

Apresentação da Proposta de Problema e Solução

Proposta: Análise e Visualização de Informações do Mercado de Bitcoin utilizando um Pipeline de Dados Diversificado

Universidade Presbiteriana Mackenzie
Mestrado Profissional em Computação Aplicada
Disciplina: Coleta, Armazenamento e Visualização de Dados
Aluno: Diogo Luís Campanhã
RA: 10403308
Data: Abril/ 2025

1. Introdução

O mercado de criptomoedas é conhecido por sua alta volatilidade e complexidade, sendo influenciado por uma vasta gama de fatores, incluindo notícias, sentimentos do mercado e dados de preços em tempo real. A capacidade de coletar, armazenar e analisar eficientemente essa diversidade de informações é crucial para obter insights valiosos e potencialmente auxiliar na tomada de decisões. Este projeto propõe a construção de um pipeline de dados robusto para integrar dados de diferentes fontes relacionadas ao mercado de criptomoedas, visando sua análise exploratória e visualização.

2. Problema Observado

A dispersão de informações relevantes sobre criptomoedas em múltiplas plataformas e formatos dificulta a obtenção de uma visão holística do mercado. Analisar o impacto de notícias, o sentimento geral dos investidores e a dinâmica de preços de forma integrada representa um desafio significativo. A falta de uma infraestrutura unificada para coletar e processar esses dados impede a identificação rápida de padrões e potenciais oportunidades ou riscos.

3. Solução Proposta

Propõe-se a implementação de um pipeline de dados que automatize a coleta, o armazenamento e a visualização de informações cruciais do mercado de criptomoedas. A arquitetura do pipeline consiste nos seguintes componentes principais:

- **Coleta de Dados:** Utilização de diversas APIs para obter dados em tempo real e de forma automatizada. As APIs selecionadas incluem:
 - **CoinMarketCap:** Fornece dados de cotação atual de diversas criptomoedas, como o Bitcoin.
 - **API de Índice de Medo e Ganância:** Oferece uma métrica agregada do sentimento do mercado de criptomoedas.
 - **Crypto Panic:** Agrega notícias relevantes do universo cripto de múltiplas fontes.
- **Ingestão de Dados:** Os dados coletados das APIs serão ingeridos em um tópico Kafka. O Kafka atuará como uma camada de desacoplamento e fornecerá escalabilidade e confiabilidade para o fluxo de dados.

- **Armazenamento de Dados:** Os dados brutos do Kafka serão persistidos em um banco de dados MongoDB. O MongoDB, um banco de dados NoSQL orientado a documentos, oferece a flexibilidade necessária para armazenar dados com diferentes estruturas provenientes de diversas APIs. Adicionalmente, dados processados e transformados poderão ser armazenados em um banco de dados Cassandra, uma NoSQL column-family store, adequada para dados com alta taxa de gravação e leitura, como séries temporais de indicadores.
- **Processamento e Transformação:** Utilização da linguagem Python para consumir os dados do MongoDB, realizar análises exploratórias, calcular métricas derivadas e preparar os dados para visualização e alertas.
- **Visualização de Dados:** A ferramenta Grafana será utilizada para criar painéis interativos que apresentarão as informações coletadas e processadas de forma clara e concisa, permitindo o acompanhamento em tempo real das cotações, do sentimento do mercado e das notícias relevantes.
- **Alertas:** Utilizaremos python para configurar alertas de e-mail com base em condições predefinidas, como mudanças no índice de medo e ganância.

4. Justificativa da Aplicação

Este projeto se alinha com os objetivos da disciplina de Coleta, Armazenamento e Visualização de Dados, abordando todas as etapas de um pipeline de dados, desde a coleta e ingestão até o armazenamento, processamento, análise exploratória e visualização. A utilização de diferentes tecnologias, como APIs, Kafka, MongoDB, Python, Cassandra e Grafana, demonstra a capacidade de integrar diversas ferramentas para construir uma solução analítica robusta. A análise do mercado de criptomoedas, um domínio com grande volume e variedade de dados, serve como um caso de uso prático para aplicar os conceitos aprendidos na disciplina.

5. Resultados Esperados

Espera-se que este projeto resulte em um pipeline de dados funcional capaz de:

- Coletar dados em tempo real de múltiplas fontes de informação sobre criptomoedas.

- Armazenar eficientemente dados estruturados e não estruturados em diferentes sistemas de gerenciamento de bancos de dados (MongoDB e Cassandra).
- Processar e transformar os dados coletados para gerar insights relevantes.
- Visualizar as informações de forma clara e interativa através de painéis no Grafana.
- Gerar alertas automatizados por e-mail com base em eventos predefinidos.

A implementação deste pipeline proporcionará uma compreensão mais aprofundada da dinâmica do mercado de criptomoedas e demonstrará a aplicação prática dos conceitos e ferramentas estudados na disciplina