

Pós graduação em Computação Aplicada – Universidade Presbiteriana
Mackenzie (MACKENZIE)
Rua da Consolação, 896 Prédio 45 "T" (2º subsolo) – Caixa Postal 01302-907
– São Paulo – SP – Brasil

Arquitetura para Coleta e Análise de Notícias Econômicas

Daniel M. Correia¹, Fabricio R. Zillig², Messias O. da Silva³

2 de outubro de 2024

Resumo

Este trabalho apresenta uma arquitetura para coleta e análise de notícias econômicas em tempo real utilizando Kafka, MongoDB, uma API para integração dos serviços, e a apresentação de dados com Kibana. A arquitetura foi desenvolvida para facilitar a ingestão e análise de grandes volumes de dados provenientes de diversas fontes de notícias, permitindo monitoramento e análise em tempo real de tendências econômicas.

Este trabalho propõe uma arquitetura para coleta e análise de notícias econômicas utilizando tecnologias como Apache Kafka, MongoDB, e Kibana. A proposta visa processar grandes volumes de dados em tempo real e apresentar as informações de forma visualmente acessível.

1 Introdução

Nos dias atuais, o volume de dados gerados a partir de notícias e publicações na internet cresce exponencialmente. O setor econômico, em particular, está repleto de informações que influenciam diretamente as decisões de mercado. Para acompanhar as mudanças e analisar as tendências econômicas de forma eficiente, propomos uma arquitetura que coleta, armazena e analisa notícias econômicas em tempo real.

A arquitetura proposta utiliza Apache Kafka para ingestão de dados, MongoDB como base de dados NoSQL para armazenamento, uma API que integra o fluxo de dados entre os diferentes componentes, e o Kibana para visualização e análise dos dados processados.

2 Arquitetura Proposta

2.1 Ingestão de Dados com Apache Kafka

O Apache Kafka é uma plataforma de streaming distribuída projetada para lidar com dados em tempo real de maneira eficiente. Na nossa arquitetura, o Kafka é responsável por consumir notícias de diversas fontes, como feeds RSS, APIs de portais de notícias e redes sociais.

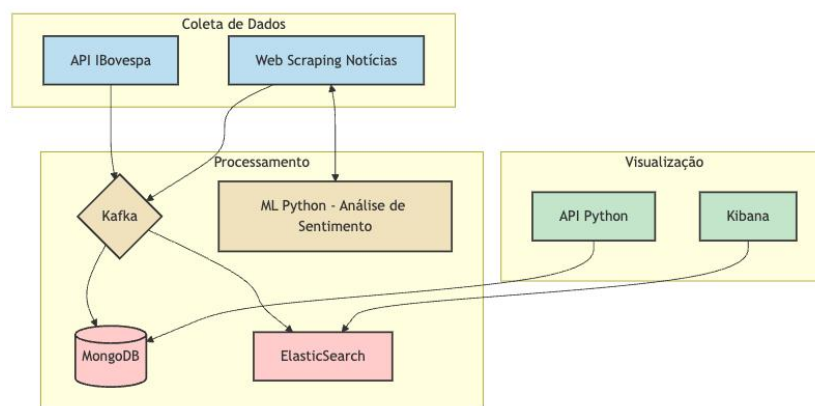


Figura 1: Arquitetura proposta para coleta e análise de notícias econômicas

Os dados são organizados em tópicos de acordo com o tipo de informação (notícias econômicas, indicadores financeiros, etc.). Cada produtor de dados (feed de notícias, por exemplo) envia os dados para o Kafka, que os distribui para os consumidores, que no caso são os outros componentes do sistema.

2.2 Armazenamento de Dados com MongoDB

O MongoDB foi escolhido como o sistema de armazenamento por ser um banco de dados NoSQL altamente escalável e eficiente para lidar com grandes volumes de dados não estruturados, como notícias. As notícias captu-

radas pelo Kafka são armazenadas no MongoDB para consultas e análises posteriores. A flexibilidade do modelo de documentos do MongoDB permite armazenar metadados das notícias (título, fonte, data, etc.) de forma eficiente.

2.3 Integração dos Componentes com API

A comunicação entre os diferentes componentes da arquitetura é realizada através de uma API RESTful. Essa API é responsável por expor os dados armazenados no MongoDB para aplicações de análise e visualização, bem como por gerenciar as requisições de novos dados ao Kafka. A API atua como um ponto central de integração, facilitando a expansão da arquitetura com novos serviços ou fontes de dados.

2.4 Visualização com Kibana

O Kibana, uma ferramenta de análise e visualização de dados, é utilizado para apresentar as informações coletadas de forma interativa e visualmente acessível. Através do Kibana, é possível criar dashboards que mostram tendências econômicas baseadas nas notícias, permitindo uma análise detalhada dos dados com gráficos de linha, barras e mapas de calor, por exemplo.



Figura 2: Exemplo de dashboard no Kibana mostrando tendências econômicas

3 Conclusão

Neste artigo, foi proposta uma arquitetura escalável para a coleta e análise de notícias econômicas em tempo real. A utilização de tecnologias modernas como Apache Kafka, MongoDB, e Kibana permite processar grandes volumes de dados de forma eficiente, garantindo que as informações sejam apresentadas em tempo hábil para análise de tendências econômicas. Trabalhos futuros podem incluir a integração com algoritmos de aprendizado de máquina para previsão de eventos econômicos.

4 Referências

Referências

- [1] Apache Kafka: A Distributed Streaming Platform. Disponível em: <https://kafka.apache.org>. Acesso em: 30 set. 2024.
- [2] MongoDB: The Database for Modern Applications. Disponível em: <https://www.mongodb.com>. Acesso em: 30 set. 2024.
- [3] Kibana: Explore, Visualize, and Analyze Data. Disponível em: <https://www.elastic.co/kibana>. Acesso em: 30 set. 2024.
- [4] Fielding, R. T. (2000). Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine.