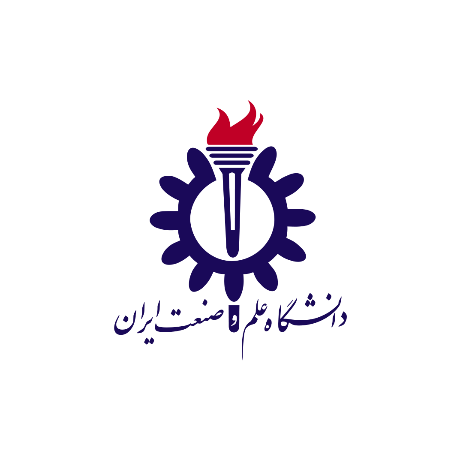


**سیستم تحلیل مالی و معاملات بلادرنگ با هوش مصنوعی**

پروژه نهایی درس تحلیل‌ها و سیستم‌های داده حجیم



بهمن ۱۴۰۳

****

**استاد درس: دکتر حسن نادری**

**دانشجویان: علی احمدی، امیر کریمی، مهدی شکاری و اهورا امینی**

فهرست مطالب

[**مقدمه** 2](#_Toc189755340)

[هدف پروژه 2](#_Toc189755341)

[اهمیت پروژه 2](#_Toc189755342)

[چالش‌ها 2](#_Toc189755343)

[**معماری سیستم** 3](#_Toc189755344)

[API Nobitex 4](#_Toc189755345)

[Data Ingestion Service 4](#_Toc189755346)

[Apache Kafka 4](#_Toc189755347)

[Stream Data Processing Service (Apache Spark) 4](#_Toc189755348)

[محاسبه شاخص‌ها 5](#_Toc189755349)

[تولید سیگنال‌ها 5](#_Toc189755350)

[پردازش میکروبچ‌ها 6](#_Toc189755351)

[خواندن و نوشتن با Spark 6](#_Toc189755352)

[مدیریت استریم 6](#_Toc189755353)

[سرویس میانی Kafka-to-QuestDB 6](#_Toc189755354)

[پردازش پیام‌ها 6](#_Toc189755355)

[جایگزینی nan و تبدیل مقادیر 6](#_Toc189755356)

[ساخت قالب ILP و ارسال داده‌ها به QuestDB 7](#_Toc189755357)

[QuestDB 7](#_Toc189755358)

[Aggregation Service 8](#_Toc189755359)

[Aggregate API 8](#_Toc189755360)

[Summarize API 9](#_Toc189755361)

[Summarize Multiple API 9](#_Toc189755362)

[Visualization Service (Grafana) 10](#_Toc189755363)

# **مقدمه**

## هدف پروژه

پروژه تحلیل داده‌های مالی در بازارهای مالی مانند رمزارز و نمادهای سهام مختلف به منظور تجزیه و تحلیل داده‌ها در زمان واقعی (Real-time) طراحی و پیاده‌سازی شده است. این سیستم با استفاده از تکنولوژی‌های توزیع شده و پردازش بلادرنگ داده‌ها، امکان تحلیل سریع و تصمیم‌گیری به موقع برای معامله گران بازارهای مالی را فراهم می‌کند.

بازارهای مالی به دلیل نوسانات شدید و لحظه‌ای قیمت‌ها، نیاز به ابزارهایی دارند که بتوانند به سرعت و به صورت دقیق، اطلاعات را پردازش کرده و سیگنال‌های مناسب برای خرید و فروش ارائه دهند. در این پروژه، هدف اصلی این است که با استفاده از پردازش‌های پیچیده و بلادرنگ، سیگنال‌های دقیق و به موقع برای معامله‌گران فراهم شود تا بتوانند تصمیمات بهتری اتخاذ نمایند.

## اهمیت پروژه

تحلیل داده‌های مالی از اهمیت بسیار بالایی برخوردار است. با افزایش حجم داده‌ها و نیاز به پردازش بلادرنگ، استفاده از سیستم‌های مقیاس‌پذیر و توزیع شده مانند Apache Kafka، Apache Spark و QuestDB به ویژه در تحلیل‌های مالی اهمیت پیدا می‌کند. این پروژه نه تنها بر مقیاس‌پذیری تمرکز دارد، بلکه توانایی پردازش داده‌ها در لحظه را نیز فراهم می‌آورد.

به طور ساده، این پروژه یک سیستم بلادرنگ است که به صورت خودکار، داده‌ها را از بازارهای مالی دریافت کرده، پردازش کرده و سپس سیگنال‌های مناسب خرید و فروش و همچنین با استفاده از هوش مصنوعی پیشبینی روند را نیز به کاربران نمایش می‌دهد.

این توانایی می‌تواند برای کسانی که در بازارهای مالی فعالیت می‌کنند، به ویژه برای کسانی که به دنبال تصمیم‌گیری‌های سریع و بهینه هستند، بسیار ارزشمند باشد.

## چالش‌ها

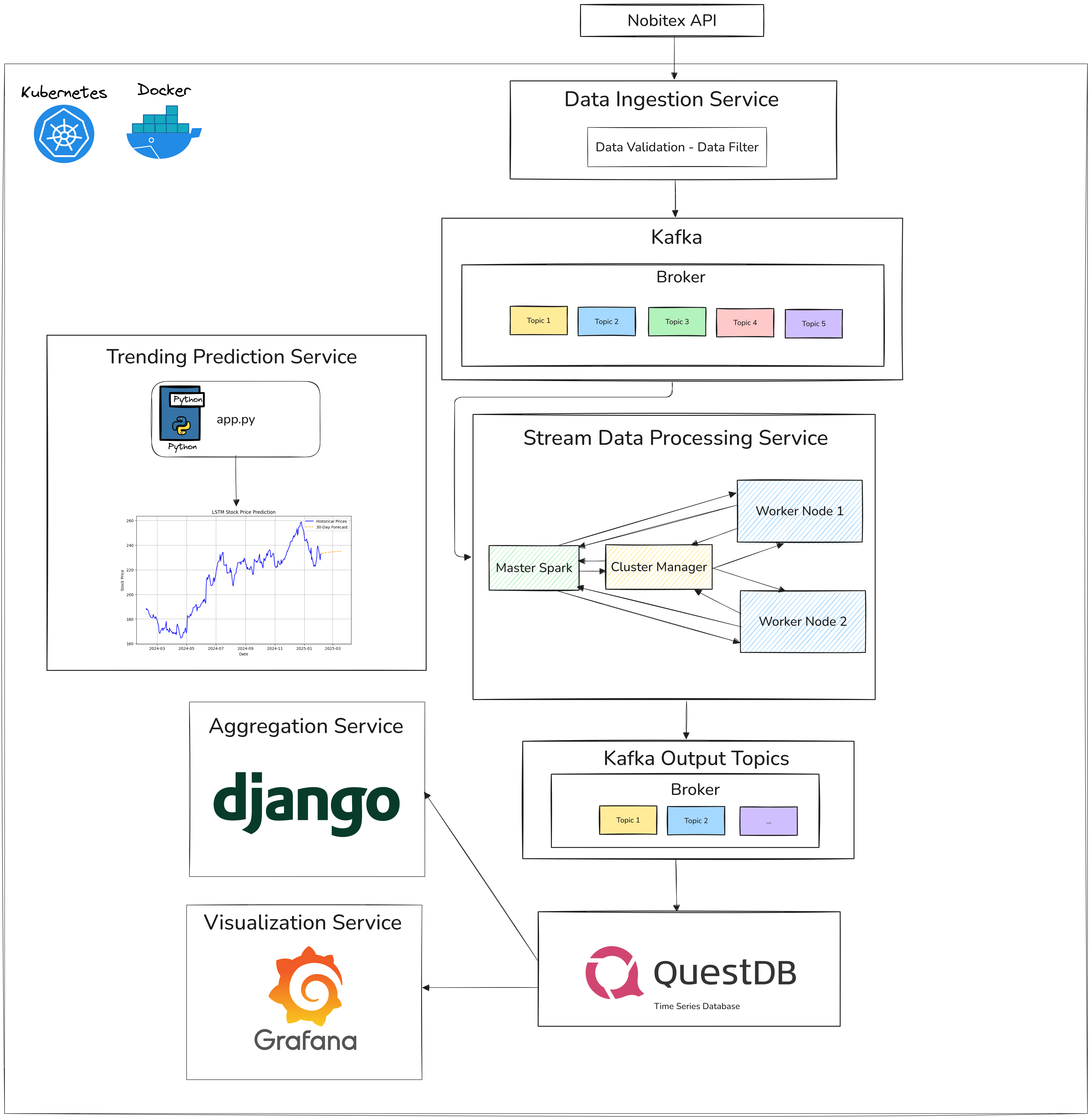
در این پروژه، برخی از چالش‌ها به شرح زیر هستند:

* **حجم بالای داده‌ها:** با توجه به حجم عظیم داده‌های موجود در بازارهای مالی، پردازش سریع داده‌ها به صورت بلادرنگ نیازمند استفاده از سیستم‌های توزیع شده است.
* **تاخیر در پردازش (Latency):** سیستم باید قادر باشد تا داده‌ها را با کمترین تاخیر پردازش کرده و سیگنال‌های دقیق و به موقع برای کاربران ارسال نماید.
* **مقیاس‌پذیری:** با توجه به اینکه داده‌ها به صورت لحظه‌ای وارد سیستم می‌شوند، مقیاس‌پذیری سیستم برای مدیریت تعداد زیاد درخواست‌ها و داده‌ها ضروری است.
* **پیشبینی روند با استفاده از هوش مصنوعی:** به دلیل پیچیدگی ذاتی داده‌های مالی و همچنین پیشبینی آن‌ها، استفاده از الگوریتم مناسب برای تحلیل روند در آینده بسیار چالش برانگیز و محل بحث است.

هدف از این پروژه این است که با استفاده از معماری‌های مقیاس‌پذیر، پردازش‌های توزیع شده و استفاده از هوش مصنوعی (یادگیری عمیق) این چالش‌ها را پشت سر گذاشته و سیستمی کارآمد و سریع برای تحلیل داده‌های بازارهای مالی ارائه دهد.

# **معماری سیستم**

معماری سیستم به شرح زیر است:



این معماری یک سیستم توزیع شده برای پردازش داده‌ها به صورت بلادرنگ است که شامل اجزای مختلفی می‌باشد. در ادامه توضیحات معماری ارائه شده به تفصیل شرح داده خواهد شد.

## API Nobitex

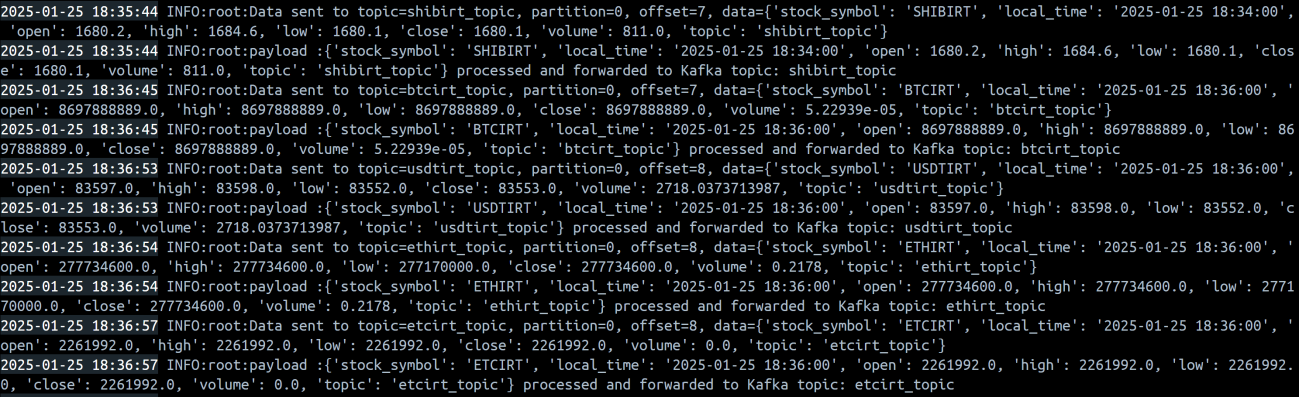
این API به عنوان منبع داده‌های رمز ارز عمل می‌کند. داده‌ها شامل اطلاعات معاملات، قیمت‌ها، حجم بازار و سایر معیارهای مالی مرتبط با رمزارزها می‌باشد. ساختار داده‌های دریافتی به شرح زیر است:

|  |  |  |
| --- | --- | --- |
| **نام متغیر** | نوع متغیر | توضیحات |
| **stock\_symbol** | object | نماد سهام |
| **local\_time** | timestamps | تاریخ و ساعت دقیق دریافت داده |
| **open\_prices** | float | قیمت باز شدن (بر اساس کندل) |
| **high\_prices** | float | بالاترین قیمت |
| **low\_prices** | float | پایین‌ترین قیمت |
| **close\_prices** | float | قیمت بسته شدن |
| **volumes** | int | حجم معامله بازار |

## Data Ingestion Service

وظیفه این سرویس جمع‌آوری و انتقال داده از API Nobitex است. در این فرآیند، داده‌ها مورد اعتبارسنجی (Validation) و فیلترگذاری قرار می‌گیرند تا فقط داده‌های معتبر به مرحله بعدی ارسال شوند. در صورت صحت داده‌ها، توسط کد producer هر داده جدید بر اساس فیلد stock\_symbol به یک Topic در Kafka ارسال می‌شود.

برای مثال داده‌های مربوط به BTCIRT به تاپیک btcirt\_topic ارسال خواهد شد.



## Apache Kafka

داده‌های تصحیح شده توسط سرویس Data Ingestion به Kafka به عنوان یک سیستم مقیاس‌پذیر و پایدار وارد می‌شوند. Kafka امکان انتقال داده‌ها به اجزای مختلف را به شیوه‌ای بسیار کارآمد فراهم می‌کند.

Broker: مسئول توزیع داده‌ها در قالب Topicهای مختلف است. در معماری سیستم چندین Topic نمایش داده شده است که هر کدام ممکن است یک جزئی از داده را نمایش دهند.

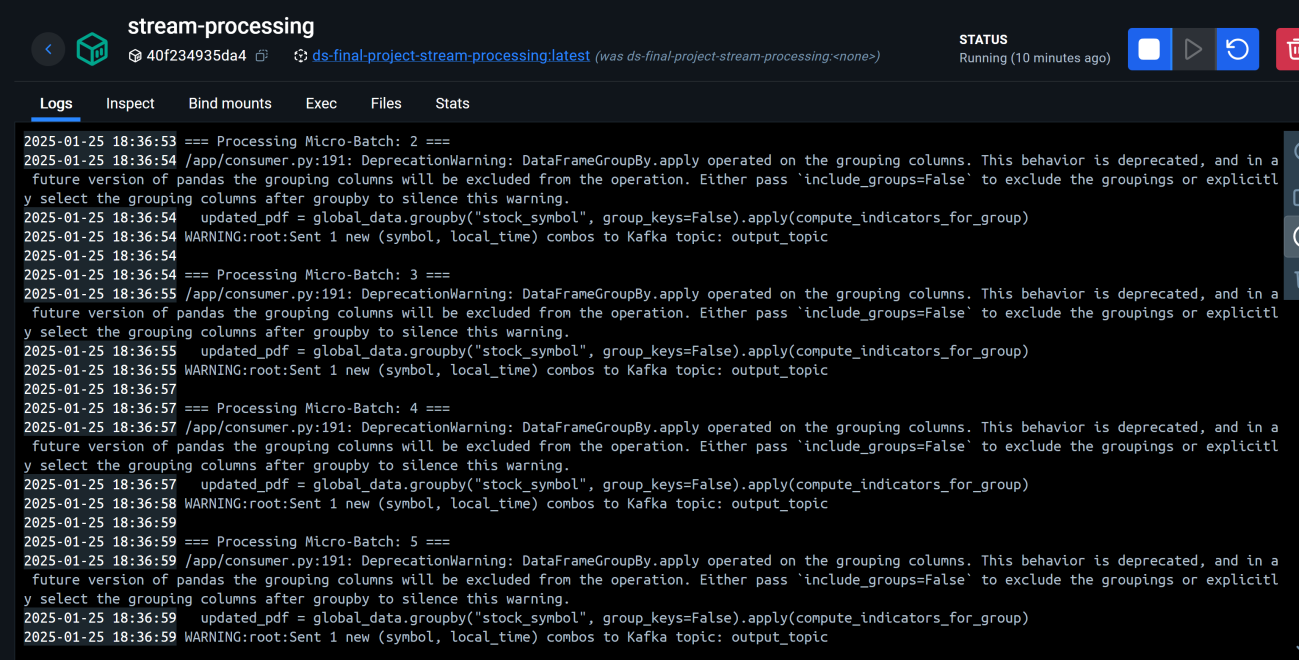
## Stream Data Processing Service (Apache Spark)

داده‌های ارسال شده از Kafka در اینجا پردازش می‌شوند. Apache Spark به عنوان یک محیط پردازش موازی برای انجام عملیات پیچیده روی داده‌ها عمل می‌کند.

Master Spark و Cluster Spark: مدیریت عملیات پردازش توزیع شده را برعهده دارند که داده‌ها میان Worker Node1 و Worker Node2 تقسیم می‌شوند.

Kafka Output Topics: داده‌های پردازش شده مجددا به Kafka ارسال می‌شوند، جایی که موضوعات خروجی جدید برای این داده‌ها ایجاد می‌شود.

این سرویس شامل چندین بخش است که به طور کلی برای پردازش و داده‌های استریم از Kafka، محاسبه شاخص‌های معاملاتی یا همان اندیکاتورها (SMA, EMA, RSI) و ارسال داده به Kafka و مدیریت میکروبچ‌ها با Spark طراحی شده است.



### محاسبه شاخص‌ها

این بخش برای هر گروه از داده‌ها شاخص‌های زیر را محاسبه می‌کند:

* **SMA (5):** میانگین ساده در دوره‌های ۵ دقیقه‌ای
* **EMA (10):** میانگین نمایی دوره‌های ۱۰ دقیقه‌ای
* **RSI (10):** شاخص قدرت نسبی در دوره‌های ۱۰ دقیقه‌ای
* **Average gain & Average Loss:** در دوره‌های ۱۰ دقیقه‌ای

### تولید سیگنال‌ها

در این بخش سیگنال‌های خرید، فروش یا نگه‌داشتن بر اساس شرایط زیر تولید می‌شود:

* **سیگنال خرید:** اگر SMA (5) > EMA(10) باشد و RSI (10) نیز کمتر از ۷۰ باشد.
* **سیگنال فروش:** اگر SMA (5) < EMA (10) باشد و RSI (10) بیشتر از ۳۰ باشد.
* در غیر این صورت **سیگنال «نگه داشتن»** یا همان hold است.

### پردازش میکروبچ‌ها

هر میکروبچ داده‌های زیر را پردازش می‌کند:

* تبدیل داده‌های Spark به Pandas
* محاسبه شاخص‌ها و سیگنال‌ها
* فیلتر کردن داده‌هایی که قبلا ارسال شده‌اند.
* ارسال داده‌های جدید به Kafka

### خواندن و نوشتن با Spark

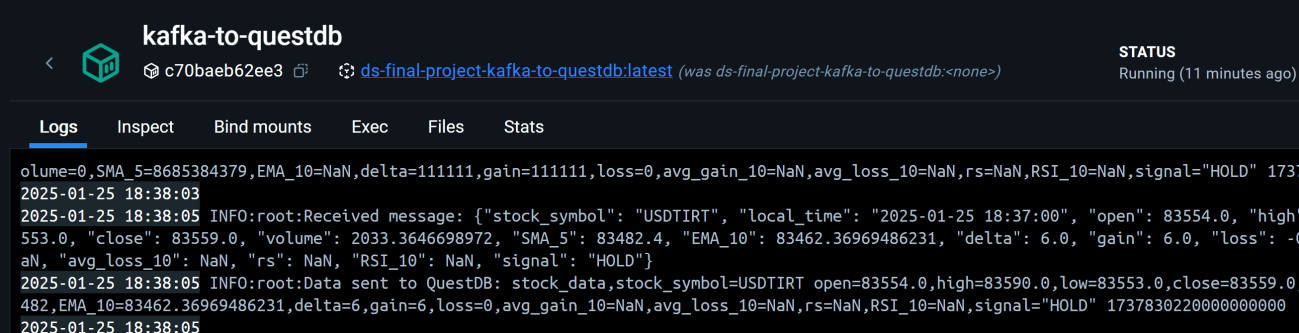
* داده‌ها از Kafka با فرمت Json خوانده می‌شوند و با استفاده از foreachBatch به صورت دسته‌ای پردازش می‌شوند.
* یک اسکیمای مشخص برای ستون‌های داده تعریف شده است.

### مدیریت استریم

داده‌ها به صورت استریم از Kafka خوانده شده، شاخص‌ها محاسبه شده و داده‌های نهایی مجددا به output\_topic ارسال می‌شوند.

## سرویس میانی Kafka-to-QuestDB

این سرویس برای دریافت داده‌ها از Kafka، پردازش آن‌ها و ارسال آن‌ها به QuestDB از طریق Influx Line Protocol یا ILP طراحی شده است. در ادامه بخش‌های مختلف آن توضیح داده می‌شود.



* **Kafka Consumer:** برای دریافت پیام‌ها از Kafka
* **Socket:** برای ارتباط با QuestDB از طریق پروتکل ILP
* **کتابخانه‌های Pandas و math:** برای پردازش داده‌ها

### پردازش پیام‌ها

* پیام‌ها از Kafka دریافت و مقدار آن‌ها (JSON) پردازش می‌شود.
* اگر پیام شامل داده‌های نامعتبر باشد، خطا گزارش می‌شود.

### جایگزینی nan و تبدیل مقادیر

* مقدار “NaN” با مقدار float(‘nan’) جایگزین می‌شود.
* مقادیر عددی مانند volume و سایر شاخص‌ها به نوع مناسب تبدیل می‌شوند.

### ساخت قالب ILP و ارسال داده‌ها به QuestDB

* داده‌ها در قالب Influx Line Protocol برای QuestDB قالب‌بندی می‌شوند.
* هر رکورد شامل اطلاعاتی مثل قیمت‌ها (open, close,…) که پیشتر در ابتدا آمده بود و شاخص‌های محاسبه شده است.

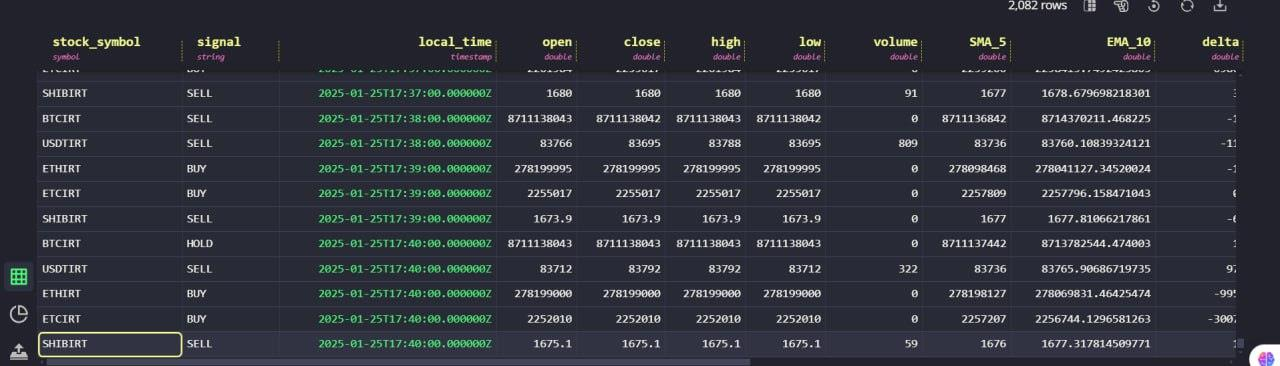
## QuestDB

داده‌های پردازش شده در QuestDB ذخیره می‌شوند. QuestDB یک پایگاه داده سری زمانی است که برای انجام تحلیل‌های پیچیده و بلادرنگ استفاده می‌شوند. این پایگاه داده برای داده‌های بازارهای مالی نیز بسیار مناسب است و stock\_symbol را به صورت خاص index گذاری می‌کند که به صورت سریع بتواند داده‌های مربوط به هر نماد سهام را در کوئری‌ها برگرداند.

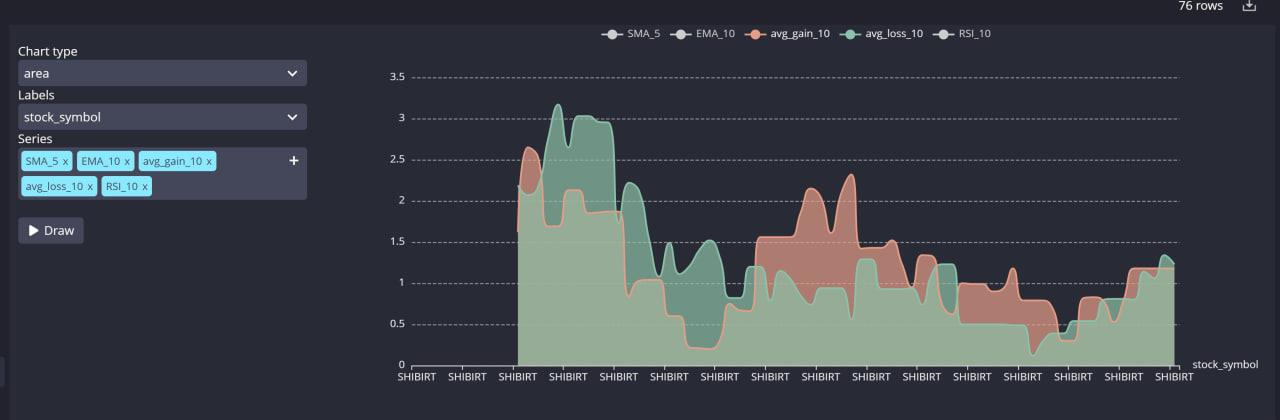
استفاده از این دیتابیس بسیار بهینه‌تر از ترکیب یک دیتابیس دیگر مثل PostgreSQL با Redis است زیرا:

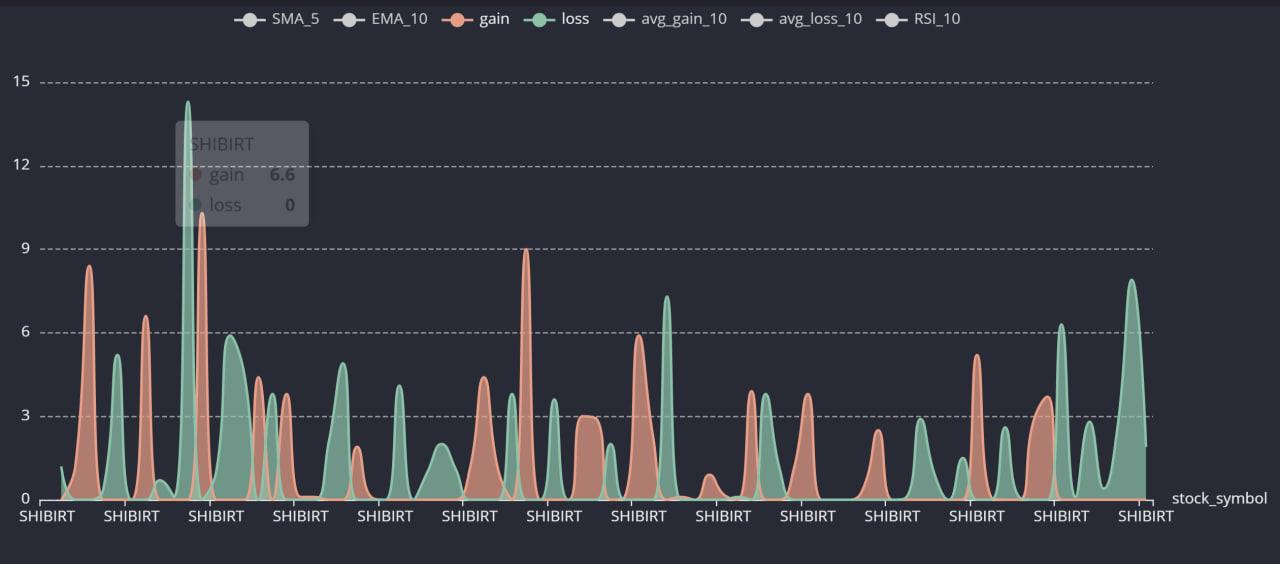
* **عملکرد بالا:** QuestDB برای پردازش داده‌های سری زمانی بهینه شده و می‌تواند حجم عظیمی از داده‌ها را با سرعت بسیار بالا وارد کوئری کند.
* **کاهش پیچیدگی معماری:** با استفاده از QuestDB دیگر نیازی به ترکیب چندین ابزار نیست.
* **مقیاس‌پذیری:** QuestDB در مدیریت حجم عظیمی از داده‌های سری زمانی مقیاس‌پذیری بالایی دارد، در حالی که Redis برای داده‌های بسیار بزرگ نیاز به حافظه بیشتر و تنظیمات پیچیده‌تری دارد.
* **پشتیبانی از ILP:** امکان ارسال داده‌ها با فرمت ILP باعث ساده‌تر شدن ارسال داده‌های بازار سهام می‌شود.
* جستجو و انجام کوئری‌های پیشرفته SQL: قابلیت استفاده از SQL استاندارد برای کوئری‌های پیچده مانند محاسبه میانگین، محاسبه شاخص‌های معاملاتی (EMA, SMA, …) و جستجوی مقادیر خاص در بازه‌های زمانی
* **پشتیبانی از داده‌های بلادرنگ (Realtime):** مناسب برای سیستم‌های بلادرنگ که داده‌های بازار سهام را در لحظه پردازش و تحلیل می‌کنند.
* **مقیاس‌پذیری بالا:** مناسب برای ذخیره و پردازش داده‌های حجیم مانند داده‌های تاریخی بازار سهام
* **پشتیبانی از چارت‌های مختلف برای تحلیل بصری داده‌**

در زیر تصویری از محیط دیتابیس نمایش داده شده است:



در زیر تصویری از چارت‌های QuestDB نمایش داده شده است:





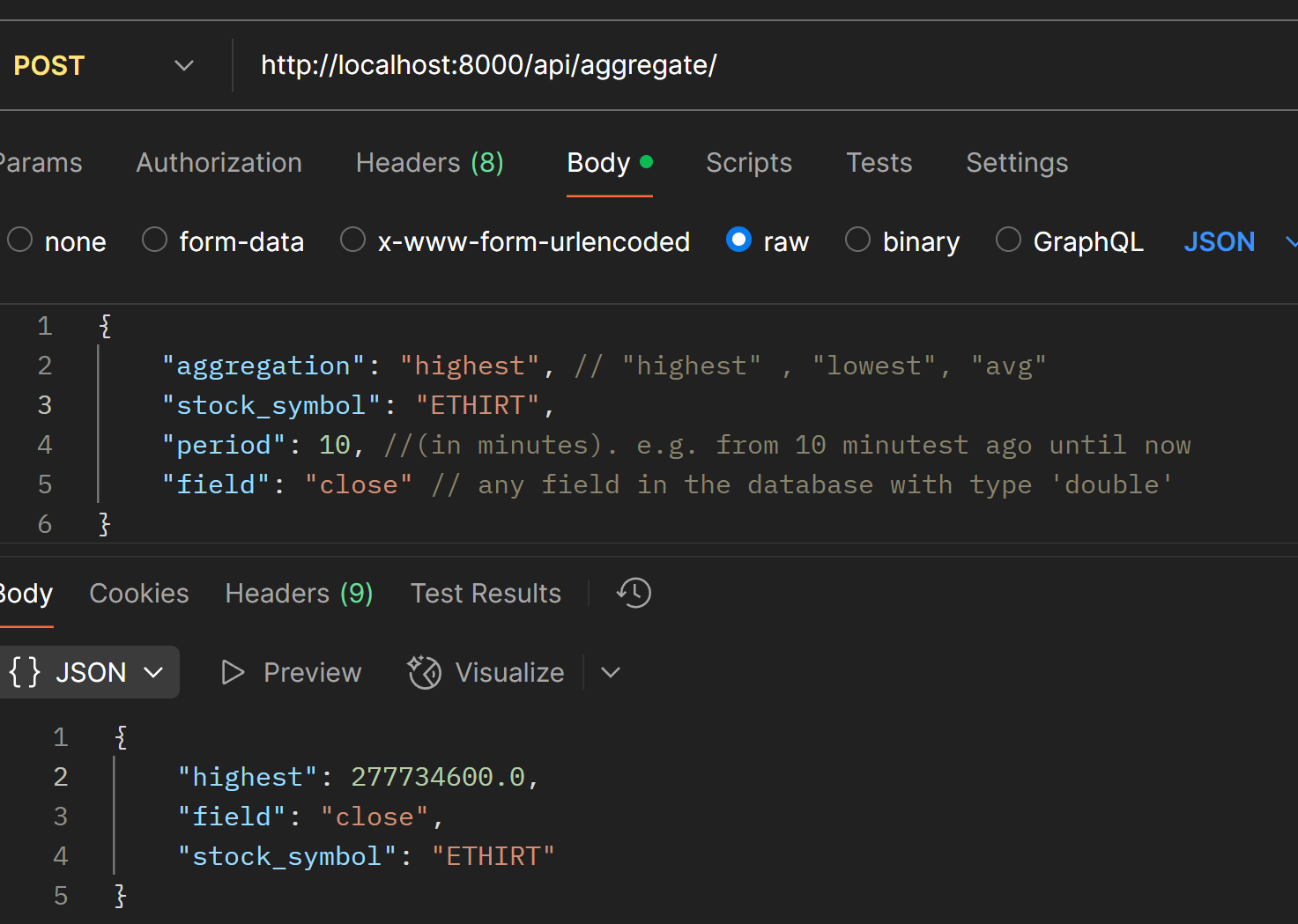


## Aggregation Service

در این سرویس از یک اپلیکیشن جنگو استفاده شده است که شامل ۳ نوع API است.

### Aggregate API

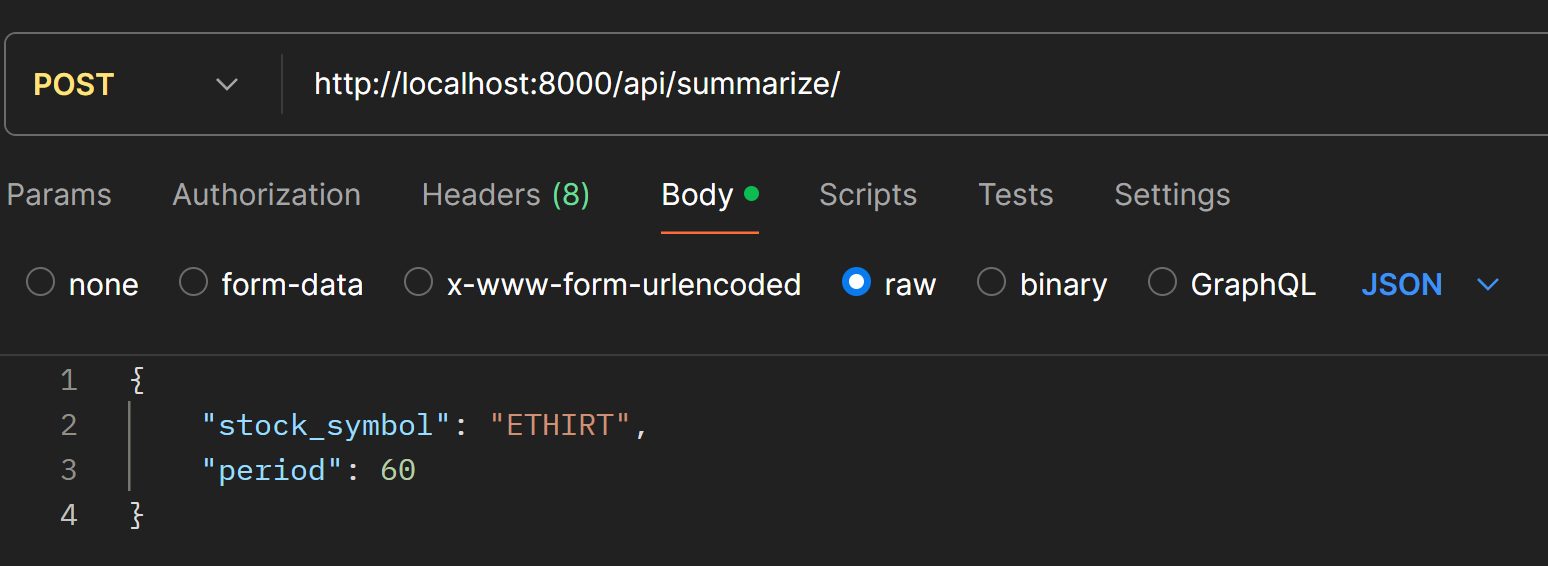
در این API با نوشتن اسم سهام مورد نظر، نوع aggregation شامل میانگین‌گیری، بیشترین یا کمترین مقدار، بازه زمانی مورد نظر و فیلد انتخابی مورد نظر جواب مناسب خواهید گرفت. برای مثال:



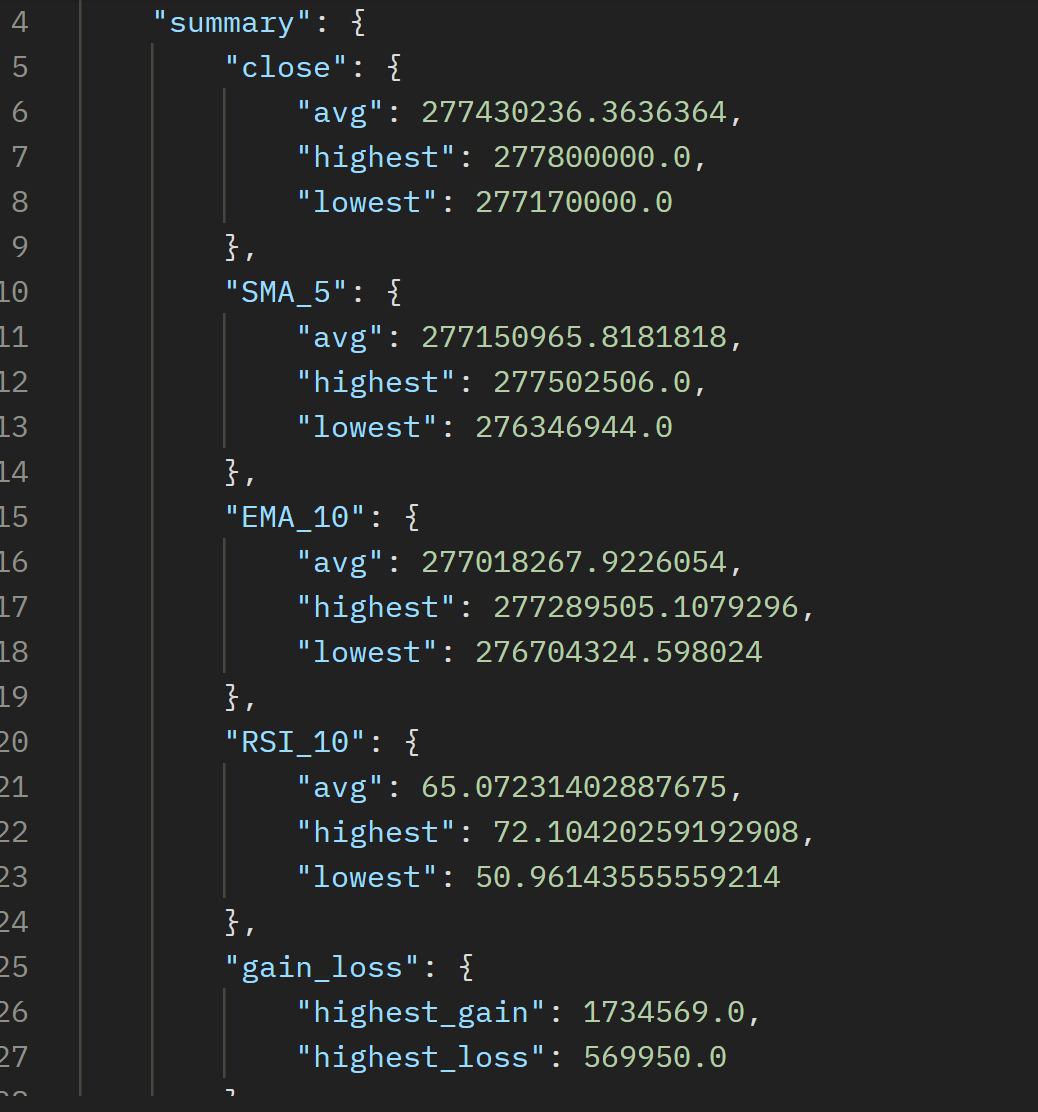
### Summarize API

خلاصه عملکرد یک سهام در بازه زمانی مشخص را برای معیارهای SMA (5), EMA (10), RSI (10), loss, gain, close تولید می‌کند.

برای مثال ورودی:



و خروجی آن:



### Summarize Multiple API

لیستی از چند نماد سهام را به عنوان ورودی گرفته و خلاصه عملکرد را برای هر کدام در بازه زمانی مشخص تولید می‌کند.

## Visualization Service (Grafana)

در سرویس گرافانا که برای نمایش نمودارها و چارت‌ها استفاده شده است این قابلیت وجود دارد که به صورت socket-based برای دریافت سیگنال و همچنین query-based برای انجام عملیات aggregation استفاده شود.

در این سرویس در گرافانا یک داشبورد برای هر سهام ساخته شده است که در یک نگاه خلاصه‌ای از عملکرد سهام و نمودار‌های آن نشان داده شده است که به شرح زیر است:

