Real-Time Market Data Analytics System

# Introduction

This document provides a comprehensive overview of the Real-Time Market Data Analytics System, a robust solution for processing and analyzing financial market data in real-time. The system utilizes Apache Spark for stream processing, Kafka for message queuing, and PostgreSQL for data storage.

# System Architecture

The system consists of several key components working together to provide real-time market data analytics:

**Data Producers:**

* Finnhub API integration for real-time market data
* Binance API integration for cryptocurrency data
* Configurable data collection intervals
* Automatic reconnection and error handling

**Data Processing:**

* Apache Spark for stream processing
* Real-time analytics calculation
* Multiple time window aggregations
* Error handling and data validation

**Data Storage:**

* PostgreSQL database for analytics storage
* Efficient data indexing
* Optimized query performance

**API Layer:**

* FastAPI-based REST API
* Real-time analytics endpoints
* Comprehensive error handling
* OpenAPI documentation

# API Endpoints

**Endpoint: /api/v1/spark-analytics/symbols**

Get list of available trading symbols

**Endpoint: /api/v1/spark-analytics/min-max/{symbol}**

Get min-max analytics for a specific symbol

**Endpoint: /api/v1/spark-analytics/rolling-metrics/{symbol}**

Get rolling window metrics for a symbol

**Endpoint: /api/v1/spark-analytics/summary**

Get analytics summary for all symbols

# Analytics Features

**Price Analytics:**

* Minimum price
* Maximum price
* Average price
* Price volatility

**Volume Analytics:**

* Total volume
* Average volume
* Volume trends

**Time-based Analytics:**

* 1-minute window metrics
* 5-minute window metrics
* Custom time window support

# Technical Implementation

The system is implemented using modern technologies and best practices:

**Python 3.13:**

Core programming language

**Apache Spark:**

Stream processing engine

**Apache Kafka:**

Message queuing system

**PostgreSQL:**

Analytics data storage

**FastAPI:**

REST API framework

**Docker:**

Containerization

**SQLAlchemy:**

Database ORM

**Pydantic:**

Data validation