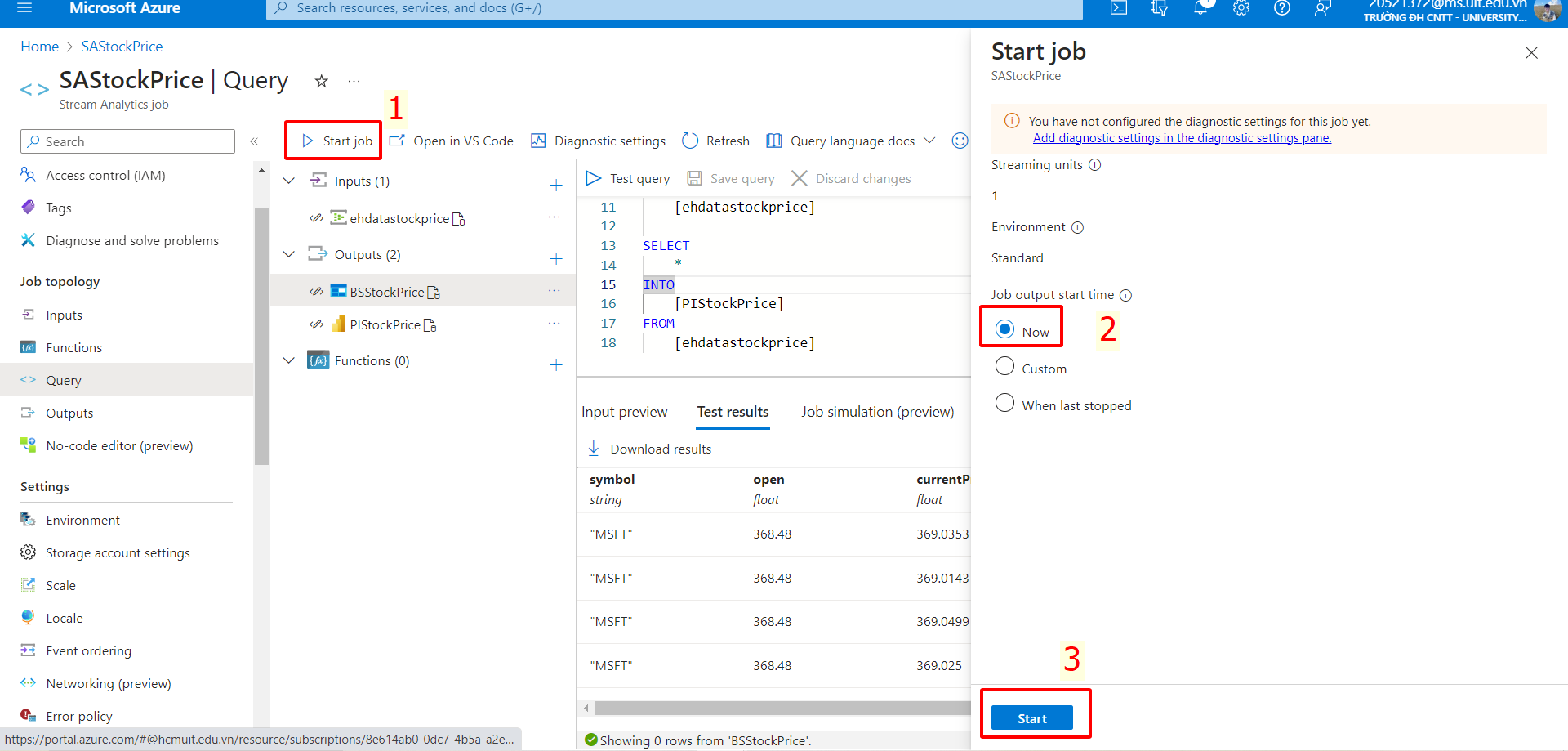
Nhánh 1:

* Bước 1: Lấy dữ liệu mẫu được từ API. Sau khi chạy file Python, chờ dữ liệu được chạy vào EventHubs.



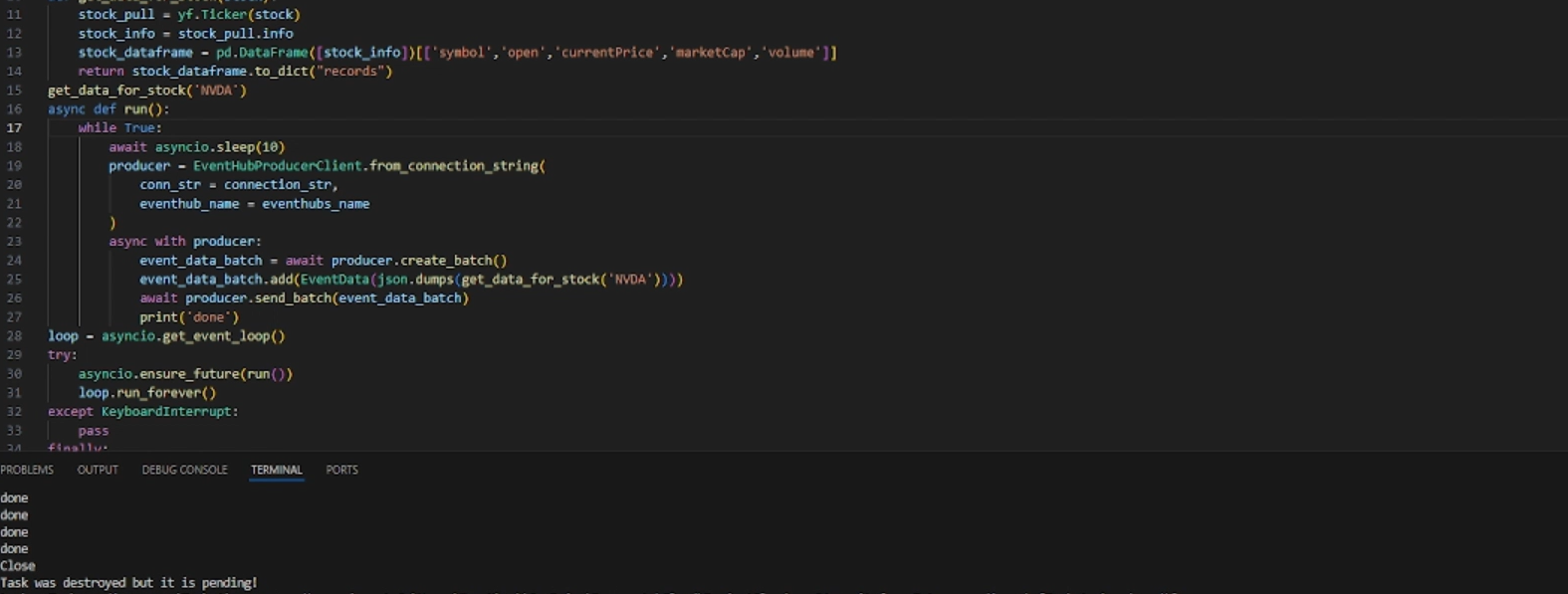
Hình 59: Màn hình hiển thị chuyển dữ liệu lên EventHubs thành công

* Bước 2: Stream Analytics nhận dữ liệu từ Event Hub và Blob Storage, sau đó xuất dữ liệu dưới dạng JSON qua Visual Power BI hoặc lưu vào Blob Storage. Chọn Start job để bắt đầu công việc cho Stream Analytics, sau đó chọn Start để bắt đầu công việc chuyển dữ liệu vào Blog Storage và Power BI.



Hình 60: Minh họa cấu hình Start Job cho Stream Analytics

* Bước 3: Trong file Python tiến hành dừng lấy dữ liệu.



Hình 61: Minh họa dừng lấy dữ liệu thành công (Close)

* Bước 4: Azure Blob Storage lưu trữ dữ liệu từ Stream Analytics dưới dạng nhị phân. Khi yêu cầu, dữ liệu sẽ được lấy từ Blob Storage qua Stream Analytics.

Nhánh 2:

* Bước 1: Sau khi cài đặt và cấu hình xong tiến hành chuyển dữ liệu từ Blod Storage sang Azure SQL Database. Sau khi cấu hình xong chọn vào debug để chạy chuyển dữ liệu sang Azure SQL Database. Chọn debug.

A computer screen shot of a computer

Description automatically generated

Hình 62: Minh họa chuyển dữ liệu sang Azure SQL Database thành công

* Bước 2: Chuyển sang trang Azure SQL đăng nhập và kiểm tra dữ liệu:

A screenshot of a computer

Description automatically generated

Hình 63: Minh họa dữ liệu ở Azure SQL database

* Bước 3: Chọn Configure & Submit để tiến hành chạy Model và chờ đợi.

A screenshot of a computer

Description automatically generated

Hình 64: Minh họa chạy model

* Bước 4: Kiểm tra kết quả và download dữ liệu bằng file csv ở Score Model.

A screenshot of a computer

Description automatically generated

Hình 65: Minh họa kiểm tra dữ liệu thành công

* Bước 5: Tiến hành visualize dữ liệu lên PowerBI.

## **4.2. Kết quả và đánh giá.**

Kiểm tra dữ liệu ở Blog Storage:

A screenshot of a computer

Description automatically generated

Hình 66: Minh họa dữ liệu tải lên Blob Storage thành công

Kiểm tra dữ liệu ở PowerBI:

A graph on a white background

Description automatically generated

Hình 67: Minh họa dữ liệu tải lên PowerBI thành công

Kiểm Tra dữ liệu trực quan của ML và Visualizations trên Power BI

A screen shot of a graph

Description automatically generated

Hình 68: Minh họa dữ liệu trực quan của ML và Visualizations trên Power BI

# **CHƯƠNG 5. TỔNG KẾT**

## **5.1. Kết luận**

Với kiến thức và nghiên cứu trong học tập thì nhóm chúng em đã hoàn thành được các mục tiêu sau:

* Azure Event Hubs: đã thu thập dữ liệu thời gian thực
* Azure Stream Analytics: Cấu hình thành công cho việc nhận dữ liệu đầu vào từ Event Hubs và đưa dữ liệu đầu ra lưu trữ ở Blob Storage và đưa qua Power BI
* Azure SQL Database: Lưu trữ dữ liệu trên cloud
* Azure Machine Learning: Xây dựng, triển khai và dự đoán trên thời gian thực
* Data Factory: Quản lý việc vận chuyển, biến đổi dữ liệu
* Power BI: Trực quan dữ liệu từ cả hai nhánh hoạt động.

## **5.2. Hướng phát triển**

Tuy đã hoàn thành được một số mục tiêu nhưng các dịch vụ và công cụ trên Azure còn nhiều tiềm năng để phục vụ cho một số bài toán hay việc tối ưu hóa luồng hoạt động. Nếu còn thời gian để tiếp tục nghiên cứu thì nhóm chúng em sẽ khai thác thêm các tiềm năng của các dịch vụ trên Azure để phục vụ cho việc học tập nghiên cứu trong tương lai.

# **BẢNG PHÂN CÔNG CÔNG VIỆC**

|  |  |  |
| --- | --- | --- |
| Thành viên | Công việc | Hoàn thành |
| Nguyễn Xuân Tuấn Kiệt | Thuyết trình, hỗ trợ cấu hình đồ án | 100% - Hoàn thành tốt |
| Trần Duy Khánh | Tìm tài liệu, cấu hình đồ án | 100% - Hoàn thành tốt |
| Trần Thạnh Phong | Viết báo cáo, hỗ trợ cấu hình đồ án | 100% - Hoàn thành tốt |

Bảng 2: Bản phân công công việc

# **TÀI LIỆU THAM KHẢO**

* <https://learn.microsoft.com/en-us/azure/stream-analytics/event-hubs-parquet-capture-tutorial>
* <https://learn.microsoft.com/en-us/azure/stream-analytics/stream-analytics-real-time-fraud-detection>
* https://learn.microsoft.com/en-us/azure/architecture/solution-ideas/articles/demand-forecasting