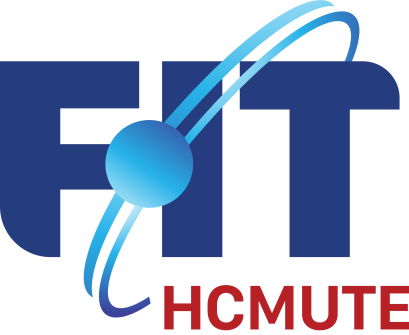
**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

✩✩✩✩✩



**Ứng dụng Big Data: Truyền dữ liệu trong thời gian thực**

**BÁO CÁO**

**ĐỀ TÀI: TÌM HIỂU VỀ APACHE KAFKA,**

**APACHE SPARK VÀ ỨNG DỤNG**

**GVHD: ThS. LÊ MINH TÂN**

**NHÓM SINH VIÊN THỰC HIỆN: NHÓM 8**

|  |  |
| --- | --- |
| **SVTH** | **MSSV** |
| Nguyễn Thanh Tuấn | 20133110 |
| Phạm Trung Nghĩa | 20133073 |
| Ngô Hoàng Khắc Tường | 20133112 |
| Nguyễn Thị Tường Vi | 20133113 |

**Tp.Hồ Chí Minh, tháng 10 năm 2023**

**LỜI CẢM ƠN**

*Đầu tiên, nhóm em xin chân thành gửi lời cảm ơn với Lê Minh Tân – Giảng viên phụ trách môn Ứng dụng Bigdata Truyền dữ liệu trong thời gian thực – Trường Đại học Sư phạm kỹ thuật TP.HCM.*

*Trong thời gian học tập, nhóm em nhận được sự nhiều sự giúp đỡ từ Thầy. Thầy đã cung cấp cho chúng em đầy đủ kiến thức, chỉ bảo và đóng góp những ý kiến quý báu giúp chúng em có thêm hiểu biết để hoàn thành được báo cáo cuối kỳ cho môn học. Thầy đã cung cấp tài liệu và hướng dẫn tận tình cho chúng em trong suốt quá trình học tập. Thầy luôn nhiệt tình vui vẻ chỉ dạy, chia sẻ giúp chúng em trong môn học. Chúng em rất trân quý sự tâm huyết và trách nhiệm của Thầy trong công việc giảng dạy và truyền đạt kiến thức.*

*Trong quá trình thực hiện đồ án, dựa trên kiến thức được Thầy cung cấp qua các buổi học lý thuyết cũng như thực hành trên lớp, kết hợp với việc tự tìm hiểu những công cụ và kiến thức mới , nhóm đã cố gắng thực hiện đồ án một cách tốt nhất .Tuy nhiên, đồ án còn chưa được hoàn hiện và có thể còn có sai sót.*

*Nhóm rất mong nhận được sự góp ý từ Thầy nhằm rút ra những kinh nghiệm quý báu và hoàn thiện vốn kiến thức để nhóm có thể hoàn thành những đồ án, dự án khác trong tương lai .*

*Nhóm chúng em xin chân thành cảm ơn Thầy!*

**MỤC LỤC**

[CHƯƠNG 1 TỔNG QUAN VỀ APACHE KAFKA 1](#_Toc150894408)

[1.1 Khái niệm Apache Kafka 1](#_Toc150894409)

[1.2 Cấu trúc của Apache Kafka 1](#_Toc150894410)

[1.3 Cơ chế hoạt động của Apache Kafka 4](#_Toc150894411)

[1.4 Ưu nhược điểm Apache Kafka 4](#_Toc150894412)

[1.5 Ứng dụng của Apache Kafka 6](#_Toc150894413)

[CHƯƠNG 2 TỔNG QUAN VỀ APACHE SPARK 7](#_Toc150894414)

[2.1 Khái niệm Apache Spark 7](#_Toc150894415)

[2.2 Thành phần Apache Spark 7](#_Toc150894416)

[2.3 Ưu nhược điểm của Apache Spark 8](#_Toc150894417)

[2.4 Ứng dụng của Apache Spark 9](#_Toc150894418)

[CHƯƠNG 3 ỨNG DỤNG KAFKA KẾT HỢP SPARK 11](#_Toc150894419)

[3.1 Giới thiệu tổng quan về tập dữ liệu 11](#_Toc150894420)

[3.2 Giới thiệu tổng quan về ứng dụng 11](#_Toc150894421)

**DANH SÁCH HÌNH ẢNH**

[Hình 1.2.1 Mô hình cấu trúc kafka đơn giản 1](#_Toc150894485)

[Hình 1.2.2 Mô hình cấu trúc kafka chi tiết 2](#_Toc150894486)

[Hình 1.2.3 Cơ chế hoạt động của Kraft Mode 3](#_Toc150894487)

[Hình 1.3.1 Mô hình cơ chế hoạt động của Apache Kafka 4](#_Toc150894488)

[Hình 2.1.1 Thông tin về mã nguồn mở Apache Spark 7](#_Toc150894489)

[Hình 2.2.1 Các thành phần chính của Apache Spark 8](#_Toc150894490)

[Hình 3.2.1 Mô hình Kafka kết hợp Spark 12](#_Toc150894491)

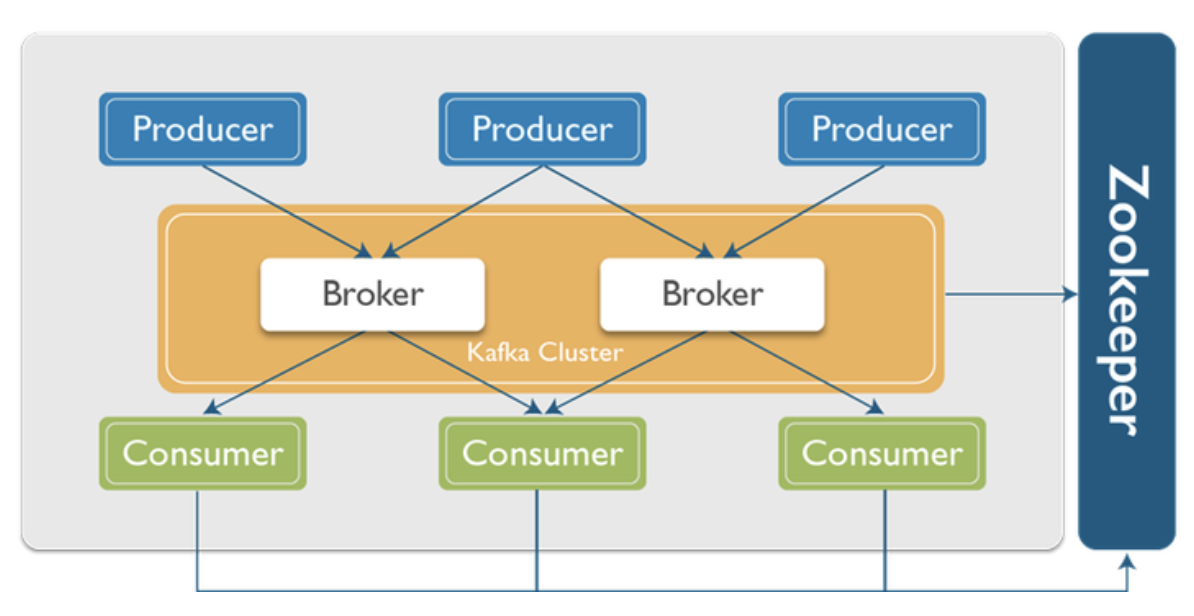
# TỔNG QUAN VỀ APACHE KAFKA

## Khái niệm Apache Kafka

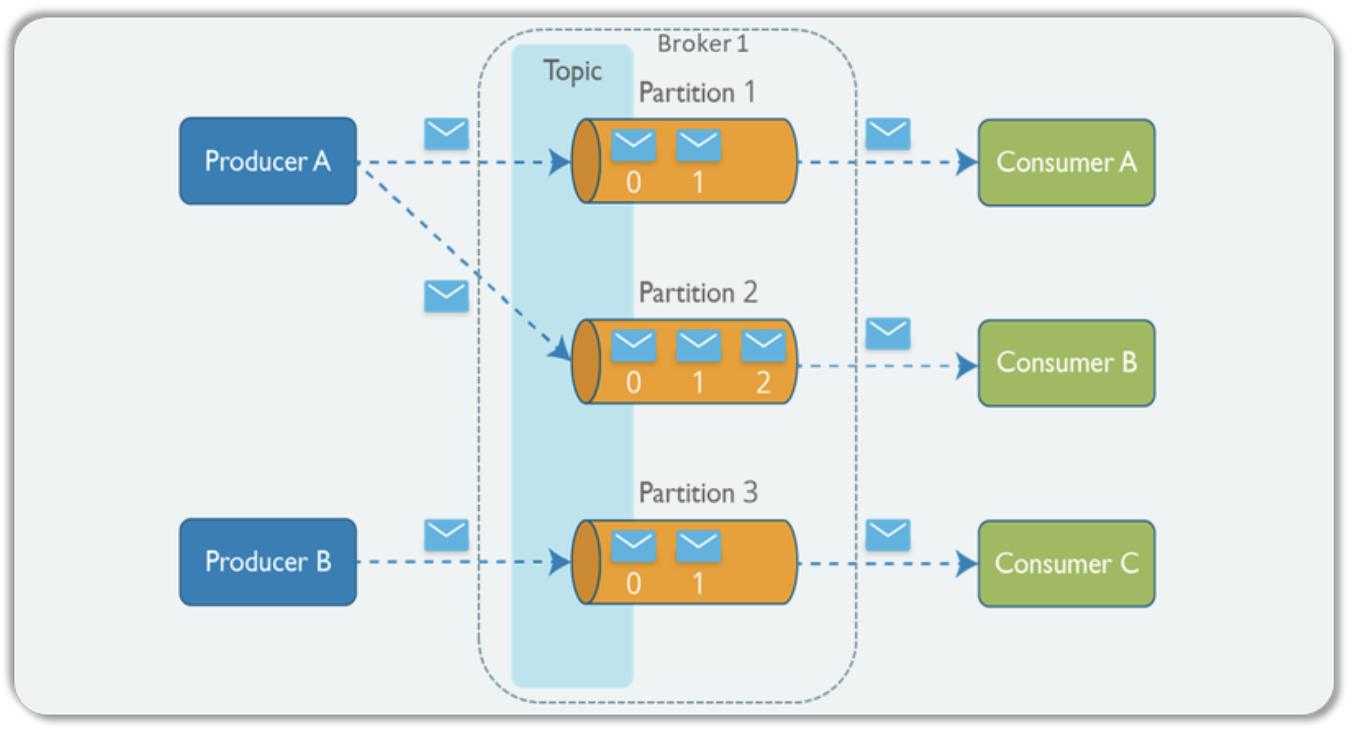
Apache Kafka là một nền tảng phát trực tuyến sự kiện phân tán mã nguồn mở. Apache Kafka được xây dựng nhằm mục đích xử lý dữ liệu streaming real-time (theo thời gian thực). Kafka hoạt động như một hệ thống hàng đợi (message queue) phân tán, cho phép các ứng dụng gửi và nhận các thông điệp (messages) theo mô hình xuất bản-đăng ký (publish-subscribe) hoặc theo mô hình chủ đề (topic). Kafka được viết bằng Java và Scala.

Với Kafka, bạn có thể tạo ra các hệ thống thời gian thực, từ việc giám sát dữ liệu trực tiếp đến việc phân tích dữ liệu lớn để tìm ra xu hướng và mẫu. Ví dụ, hãy tưởng tượng bạn đang giám sát các trạng thái của máy móc trong một nhà máy. Các máy móc sẽ gửi thông tin về trạng thái của chúng vào Kafka, và bạn có thể có các ứng dụng đọc thông tin này để phát hiện sớm sự cố và thậm chí dự đoán khi nào máy móc cần bảo trì.

## Cấu trúc của Apache Kafka



### Mô hình cấu trúc kafka đơn giản



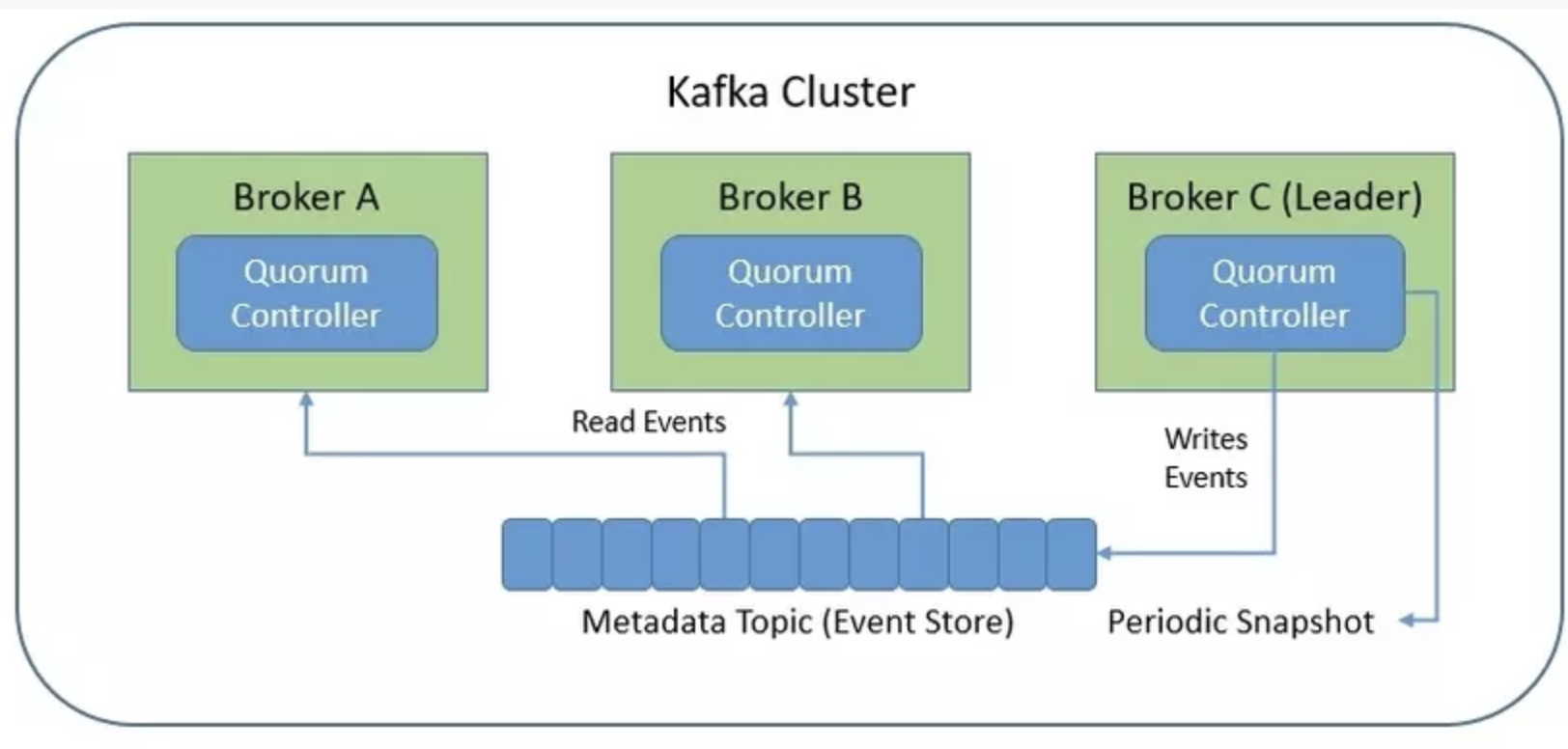
### Mô hình cấu trúc kafka chi tiết

Mô hình cấu trúc Kafka bao gồm các thành phần sau:

* Message: Thông tin được gửi đi, có thể là text, binary, Json.
* Producer: Chương trình/service tạo ra message, đẩy message publish vào Topic.
* Broker: Một host có thể chạy nhiều server kafka, mỗi server như vậy gọi là một broker. Các broker này cùng trỏ tới chung 1 zookeeper, gọi là cụm broker(hay là Clusters). Broker là nơi chứa các partition. Một broker có thể chứa nhiều partition.
* Partition: Nơi lưu trữ message của topic. Một topic có thể có nhiều partition
* Consumer: Chương trình/service có chức năng subscribe vào một Topic để tiêu thụ, xử lý các message đó.
* Zookeeper: được dùng để quản lý và bố trí các broker.

Zookeeper có những lợi ích nhất định đối với hệ thống, tuy nhiên trong một số trường hợp zookeeper làm tăng độ phức tạp của hệ thống, tăng khả năng lỗi do có nhiều thành phần,… Vì vậy, Kafka Kraft đã xuất hiện.

Kafka Kraft là một giao thức đồng thuận mới được phát triển nhằm mục đích loại bỏ đi sự hiện diện của Zookeeper trong hệ thống. Kafka sẽ tự lưu trữ các metadata trên chính nó. Sử dụng giao thức này ta cần bật chế độ Kraft trên cụm Kafka, Kraft sẽ sử dụng Qurom Controller mới với giao thức đồng thuận Raft để bầu chọn leader.



### Cơ chế hoạt động của Kraft Mode

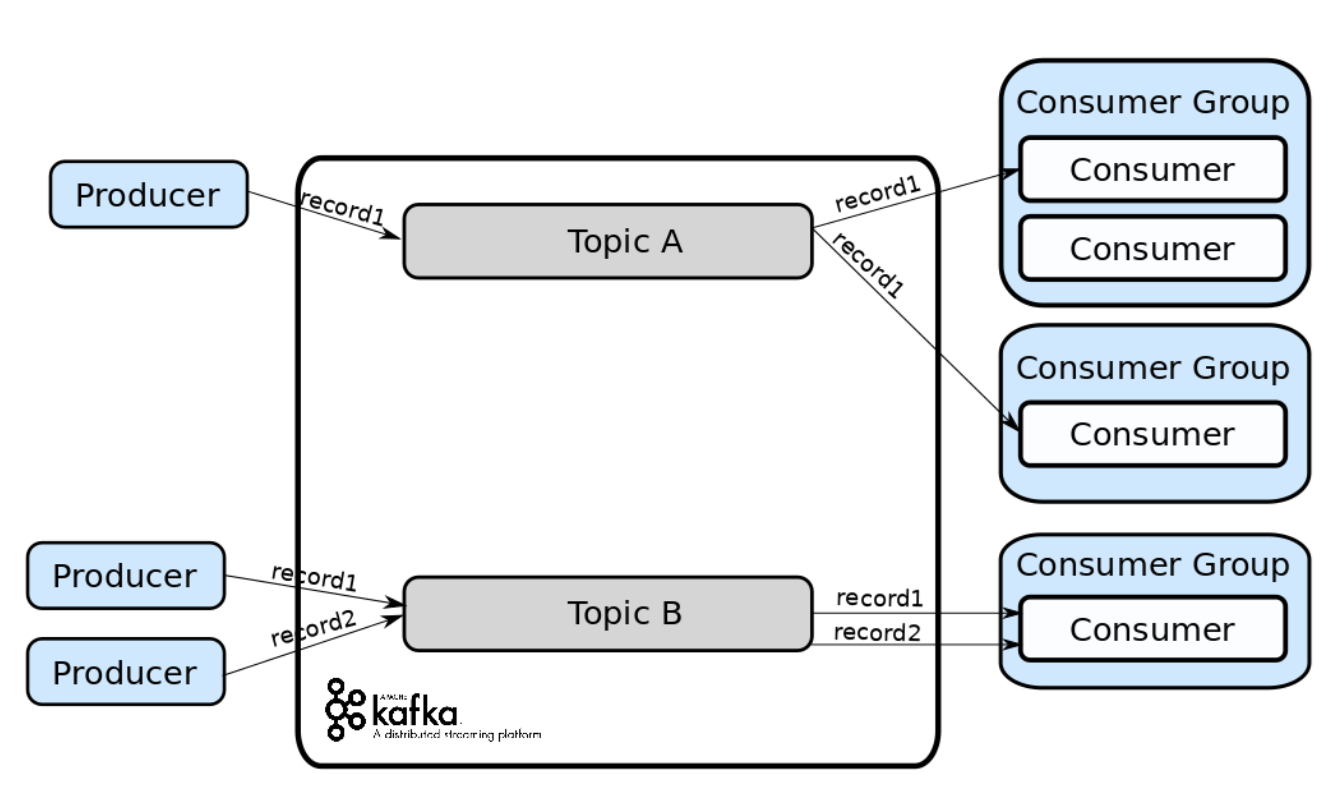
Một trong số Quorum Controller sẽ đóng vai trò leader tạo ra event trong metadata topic, các Quorum Controller còn lại sẽ đồng bộ từ leader bằng cách phản hồi các event này. Vì vậy khi một broker bị lỗi và cần tham gia lại có thể bắt kịp các sự kiện bị mất từ nhật ký event.

Không như với mô hình dựa trên Zookeeper, khi có sự thay đổi về Leader thì Quorum Controller mới đã có sẵn các dữ liệu metadata đã được đồng bộ trong bộ nhớ thay vì phải load lại trạng thái từ Zookeeper.

**Lợi ích khi sử dụng Kraft Mode**

* Tăng khả năng mở rộng của Kafka, dễ dàng mở rộng đến hàng triệu partition
* Dễ dàng cài đặt, bảo trì
* Tăng sự ổn định cho hệ thống, dễ dàng giám sát và quản lý
* Khởi động, Restart Kafka nhanh hơn
* Tăng khả năng phục hồi khi xảy ra sự cố

## Cơ chế hoạt động của Apache Kafka



### Mô hình cơ chế hoạt động của Apache Kafka

Các message được gửi tới Kafka theo Topic, các Topic giống như là các kênh lưu trữ message từ Producer gửi đến Kafka. Người tiêu dùng (Consumer) đăng ký một hoặc nhiều Topic để tiêu thụ những message đó.

Kafka có thời gian lưu giữ, vì vậy nó sẽ lưu trữ message theo thời gian hoặc kích thước đã được cấu hình và có thể được chỉ định gửi theo Topic.

Những người tiêu dùng (Consumer) sẽ được gán nhãn theo Consumer Group (nhóm người tiêu dùng). Trong một nhóm người tiêu dùng, một hoặc nhiều người tiêu dùng làm việc cùng nhau để xử lý một Topic. Khi có một message mới đến Topic, nó sẽ được gửi đến một Consumer trong một nhóm Consumer Group.

## Ưu nhược điểm Apache Kafka

**Ưu điểm:**

* Xử lý dữ liệu thời gian thực: Kafka được thiết kế để xử lý dữ liệu thời gian thực và streaming, cho phép đáp ứng nhanh chóng đối với sự kiện mới xảy ra. Điều này rất hữu ích cho việc giám sát, phân tích, và ứng phó tức thì với các tình huống thay đổi.
* Tính nhất quán và độ tin cậy cao: Kafka giữ cho bạn tính nhất quán dữ liệu thông qua quá trình sao lưu và sao chép thông điệp giữa các broker. Điều này đảm bảo dữ liệu không bị mất mát và luôn sẵn sàng cho các ứng dụng tiêu thụ.
* Mở rộng dễ dàng: Kafka có khả năng mở rộng dễ dàng bằng cách thêm các broker vào cluster. Điều này cho phép tăng khả năng xử lý dữ liệu mà không cần thay đổi cấu trúc toàn bộ hệ thống.
* Dữ liệu đa dạng: Kafka không chỉ hỗ trợ dữ liệu thông thường mà còn có khả năng xử lý dữ liệu đa dạng như logs, trạng thái ứng dụng, sự kiện liên quan đến giao dịch tài chính, và nhiều loại dữ liệu khác.
* Khả năng chia lượng lớn: Nhờ việc sử dụng phân vùng (partition), Kafka có khả năng xử lý lượng dữ liệu lớn một cách hiệu quả. Điều này làm cho nó phù hợp cho các tình huống cần xử lý hàng tỷ sự kiện mỗi ngày.
* Tích hợp linh hoạt: Kafka có khả năng tích hợp với nhiều công nghệ và ứng dụng khác. Có thể sử dụng các ngôn ngữ lập trình khác nhau để viết các ứng dụng producer và consumer, và Kafka cũng hỗ trợ các giao thức truyền tải khác nhau.
* Lưu trữ dữ liệu lâu dài: Ngoài việc xử lý dữ liệu thời gian thực, Kafka cũng có khả năng lưu trữ dữ liệu lâu dài. Điều này cho phép lưu trữ và truy xuất lại dữ liệu sự kiện trong tương lai để phân tích và kiểm tra.
* Hệ thống log: Kafka hoạt động như một hệ thống log, cho phép bạn lưu trữ và tìm kiếm thông tin sự kiện theo thời gian. Điều này rất hữu ích cho việc gỡ lỗi, phân tích dữ liệu và theo dõi hoạt động hệ thống.

**Nhược điểm:**

* Phức tạp trong cài đặt ban đầu: Cài đặt và cấu hình ban đầu của Kafka có thể phức tạp đối với những người mới bắt đầu sử dụng.
* Yêu cầu tài nguyên: Kafka yêu cầu một số lượng tài nguyên tương đối lớn để hoạt động tốt, bao gồm bộ nhớ và khả năng xử lý.
* Yêu cầu hiểu biết sâu về hệ thống: Để triển khai và quản lý Kafka hiệu quả, cần có kiến thức về hệ thống và mạng.

## Ứng dụng của Apache Kafka

Quản lý Sự kiện và Giao dịch: Các ngân hàng và tổ chức tài chính sử dụng Kafka để quản lý giao dịch và sự kiện liên quan đến tài khoản, chuyển tiền, giao dịch chứng khoán, giám sát gian lận, và theo dõi hoạt động tài chính.

Phân Tích Dữ liệu lớn và Data Pipeline: Kafka được sử dụng để xây dựng các data pipeline cho việc thu thập, xử lý và chuyển tiếp dữ liệu từ nhiều nguồn khác nhau đến các hệ thống phân tích dữ liệu lớn như Hadoop hoặc Spark.

Internet of Things (IoT): Thu thập và xử lý lượng lớn dữ liệu từ các thiết bị kết nối mạng như cảm biến, thiết bị điện tử thông minh, và máy móc để theo dõi và quản lý.

Ghi nhật ký (Logging): Kafka hoạt động như một hệ thống ghi nhật ký phân tán cho các ứng dụng, giúp ghi lại các sự kiện, lỗi và thông tin quan trọng trong ứng dụng.

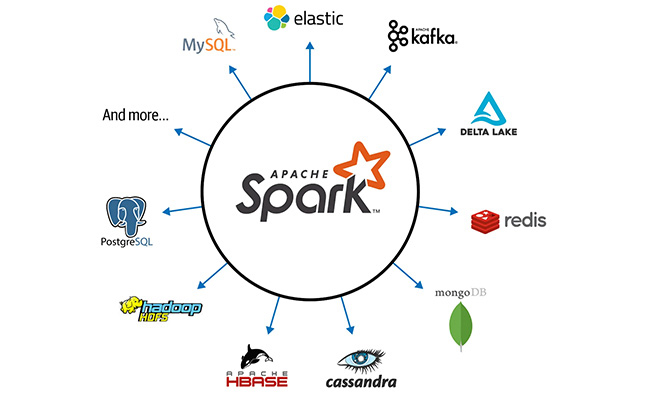
# TỔNG QUAN VỀ APACHE SPARK

## Khái niệm Apache Spark

Apache spark là một loại framework có chức năng xử lý dữ liệu mã nguồn mở trên quy mô lớn. Nó cho phép các lập trình viên xây dựng các mô hình dự đoán chính xác, nhanh chóng và trên diện rộng.

Với khả năng thực hiện trên nhiều máy tính, Apache spark tổng hợp toàn bộ các tệp dữ liệu. Tuy nhiên, chúng không cần phải trích xuất các mẫu tính toán thử nghiệm.

Framework này không sở hữu hệ thống file dữ liệu riêng mà sử dụng các hệ thống file khác nhau. Ví dụ như HDFS, Cassandra, S3,… Chúng cho phép người dùng xử lý dữ liệu theo thời gian thực.



### Thông tin về mã nguồn mở Apache Spark

## Thành phần Apache Spark

**Spark Core:**Là thành phần cơ bản của Spark, cung cấp các chức năng cơ bản như quản lý tác vụ, quản lý bộ nhớ, lập kế hoạch phân phối, và tương tác với hệ thống lưu trữ. Spark Core cũng bao gồm RDDs (Resilient Distributed Datasets), một cấu trúc dữ liệu chính trong Spark cho việc xử lý dữ liệu phân tán.

**Spark SQL:** Cho phép xử lý dữ liệu có cấu trúc bằng cách sử dụng ngôn ngữ SQL. Nó cung cấp một giao diện cho việc truy vấn dữ liệu từ các nguồn khác nhau như Hive, Avro, Parquet và các nguồn dữ liệu có cấu trúc khác.

**Spark Streaming:** Cho phép xử lý dữ liệu theo thời gian thực. Nó có thể xử lý và phân tích dữ liệu đầu vào liên tục từ các nguồn như Kafka, Flume và HDFS.

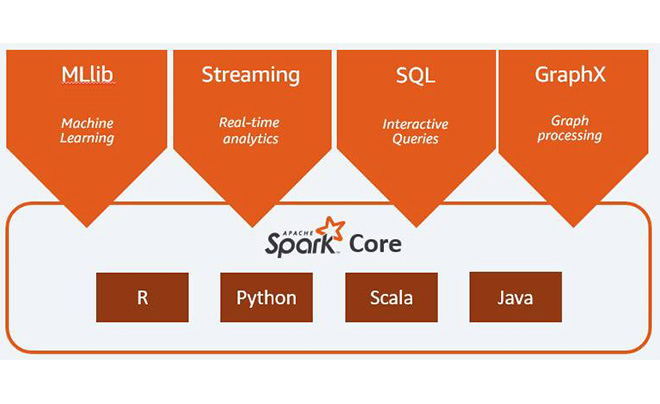
**MLlib (Machine Learning Library):** Thư viện machine learning tích hợp sẵn trong Spark để hỗ trợ các nhiệm vụ machine learning và data mining.

**GraphX:** Thư viện cho xử lý đồ thị và tính toán đồ thị phân tán. Nó cung cấp các API để xây dựng và tính toán đồ thị.

**SparkR:** Một giao diện R cho Spark, cho phép người dùng sử dụng R để thực hiện các phân tích dữ liệu trên Spark.

**Cluster Manager:** Apache Spark có thể chạy trên các cluster manager như Apache Mesos, Hadoop YARN hoặc Spark's native cluster manager để quản lý và phân phối các tác vụ xử lý dữ liệu.

**Spark Packages:** Là các thư viện và công cụ mở rộng được cộng đồng phát triển để mở rộng khả năng của Spark, bao gồm các gói cho xử lý dữ liệu, machine learning, đồ thị, và nhiều lĩnh vực khác.



### Các thành phần chính của Apache Spark

## Ưu nhược điểm của Apache Spark

**Ưu điểm:**

* *Khả năng tương thích rất tốt:* Apache spark là phần mềm mã nguồn mở miễn phí, bởi vậy nó tương thích với mọi hệ điều hành. Đồng thời phần mềm này tích hợp với mọi định dạng tệp và các nguồn dữ liệu do Hadoop hỗ trợ.
* *Khả năng xử lý dữ liệu nhanh, quy mô lớn:* Apache spark xử lý cùng lúc nhiều dữ liệu trên nhiều máy tính khác nhau dựa trên thời gian thực.
* *Hỗ trợ đa ngôn ngữ:* Người dùng có thể sử dụng rất nhiều ngôn ngữ khác nhau với Apache spark như: Java, Python, Scala, R.
* *Điều chỉnh độ trễ linh hoạt:* Phần mềm có thể tạo sẵn các sparkcontext giúp tăng, giảm độ trễ thực thi công việc (theo giây). Người dùng có thể thực hiện cùng lúc nhiều tính toán mà không tạo sự chênh lệch quá lớn.
* Phần mềm dễ dàng quản lý trên giao diện REST.

**Nhược điểm:**

* *Yêu cầu bộ nhớ lớn:* Apache Spark đòi hỏi rất nhiều RAM để chạy trong bộ nhớ, do đó chi phí của Spark khá cao.
* *Không thực sự real-time:* Spark Streaming không thực sự real-time.
* *Phụ thuộc vào các nền tảng khác:* Spark không có hệ thống file của riêng mình, do đó, nó phụ thuộc vào một số nền tảng khác như Hadoop hoặc một nền tảng dựa trên đám mây (S3, Google Cloud Storage,…).

## Ứng dụng của Apache Spark

Apache spark có khả năng phát hiện và cung cấp các dữ liệu theo thời gian thực. Bởi vậy, nó được ứng dụng rất phổ biến trong rất nhiều lĩnh vực khác nhau.

* *Trong lĩnh vực tài chính:* phần mềm được áp dụng để dự đoán tỷ lệ chuyển đổi khách hàng. Họ sẽ biết được ai tham gia hoặc hủy bỏ các sản phẩm mà họ cung cấp.
* *Trong lĩnh vực y tế:* công cụ này xây dựng dịch vụ chăm sóc bệnh nhân một cách toàn diện. Nó cho phép người dùng và tổ chức y tế tương tác và trao đổi thông tin trực tuyến.
* *Trong lĩnh vực sản xuất:* phần mềm giúp loại bỏ thời gian chết của thiết bị kết nối mạng internet.
* *Trong hoạt động thương mại:* Apache spark giúp thu hút và tìm kiếm khách hàng tiềm năng. Nó dựa trên các dịch vụ và ưu đãi cá nhân hóa.

# ỨNG DỤNG KAFKA KẾT HỢP SPARK

## Giới thiệu tổng quan về tập dữ liệu

Tập dữ liệu sử dụng là dữ liệu chứng khoán lịch sử của Công ty cổ phần Tập đoàn Hòa Phát (HPG). Tập dữ liệu được lấy từ ngày 16/11/2007 đến ngày 16/10/2023 và được lấy trực tiếp từ link: [investing.com](https://www.investing.com/equities/hoa-phat-group-jsc-historical-data). Tập dữ liệu gồm 3995 hàng và 7 trường bao gồm:

* *"Date" (Ngày):* Đây là trường dữ liệu chứa thông tin về ngày mà dữ liệu được thu thập hoặc ghi nhận.
* *"Price" (Giá):* Trường này chứa thông tin về giá đóng cửa của tài sản tài chính (ví dụ: cổ phiếu) tại ngày tương ứng.
* *"Open" (Giá mở cửa):* Đây là giá tài sản tài chính tại thời điểm mở cửa phiên giao dịch. Nó là giá đầu tiên của phiên giao dịch đó.
* *"High" (Giá cao nhất):* Trường này cho biết giá cao nhất mà tài sản tài chính đã đạt được trong suốt phiên giao dịch đó.
* *"Low" (Giá thấp nhất):* Đây là giá thấp nhất mà tài sản tài chính đã đạt được trong suốt phiên giao dịch đó.
* *"Vol." (Khối lượng giao dịch):* Trường này chứa thông tin về khối lượng giao dịch, tức là số lượng tài sản tài chính đã được giao dịch trong phiên giao dịch tương ứng.
* *"Change %" (Tỷ lệ thay đổi):* Trường này cho biết tỷ lệ thay đổi giữa giá đóng cửa tại ngày đó so với ngày giao dịch trước đó dưới dạng phần trăm.

## Giới thiệu tổng quan về ứng dụng

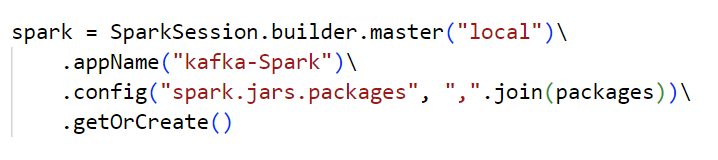
****

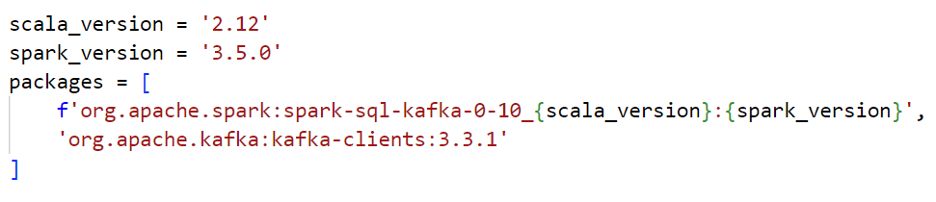
### Mô hình Kafka kết hợp Spark

Ở đây Kafka đóng vai trò là hệ thống hàng đợi dữ liệu (message queue) phục vụ chức năng thu thập dữ liệu đầu vào. Còn Spark giữ vai trò là hệ thống xử lý thực hiện phân tích các mẫu dữ liệu nhỏ trên luồng dữ liệu.

Các bước thực hiện:

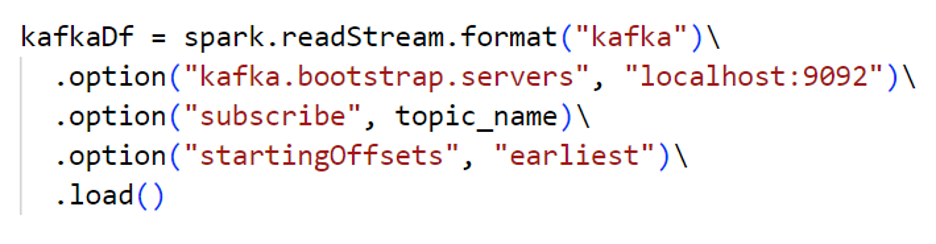
* ***Khởi tạo Spark Session:***





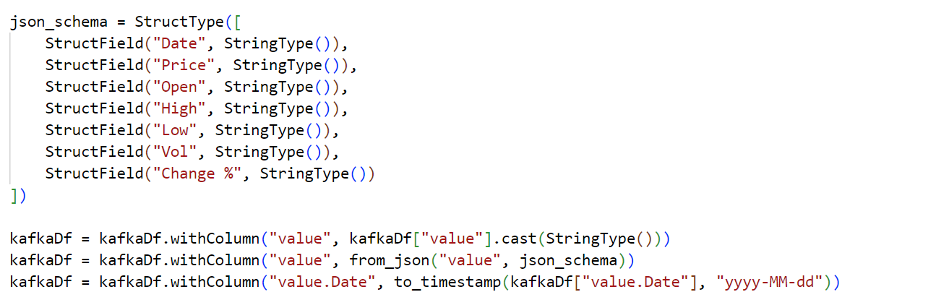
Trong phần này, khởi tạo một phiên Spark sử dụng SparkSession.builder. Điều này là cần thiết để làm việc với Apache Spark. Các cấu hình như master, appName, và config được sử dụng để xác định các thiết lập cho phiên Spark. Đặc biệt, spark.jars.packages được sử dụng để chỉ định các gói cần thiết, bao gồm gói Kafka để xử lý dữ liệu streaming từ Kafka.

* ***Đọc Dữ liệu từ Kafka:***



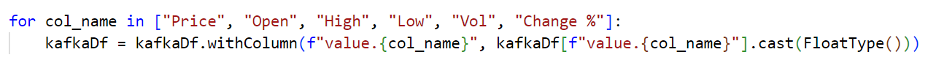
Trong phần này, sử dụng readStream để khai báo rằng dữ liệu đang được đọc trong chế độ streaming. Các tùy chọn như kafka.bootstrap.servers, subscribe, và startingOffsets đặt cấu hình kết nối đến Kafka và cách xử lý các offset khi bắt đầu streaming.

* ***Xử Lý Dữ Liệu JSON:***



Định nghĩa một schema để hiểu cấu trúc của dữ liệu JSON đầu vào từ Kafka. Sau đó, thông qua withColumn, nó chuyển đổi dữ liệu JSON từ cột "value" thành các cột riêng lẻ. Điều này giúp làm cho dữ liệu dễ dàng truy cập và sử dụng trong các phần sau.

* ***Chuyển Đổi Kiểu Dữ Liệu:***



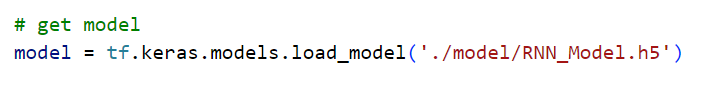
Phần này sử dụng một vòng lặp để chuyển đổi kiểu dữ liệu của các trường số liệu từ chuỗi sang số thực. Điều này là quan trọng để có thể thực hiện các phép toán số học và tạo biểu đồ dữ liệu một cách chính xác.

* ***Truy vấn dữ liệu:***

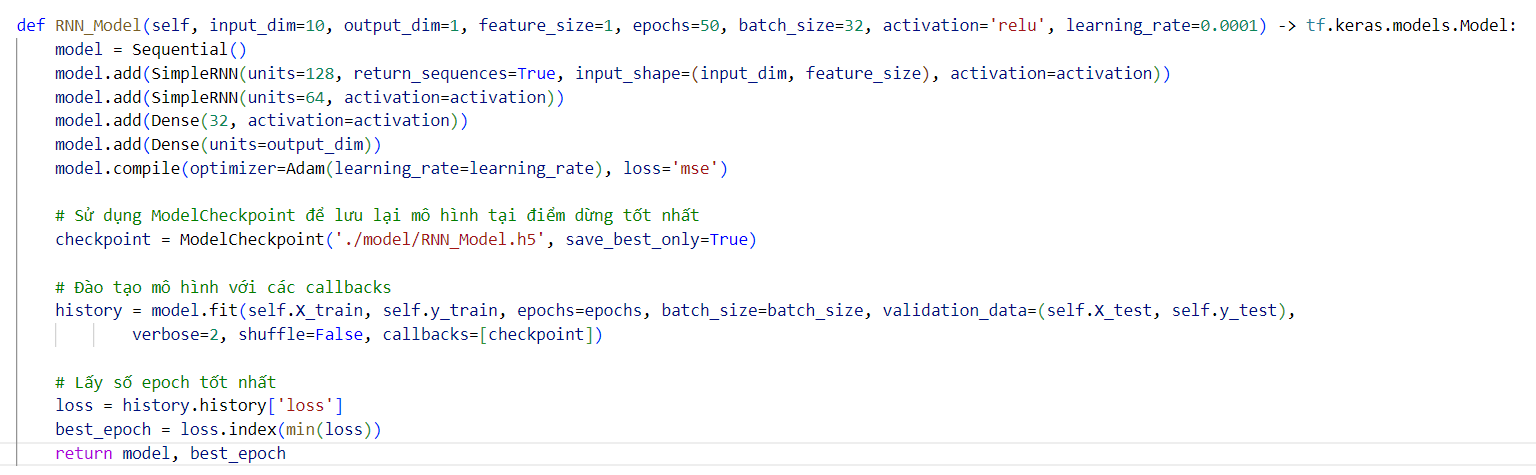


Chọn các trường cần thiết từ DataFrame và bắt đầu truy vấn luồng để xử lý dữ liệu. Điều này được thực hiện bằng cách sử dụng phương thức select và writeStream.

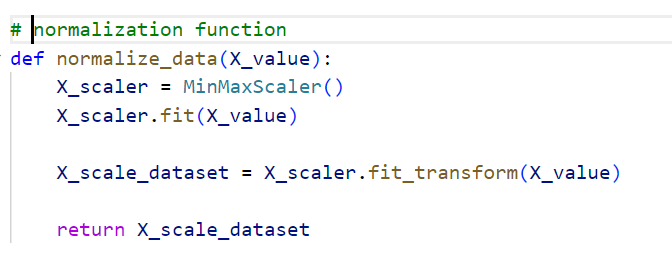
* ***Tạo mô hình RNN***



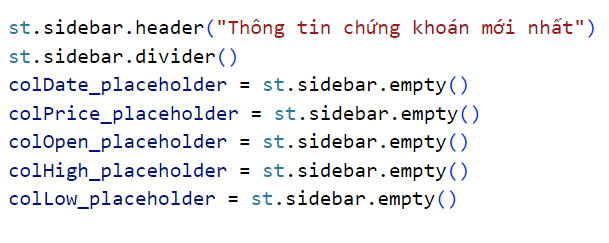
Phần này sử dụng mô hình RNN đã được huấn luyện và lưu trong thư mục model, mô hình RNN được tạo bằng cấu trúc sau:



* ***Viết hàm chuẩn hóa dữ liệu, dữ liệu được chuẩn hóa bằng cách sử dụng MinMaxScaler.***

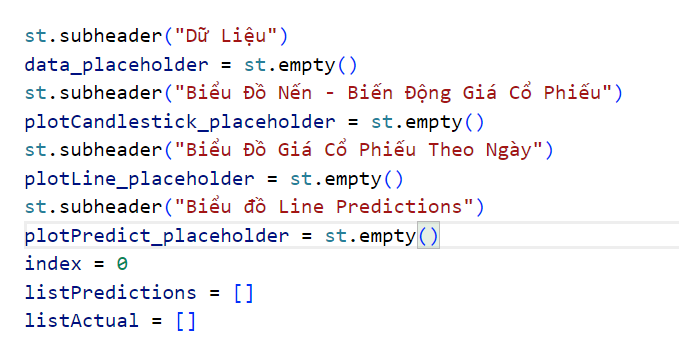


* ***Hiển thị Thông Tin Chứng Khoán Mới Nhất trong Sidebar***

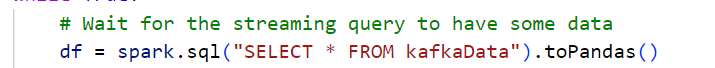


Phần này dùng Sidebar để hiển thị thông tin chứng khoán mới nhất, bao gồm các mục như ngày, giá mở, giá cao, giá thấp, v.v.

* ***Chuẩn bị Placeholder cho Hiển thị Dữ Liệu và Biểu Đồ***

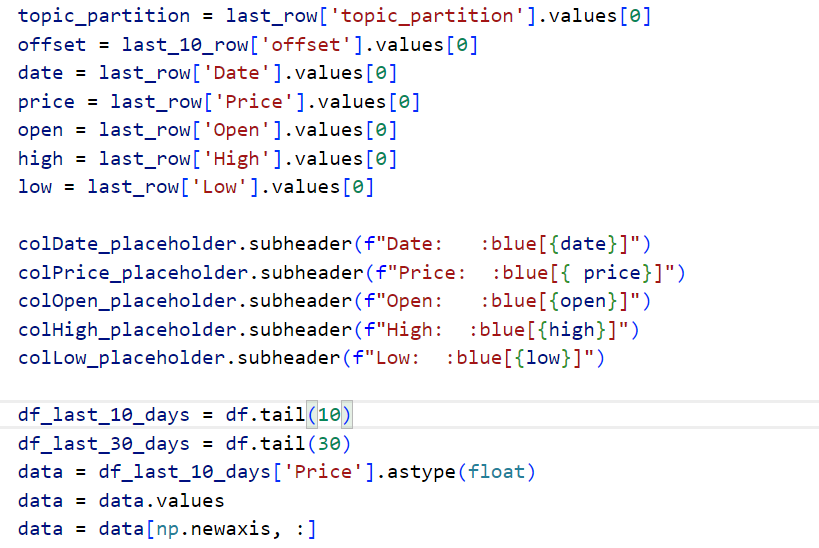


* ***Thực hiện lấy dữ liệu từ kafka***

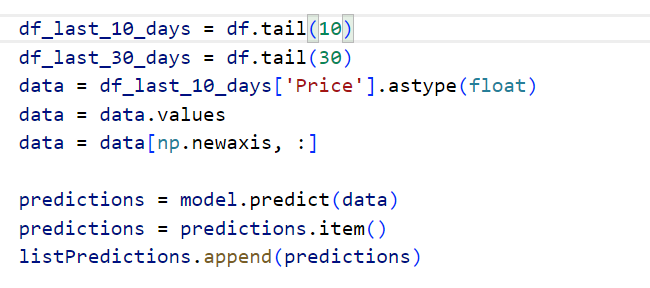


Phần này truy vấn dữ liệu từ DataFrame "kafkaData" và chuyển nó thành DataFrame của thư viện pandas.

* ***Lấy thông tin giá chứng khoán từ DataFrame và hiển thị lên giao diện***

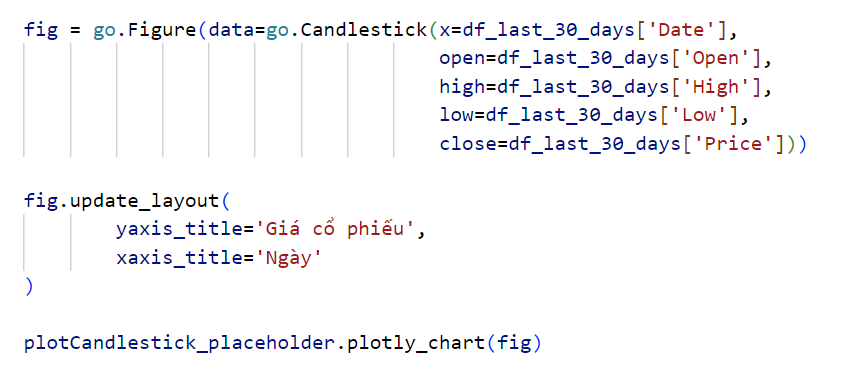


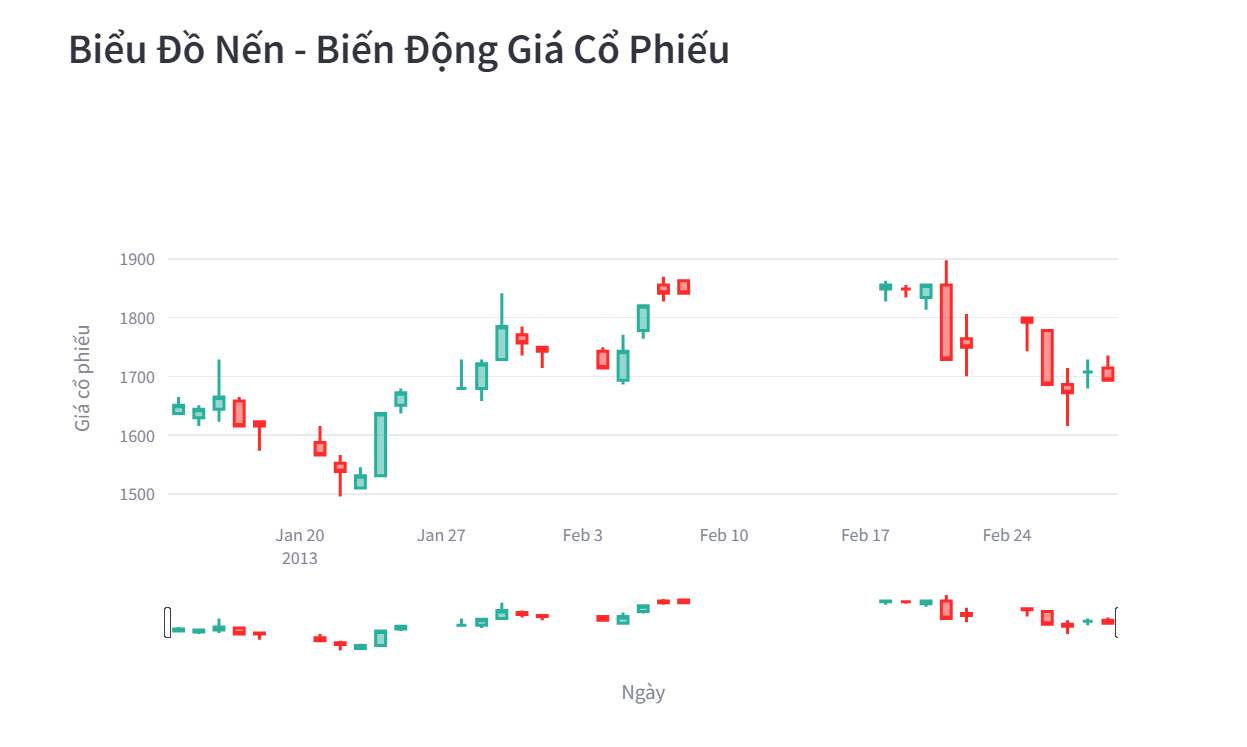
* ***Chuẩn bị dữ liệu để dự đoán và sau đó tiến hành dự đoán giá chứng khoán***



Phần này lấy dữ liệu giá cổ phiếu của 10 dòng cuối cùng và chuẩn bị nó để đưa vào mô hình dự đoán, sau đó dự đoán giá cổ phiếu sử dụng mô hình đã được tải trước đó và thêm kết quả dự đoán vào danh sách dự đoán listPrediction để thực hiện vẽ biểu đồ.

* ***Vẽ biểu đồ nến***

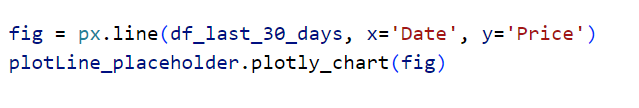




Biểu đồ Candlestick là một công cụ phân tích kỹ thuật phổ biến trong thị trường chứng khoán. Biểu đồ Candlestick giúp nhà đầu tư nhận biết các xu hướng thị trường và dự đoán các xu hướng tương lai dựa trên các mô hình nến cụ thể. Đây là một công cụ hữu ích để hỗ trợ việc ra quyết định giao dịch.

Mỗi “nến” trên biểu đồ biểu diễn bốn yếu tố quan trọng: giá mở cửa, giá đóng cửa, giá cao nhất và giá thấp nhất trong một ngày giao dịch cụ thể. Điều này giúp nhà đầu tư nắm bắt nhanh chóng và dễ dàng sự biến động của giá cổ phiếu trong thời gian.

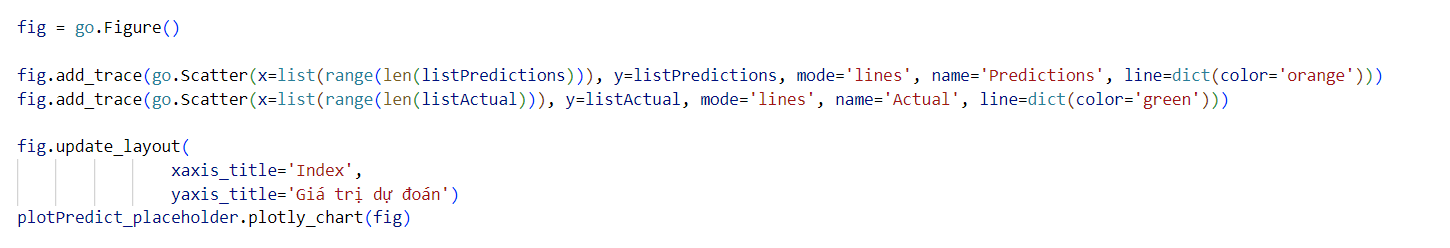
* ***Vẽ biểu đồ đường thể hiện giá trị thực tế của giá đóng cửa theo ngày***

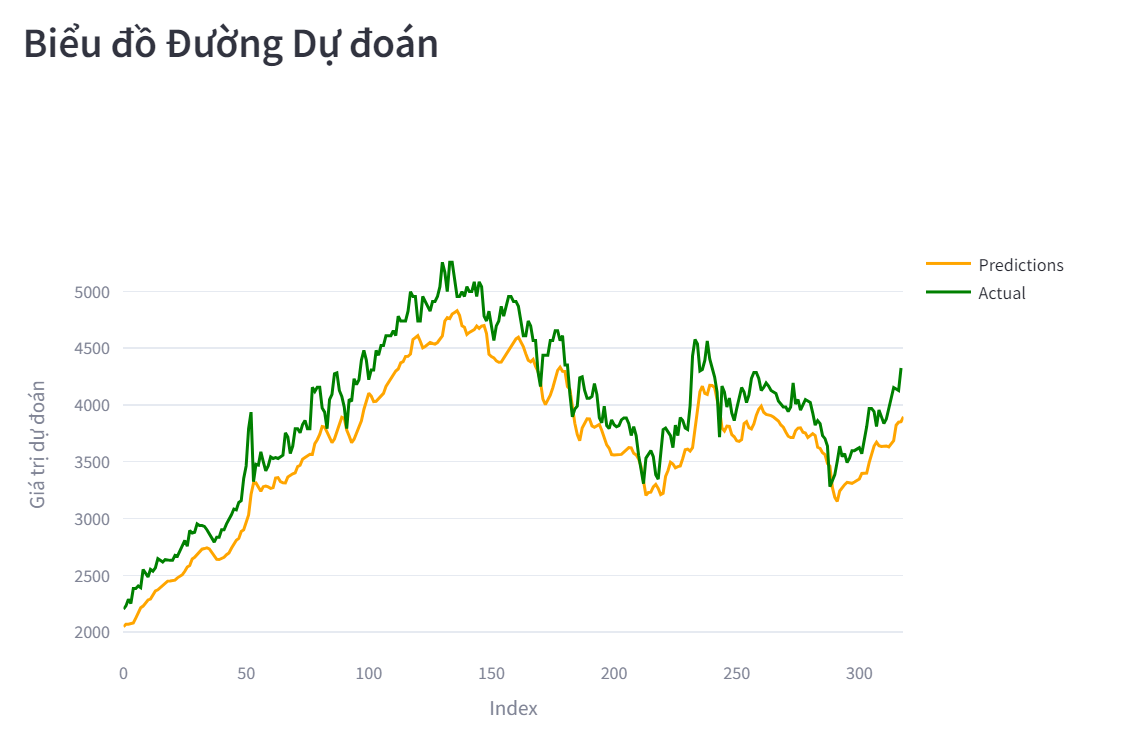




Biểu đồ đường thực tế: Biểu đồ đường trên được tạo ra để hiển thị giá cổ phiếu theo ngày trong khoảng thời gian 30 ngày gần đây, giúp người xem dễ dàng nhận thức xu hướng và biến động gần đây của thị trường chứng khoán. Trục X (hoành độ) của biểu đồ là thời gian (Date), và trục Y (tung độ) là giá cổ phiếu (Price).

* ***Vẽ biểu đồ đường so sánh giá trị thực tế và giá trị dự đoán***



****

Biểu đồ so sánh: Biểu đồ này giúp hiển thị và so sánh giữa các giá trị dự đoán và giá trị thực tế của mô hình. Mỗi đường màu tương ứng với một loạt các giá trị liên tục. Việc so sánh giữa giá trị dự đoán và giá trị thực tế giúp đánh giá hiệu suất của mô hình.

**TÀI LIỆU THAM KHẢO**

1. Vinh Pham. (2023, August 8 15). Apache Kafka là gì? Ưu nhược điểm của Kafka. Viblo. <https://viblo.asia/p/apache-kafka-la-gi-uu-nhuoc-diem-cua-kafka-r1QLxxWgLAw>
2. Bizfly Cloud. (2022, August 10). Kafka là gì? Tìm hiểu khái niệm cơ bản về Apache Kafka. <https://bizflycloud.vn/tin-tuc/kafka-la-gi-gioi-thieu-tong-quan-ve-kafka-20220810160652162.htm>
3. Tram Ho. (2023, May 4). Kafka không cần Zookeeper? Kafka Kraft mode [Bài viết blog]. ITZone. <https://itzone.com.vn/vi/article/kafka-khong-can-zookeeper-kafka-kraft-mode/>
4. Hung Tran. (December 9, 2020). Tìm hiểu cơ chế hoạt động của Apache Kafka - Phần 1. Cafeincode. <https://cafeincode.com/tim-hieu-co-che-hoat-dong-cua-apache-kafka-phan-1/>
5. Phuc Ngoc Nghia. (2020, March 22). Tìm hiểu về Apache Spark. Viblo. <https://viblo.asia/p/tim-hieu-ve-apache-spark-ByEZkQQW5Q0>
6. Vega Fintech, T. (2023, November 14). Apache Spark là gì? Vai trò và ứng dụng của Apache Spark. Vega Fintech. <https://vegafintech.vn/vi/apache-spark-la-gi-vai-tro-va-ung-dung-cua-apache-spark>