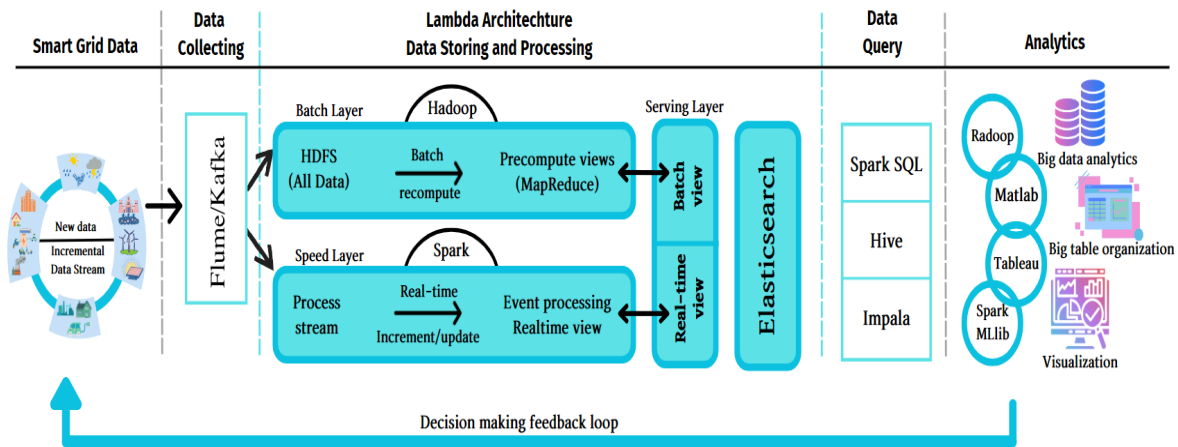


Final Project

1. Kiến trúc

Xây dựng kiến trúc lưu trữ và xử lý dữ liệu lớn gồm đầy đủ các thành phần như hình bên dưới bao gồm: data crawling, lưu trữ, phân tích, và trực quan hoá.



System architecture

- Data Collecting: Kafka, ..
- Database: GG bucket, Elasticsearch, MongoDB, ...
- Data query and process: Spark
- Analytics: Kibana, ...

2. Yêu cầu

Về mặt chức năng:

- Xây dựng và cài đặt được kiến trúc lambda như trong phần 1
- Thu thập được nguồn dữ liệu thoả mãn đặc điểm của bigdata (tối thiểu 3V volume-velocity-variety)
- Có dashboard trực quan hoá được dữ liệu (data visulization) với các biểu đồ giúp ích cho việc hiểu được tổng quan về dữ liệu
- Phân tích được dữ liệu: tối thiểu đưa ra được các tính toán về mặt thống kê, khuyến khích áp dụng các mô hình học máy/AI. Từ đó khai phá được các kiến thức/thông tin ẩn dưới dữ liệu, giúp giải quyết các bài toán thực tế đề ra.

Về mặt bài toán: tham khảo một số bài toán về phân tích dữ liệu như sau:

- 2.1 Phân tích xu hướng
- 2.2 Phát hiện các sự kiện bất thường/tiêu biểu
- 2.3 Dự báo biến động giá cả
- 2.4 Phát hiện giả mạo, lừa đảo

- 2.5 Đánh giá KOL trên mạng xã hội
- 2.6 Phân tích cảm xúc
- 2.7 Đánh giá/dự báo tiềm năng các sản phẩm

3. Nguồn dữ liệu

No	Loại dữ liệu	Ghi chú
1.1	Các tin tức tài chính, chứng khoán, tiền tệ, Fintech (startup, ventures, investment), tin tức của các doanh nghiệp trên sàn chứng khoán: chính thức + upcom	
1.2	Dữ liệu văn bản pháp quy và hướng dẫn về Fintech, hỏi đáp về luật liên quan tới tài chính, Fintech.	
1.3	Dữ liệu chứng khoán, dữ liệu tỷ giá FX, giá vàng, dầu, và các loại hàng hoá có giao dịch điện tử khác	
1.4	Hỏi đáp về các kiến thức liên quan tới công nghệ áp dụng trong tài chính (Blockchain, AI, RPA..)	
2.1	Dữ liệu hỏi đáp liên quan tới pháp luật nói chung, kết hợp với các văn bản pháp luật	
2.2	Các tương tác bình luận, trao đổi trên mạng xã hội	
3.1	Dữ liệu review về các sản phẩm trên các sàn thương mại điện tử	
3.2	Dữ liệu review về các dịch vụ như khách sạn, nhà hàng, quán cafe,	
3.3	Dữ liệu giá cả hàng hoá, doanh thu bán hàng	
3.4	Dữ liệu bất động sản	