Sr Engineer, Data Work Trial Project

Overview

The work trial is a way to:

1. Gauge your interest in the position
2. Assess your technical abilities related to the role
3. Showcase your skills and abilities to us
4. See how quickly you can get up to speed on the cryptocurrency and blockchain space

Job description: (link)

Project Requirements

Description

At Core Scientific, we are a vertically integrated cryptocurrency/blockchain company spanning from the operation of mining data centers to deployment of funds into various financial instruments. This means we require the collection and aggregation of domain-specific data across our operational surface area to inform business decisions. An important responsibility of this role is to build and maintain data pipelines and various lakes/warehouses which requires learning about a new data source and integrating it into our existing systems.

Your task is to research, design, and implement a proof-of-concept data pipeline to collect and analyze cryptocurrency price data. The below requirements are left intentionally broad, so feel free to use tooling, products, and frameworks you are most familiar with to accomplish the task. Your selection and communication around particular design decisions are just as important as the implementation itself.

Cryptocurrencies trade on a variety of exchanges over an order book data structure expressing live bid and ask orders which can be filled. As this data is constantly updating independently per exchange, it is useful to sample price data across multiple sources (exchanges) to calculate aggregate prices and compare liquidity of respective markets.

The main components of the deliverable:

1. Create a data pipeline/ETL to ingest and persist order book data across two (2) different exchanges for both BTC/USD and ETH/USD markets:
    a. Every 60 seconds poll each exchange API for the current order book and persist raw order book data. Hint: here is the API request to pull Coinbase Pro's order book: https://api.pro.coinbase.com/products/ETH-USD/book?level=2 .
    b. For each poll, extract $100k of bid and ask order book data for each exchange.
        i. The goal here is for each poll we should be able to reason about the price we could achieve when market buying or selling $100k of BTC or ETH.
2. The Business Intelligence team would like to consume the following data for their models in real time at 60 second granularity:

    a. What is the average mid price per market?
    b. Which exchange would we prefer to execute a $50k buy or sell order on? At what price?
3. Your presentation should cover your architectural decisions, implementation details, and solution to handling the business analytics requested in (2).

Resources

- https://www.investopedia.com/terms/o/order-book.asp

Deliverables

- A GitHub repository/repositories containing all code developed
- A presentation detailing your design and implementation
- Any supporting documentation/diagrams for your work product