

# بسم الله الرحمن الرحيم



## سیستم تحلیل مالی و معاملات بلادرنگ با هوش مصنوعی

پروژه نهایی درس تحلیل‌ها و سیستم‌های داده حجیم



استاد درس: دکتر حسن نادری

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

بهمن ۱۴۰۳

## فهرست مطالب

3	مقدمه
3	هدف پروژه
3	اهمیت پروژه
3	چالش‌ها
4	معماری سیستم
5	API Nobitex
5	Data Ingestion Service
5	Apache Kafka
5	Stream Data Processing Service (Apache Spark)
6	محاسبه شاخص‌ها
6	تولید سیگنال‌ها
7	پردازش میکروبیج‌ها
7	خواندن و نوشتن با Spark
7	مدیریت استریم
7	سرویس میانی Kafka-to-QuestDB
7	پردازش پیام‌ها
7	جایگزینی nan و تبدیل مقادیر
8	ساخت قالب ILP و ارسال داده‌ها به QuestDB
8	QuestDB
9	Aggregation Service
9	Aggregate API
9	Summarize API
10	Summarize Multiple API
10	Visualization Service (Grafana)
12	تکنولوژی‌های استفاده شده
12	API Nobitex
12	نحوه دریافت داده
12	چرا این API انتخاب شده است؟
12	Apache Kafka
12	چرا Kafka انتخاب شده است؟

- 12 ..... معماری Kafka
- 13 ..... Apache Spark
- 13 ..... چرا اسپارک انتخاب شده است؟
- 14 ..... معماری Spark
- 14 ..... QuestDB
- 14 ..... چرا از QuestDB استفاده شده است؟
- 15 ..... Minikube
- 15 ..... چرا Minikube انتخاب شده است؟

## 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 معامله‌گران 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 Spark، Apache Kafka 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

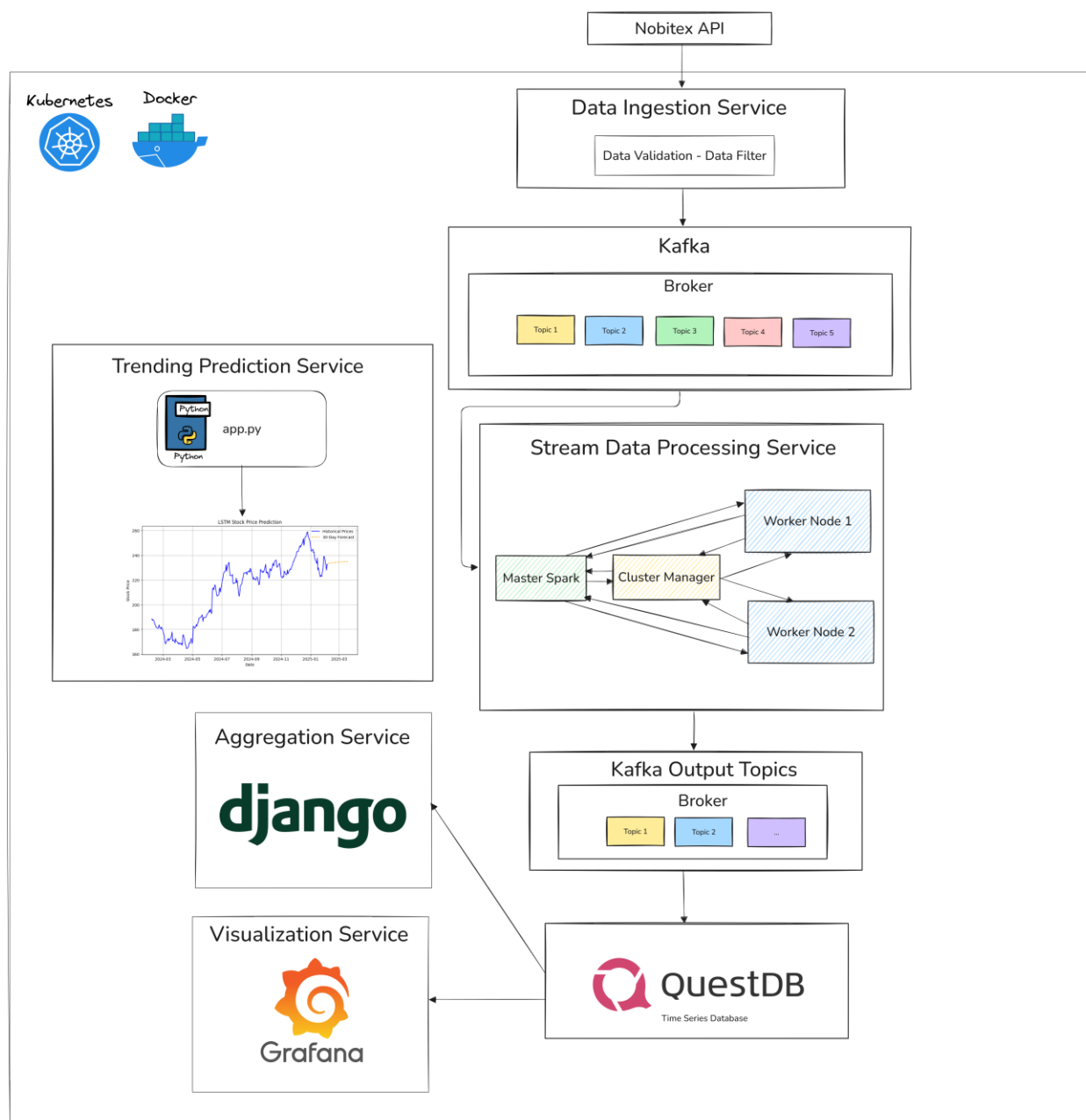
در این پروژه، برخی از چالش‌ها به شرح زیر هستند:

- **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.

```
2025-01-25 18:35:44 INFO:root:Data sent to topic=shibirt_topic, partition=0, offset=7, data={'stock_symbol': 'SHIBIRT', 'local_time': '2025-01-25 18:34:00', 'open': 1680.2, 'high': 1684.6, 'low': 1680.1, 'close': 1680.1, 'volume': 811.0, 'topic': 'shibirt_topic'}
2025-01-25 18:35:44 INFO:root:payload: {'stock_symbol': 'SHIBIRT', 'local_time': '2025-01-25 18:34:00', 'open': 1680.2, 'high': 1684.6, 'low': 1680.1, 'close': 1680.1, 'volume': 811.0, 'topic': 'shibirt_topic'} processed and forwarded to Kafka topic: shibirt_topic
2025-01-25 18:36:45 INFO:root:Data sent to topic=btctrt_topic, partition=0, offset=7, data={'stock_symbol': 'BTCIRT', 'local_time': '2025-01-25 18:36:00', 'open': 8697888889.0, 'high': 8697888889.0, 'low': 8697888889.0, 'close': 8697888889.0, 'volume': 5.22939e-05, 'topic': 'btctrt_topic'}
2025-01-25 18:36:45 INFO:root:payload: {'stock_symbol': 'BTCIRT', 'local_time': '2025-01-25 18:36:00', 'open': 8697888889.0, 'high': 8697888889.0, 'low': 8697888889.0, 'close': 8697888889.0, 'volume': 5.22939e-05, 'topic': 'btctrt_topic'} processed and forwarded to Kafka topic: btctrt_topic
2025-01-25 18:36:53 INFO:root:Data sent to topic=usdtirt_topic, partition=0, offset=8, data={'stock_symbol': 'USDIRT', 'local_time': '2025-01-25 18:36:00', 'open': 83597.0, 'high': 83598.0, 'low': 83552.0, 'close': 83553.0, 'volume': 2718.0373713987, 'topic': 'usdtirt_topic'}
2025-01-25 18:36:53 INFO:root:payload: {'stock_symbol': 'USDIRT', 'local_time': '2025-01-25 18:36:00', 'open': 83597.0, 'high': 83598.0, 'low': 83552.0, 'close': 83553.0, 'volume': 2718.0373713987, 'topic': 'usdtirt_topic'} processed and forwarded to Kafka topic: usdtirt_topic
2025-01-25 18:36:54 INFO:root:Data sent to topic=ethirt_topic, partition=0, offset=8, data={'stock_symbol': 'ETHIRT', 'local_time': '2025-01-25 18:36:00', 'open': 277734600.0, 'high': 277734600.0, 'low': 277170000.0, 'close': 277734600.0, 'volume': 0.2178, 'topic': 'ethirt_topic'}
2025-01-25 18:36:54 INFO:root:payload: {'stock_symbol': 'ETHIRT', 'local_time': '2025-01-25 18:36:00', 'open': 277734600.0, 'high': 277734600.0, 'low': 277170000.0, 'close': 277734600.0, 'volume': 0.2178, 'topic': 'ethirt_topic'} processed and forwarded to Kafka topic: ethirt_topic
2025-01-25 18:36:57 INFO:root:Data sent to topic=etctrt_topic, partition=0, offset=8, data={'stock_symbol': 'ETCIRT', 'local_time': '2025-01-25 18:36:00', 'open': 2261992.0, 'high': 2261992.0, 'low': 2261992.0, 'close': 2261992.0, 'volume': 0.0, 'topic': 'etctrt_topic'}
2025-01-25 18:36:57 INFO:root:payload: {'stock_symbol': 'ETCIRT', 'local_time': '2025-01-25 18:36:00', 'open': 2261992.0, 'high': 2261992.0, 'low': 2261992.0, 'close': 2261992.0, 'volume': 0.0, 'topic': 'etctrt_topic'} processed and forwarded to Kafka topic: etctrt_topic
```

## Apache Kafka

داده‌های تصحیح شده توسط سرویس Data Ingestion به Kafka به عنوان یک سیستم مقیاس‌پذیر و پایدار وارد می‌شوند. 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.

Cluster Spark و Master 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 مجدداً 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 میکروبیج‌ها با Spark طراحی شده است.

```
stream-processing
40f234935da4 ds-final-project-stream-processing:latest (was ds-final-project-stream-processing:<none>)
STATUS Running (10 minutes ago)
Logs Inspect Bind mounts Exec Files Stats
2025-01-25 18:36:53 === Processing Micro-Batch: 2 ===
2025-01-25 18:36:54 /app/consumer.py:191: DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass 'include_groups=False' to exclude the groupings or explicitly select the grouping columns after groupby to silence this warning.
2025-01-25 18:36:54 updated_pdf = global_data.groupby("stock_symbol", group_keys=False).apply(compute_indicators_for_group)
2025-01-25 18:36:54 WARNING:root:Sent 1 new (symbol, local_time) combos to Kafka topic: output_topic
2025-01-25 18:36:54 === Processing Micro-Batch: 3 ===
2025-01-25 18:36:55 /app/consumer.py:191: DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass 'include_groups=False' to exclude the groupings or explicitly select the grouping columns after groupby to silence this warning.
2025-01-25 18:36:55 updated_pdf = global_data.groupby("stock_symbol", group_keys=False).apply(compute_indicators_for_group)
2025-01-25 18:36:55 WARNING:root:Sent 1 new (symbol, local_time) combos to Kafka topic: output_topic
2025-01-25 18:36:57 === Processing Micro-Batch: 4 ===
2025-01-25 18:36:57 /app/consumer.py:191: DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass 'include_groups=False' to exclude the groupings or explicitly select the grouping columns after groupby to silence this warning.
2025-01-25 18:36:57 updated_pdf = global_data.groupby("stock_symbol", group_keys=False).apply(compute_indicators_for_group)
2025-01-25 18:36:58 WARNING:root:Sent 1 new (symbol, local_time) combos to Kafka topic: output_topic
2025-01-25 18:36:59 === Processing Micro-Batch: 5 ===
2025-01-25 18:36:59 /app/consumer.py:191: DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass 'include_groups=False' to exclude the groupings or explicitly select the grouping columns after groupby to silence this warning.
2025-01-25 18:36:59 updated_pdf = global_data.groupby("stock_symbol", group_keys=False).apply(compute_indicators_for_group)
2025-01-25 18:36:59 WARNING:root:Sent 1 new (symbol, local_time) combos to Kafka topic: output_topic
```

## 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 نگه داشتن بر اساس شرایط زیر تولید می‌شود:

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

## پردازش میکروبیج‌ها

هر میکروبیج داده‌های زیر را پردازش می‌کند:

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

## خواندن و نوشتن با Spark

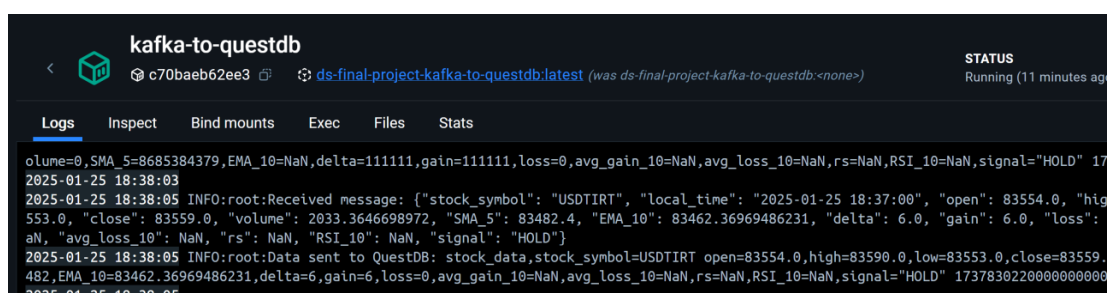
- داده‌ها از Kafka با فرمت Json خوانده می‌شوند و با استفاده از foreachBatch به صورت دسته‌ای پردازش می‌شوند.
- یک اسکیمای مشخص برای ستون‌های داده تعریف شده است.

## مدیریت استریم

داده‌ها به صورت استریم از Kafka خوانده شده، شاخص‌ها محاسبه شده و داده‌های نهایی مجدداً 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.



The screenshot shows the 'kafka-to-questdb' application interface. At the top, there's a header with the application name, a status bar indicating it's 'Running (11 minutes ago)', and navigation tabs for 'Logs', 'Inspect', 'Bind mounts', 'Exec', 'Files', and 'Stats'. The 'Logs' tab is active, displaying a log stream. The log entries show stock data processing, including fields like 'stock\_symbol', 'local\_time', 'open', 'high', 'close', 'volume', 'SMA\_5', 'EMA\_10', 'delta', 'gain', 'loss', 'avg\_gain\_10', 'avg\_loss\_10', 'rs', 'RSI\_10', and 'signal'. The log also shows messages received from Kafka and data sent to QuestDB.

- **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 بیشتر 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.

در زیر تصویری از محیط دیتابیس نمایش داده شده است:

stock_symbol	signal	local_time	open	close	high	low	volume	SMA_5	EMA_10	delta
SHIBRT	SELL	2025-01-25T17:37:00.000000Z	1680	1680	1680	1680	91	1677	1678.679698218301	...
BTCTRT	SELL	2025-01-25T17:38:00.000000Z	8711138043	8711138042	8711138043	8711138042	0	8711136842	8711378211.468225	...
USDTIRT	SELL	2025-01-25T17:38:00.000000Z	83766	83695	83788	83695	809	83736	83768.10839324121	...
ETHIRT	BUY	2025-01-25T17:39:00.000000Z	278199995	278199995	278199995	278199995	0	278098468	278041127.34520824	...
ETCIRT	BUY	2025-01-25T17:39:00.000000Z	2255817	2255817	2255817	2255817	0	2257809	2257796.158471043	...
SHIBRT	SELL	2025-01-25T17:39:00.000000Z	1673.9	1673.9	1673.9	1673.9	0	1677	1677.81066217861	...
BTCTRT	HOLD	2025-01-25T17:40:00.000000Z	8711138043	8711138043	8711138043	8711138043	0	8711137442	8713782544.474003	...
USDTIRT	SELL	2025-01-25T17:40:00.000000Z	83712	83792	83792	83712	322	83736	83765.98686719735	...
ETHIRT	BUY	2025-01-25T17:40:00.000000Z	278199000	278199000	278199000	278199000	0	278198127	278069831.46425474	...
ETCIRT	BUY	2025-01-25T17:40:00.000000Z	2252810	2252810	2252810	2252810	0	2257287	2256744.1296581263	...
SHIBRT	SELL	2025-01-25T17:40:00.000000Z	1675.1	1675.1	1675.1	1675.1	59	1676	1677.317814599771	...

در زیر تصویری از چارت‌های QuestDB نمایش داده شده است:

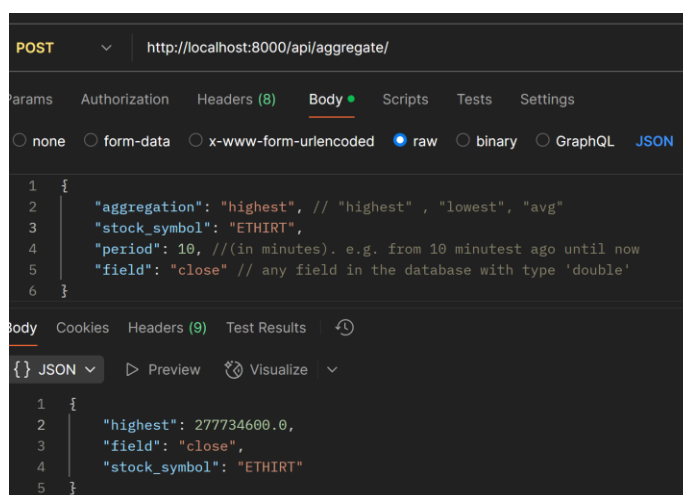


## Aggregation Service

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

### Aggregate API

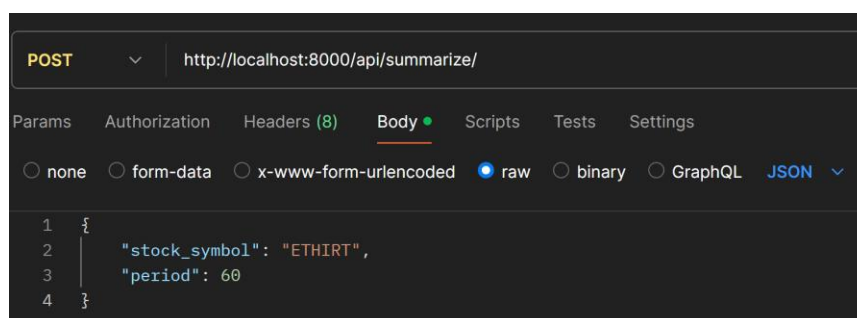
در این API با نوشتن اسم سهام مورد نظر، نوع aggregation شامل میانگین‌گیری، بیشترین یا کمترین مقدار، بازه زمانی مورد نظر و فیلد انتخابی مورد نظر جواب مناسب خواهید گرفت. برای مثال:



### 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:

```
4      "summary": {
5        "close": {
6          "avg": 277430236.3636364,
7          "highest": 277800000.0,
8          "lowest": 277170000.0
9        },
10       "SMA_5": {
11         "avg": 277150965.8181818,
12         "highest": 277502506.0,
13         "lowest": 276346944.0
14       },
15       "EMA_10": {
16         "avg": 277018267.9226054,
17         "highest": 277289505.1079296,
18         "lowest": 276704324.598024
19       },
20       "RSI_10": {
21         "avg": 65.07231402887675,
22         "highest": 72.10420259192908,
23         "lowest": 50.9614355559214
24       },
25       "gain_loss": {
26         "highest_gain": 1734569.0,
27         "highest_loss": 569950.0
28     }
```

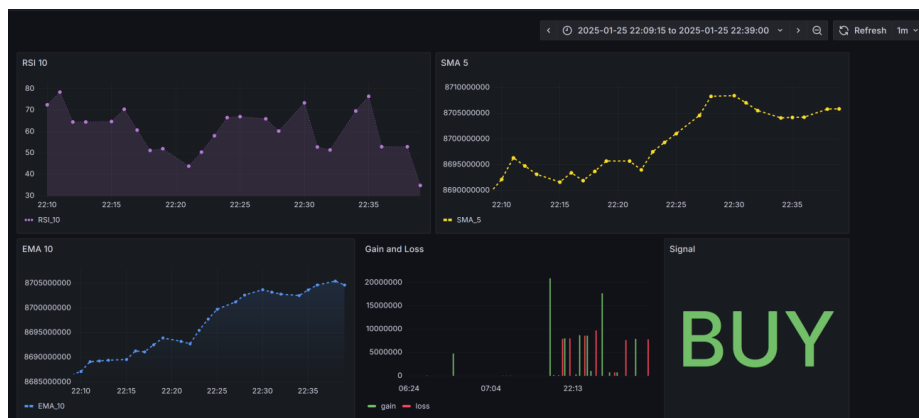
### Summarize Multiple API

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:



قابلیت نشان دادن سیگنال به صورت کوثری و یا استفاده از WebSocket که به شرح زیر است:

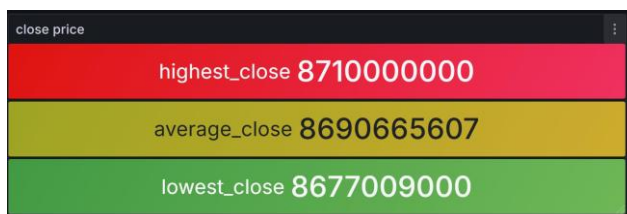


در نمودار زیر با استفاده از مقادیر دریافتی سهام، Candlestick آن را در گرافانا ساخته و مقادیر SMA، EMA را نمایش می‌دهیم:

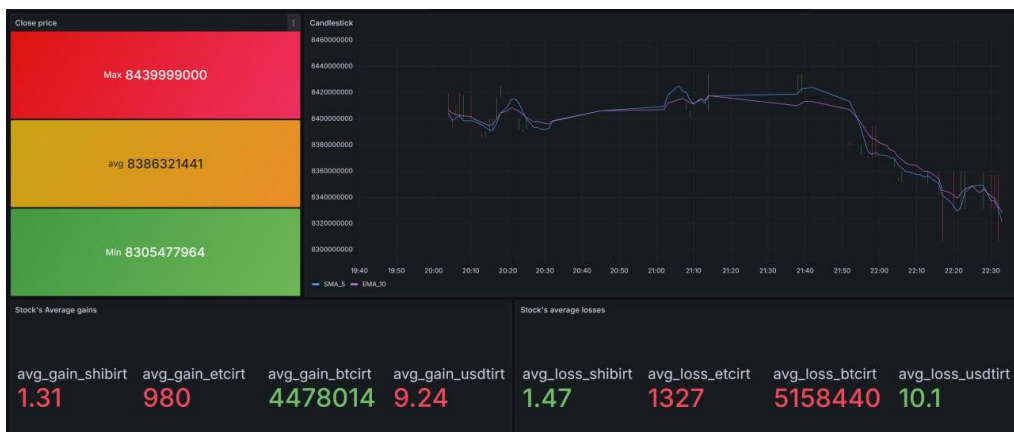


«خط بنفش نشان‌دهنده EMA و خط آبی نشان‌دهنده SMA است»

همچنین اعمال عملیات Aggregation برای نمایش خلاصه‌ای از عملکرد سهام:



همچنین در داشبورد زیر میانگین سود تو سهام‌های مختلف و میانگین ضرر در سهام‌های مختلف آورده شده است:



## تکنولوژی‌های استفاده شده

### API Nobitex

یک رابط برنامه نویسی کاربردی است که به شما اجازه می‌دهد تا اطلاعات بازار رمز ارز را از صرافی Nobitex دریافت نمایید. این API می‌تواند داده‌هایی همچون قیمت‌ها، حجم معاملات، نرخ تغییرات و اطلاعات دیگر از بازار رمز ارزها را در اختیار شما قرار دهد.

### نحوه دریافت داده

از طریق API Nobitex می‌توان درخواست‌هایی ارسال کرد که داده‌های مختلف مانند قیمت رمز ارزها و حجم معاملات، بالاترین و پایین‌ترین قیمت‌ها و تغییرات قیمت در یک بازه زمانی مشخص را دریافت کند. این داده‌ها معمولاً در قالب JSON ارسال می‌شوند که به راحتی می‌توان آن‌ها را پردازش و در سیستم‌های دیگر مورد استفاده قرار داد.

### چرا این API انتخاب شده است؟

- **دقت بالا:** Nobitex یک صرافی معتبر در ایران است که داده‌های دقیق و به روز بازارهای رمز ارز را ارائه می‌دهد.
- **سرعت بالا:** این API دارای سرعت بسیار بالایی در ارسال و دریافت داده دارد که برای تحلیل بلادرنگ بسیار حیاتی می‌باشد.
- **قابلیت‌های خاص:** این API امکان دسترسی به انواع داده‌ها را فراهم می‌آورد که برای تجزیه و تحلیل مناسب است.

### Apache Kafka

یک سیستم پیام‌رسان توزیع شده و مقیاس‌پذیر است که برای ارسال، دریافت و ذخیره‌سازی داده استفاده می‌شود. Kafka به ویژه برای پردازش داده‌ها به صورت Realtime و در مقیاس بزرگ به کار می‌رود.

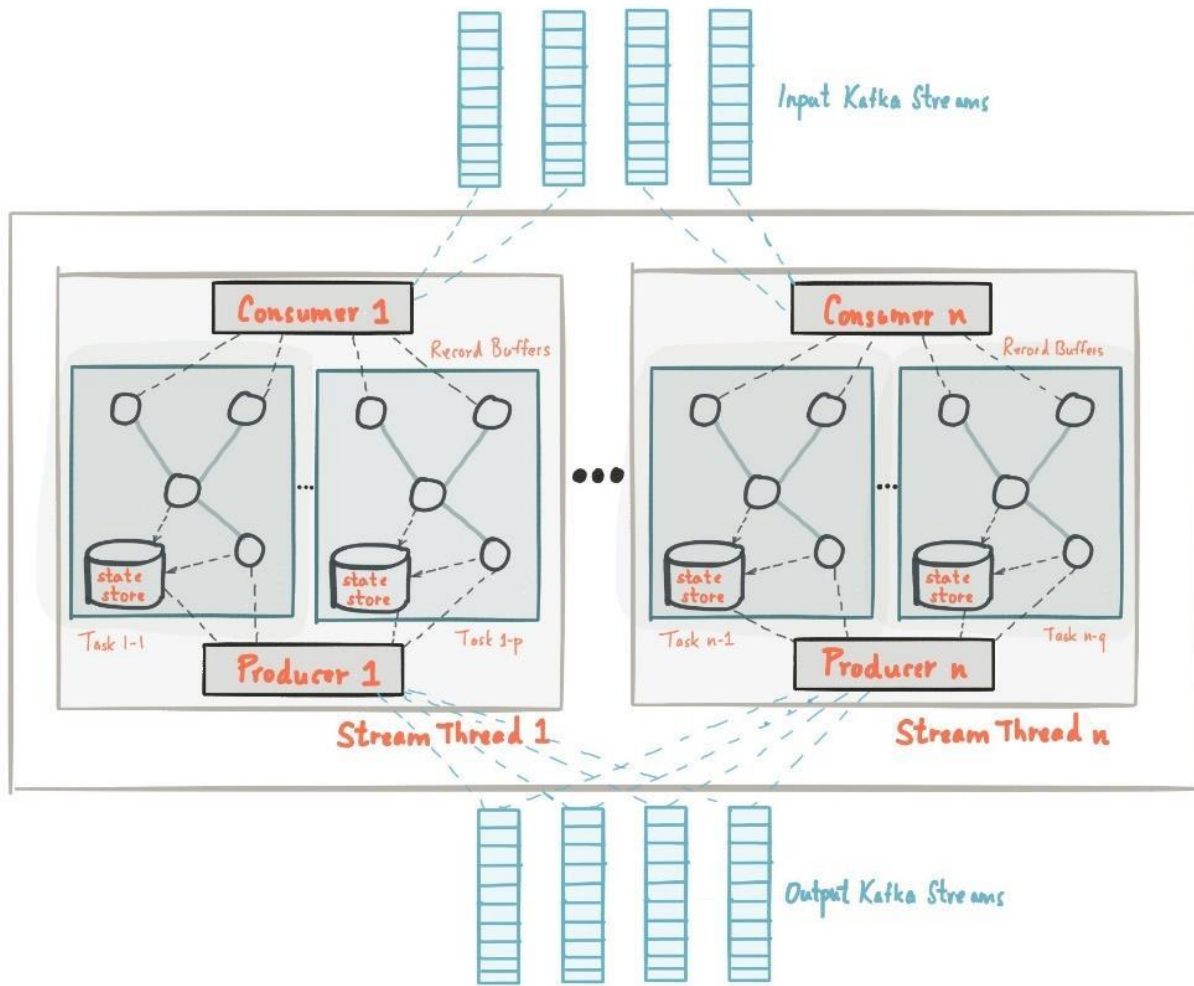
### چرا Kafka انتخاب شده است؟

- **مقاوم در برابر خرابی‌ها (Fault Tolerance):** Kafka قادر است داده‌ها را در صورت بروز مشکلات در سیستم‌های مختلف نگهداری کرده و از دست رفتن داده‌ها جلوگیری کند. این ویژگی در پروژه‌هایی که نیاز به پردازش داده‌های حیاتی دارند، ضروری است.
- **مقیاس‌پذیری:** Kafka قادر است حجم عظیمی از داده‌ها را به طور همزمان پردازش کرده و بدون کاهش کارایی، داده‌ها را به نودهای مختلف ارسال کند. این ویژگی برای سیستم‌های بلادرنگ و توزیع شده مانند پروژه‌های مالی بسیار مناسب است.

### معماری Kafka

- **Producer:** این نودها مسئول ارسال داده‌ها به Kafka هستند. در این پروژه، میکروسرویس‌ها یا نودهای پردازش داده‌ها می‌توانند به عنوان Producer عمل کنند.
- **Broker:** داده‌ها را دریافت کرده و آن‌ها را در Topic‌ها ذخیره می‌کنند. این داده‌ها به نودهای مختلف (Consumer) ارسال می‌شوند.
- **Consumer:** این نودها مسئول دریافت و پردازش داده‌ها از Topic‌ها هستند. در این پروژه، Spark و سایر سیستم به عنوان Consumer داده‌ها را پردازش کنند.

در شکل زیر معماری Kafka به صورت شماتیک آورده شده است:



## Apache Spark

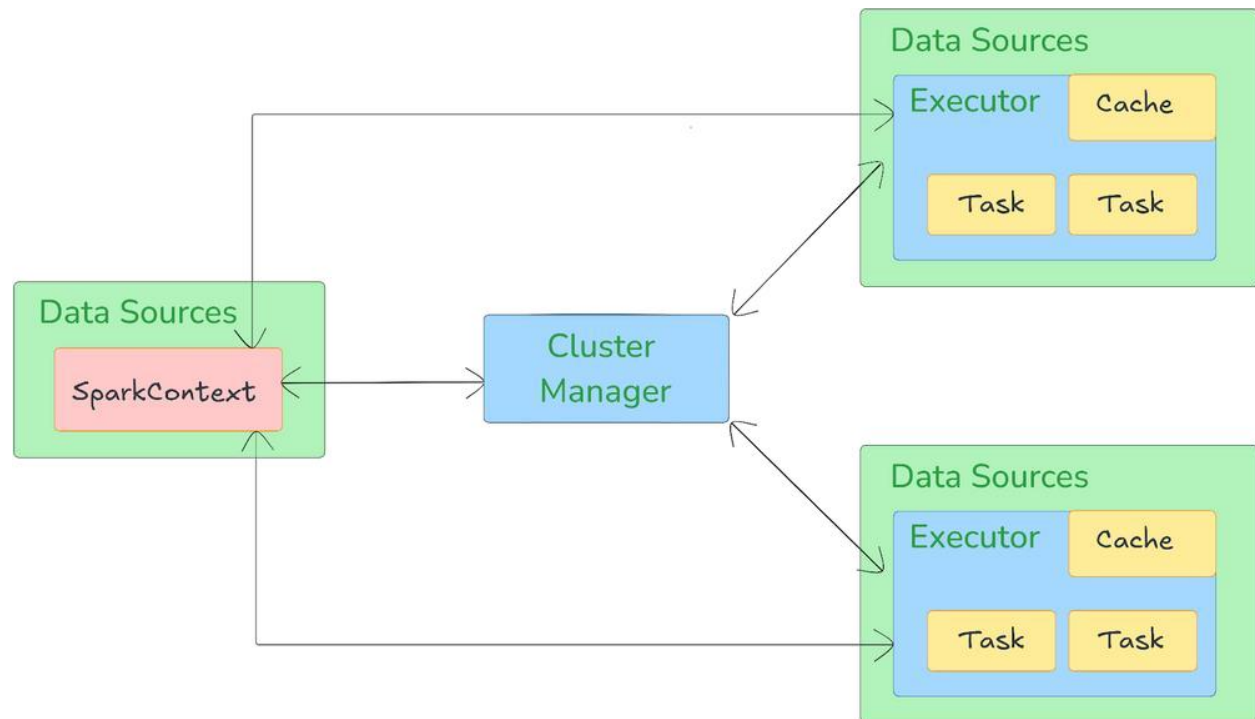
یک فریم‌ورک پردازش داده‌های توزیع شده است که قادر است داده‌ها را به صورت موازی و در مقیاس بزرگ پردازش کند. Spark در این پروژه برای تحلیل بلادرنگ داده‌های مالی به کار می‌رود.

### چرا اسپارک انتخاب شده است؟

- **پردازش بلادرنگ:** Spark به راحتی از داده‌های استریم پشتیبانی می‌کند و قادر است آن‌ها را در لحظه پردازش کند. این ویژگی به ویژه برای تحلیل داده‌های مالی که نیاز به پردازش سریع دارند، بسیار مفید است.
- **مقیاس‌پذیری:** Spark می‌تواند داده‌های حجیم را به صورت توزیع شده پردازش کند و به راحتی با افزایش حجم داده‌ها مقیاس‌پذیر می‌شود.
- **پردازش داده‌های استریم:** Spark Streaming امکان پردازش داده‌های ورودی به صورت پیوسته را فراهم می‌کند. این ویژگی برای سیستم‌هایی که نیاز به تحلیل بلادرنگ دارند، ضروری است.
- **دریاچه داده‌های توزیع شده:** Spark می‌تواند داده‌ها را در مقیاس وسیع توزیع کرده و پردازش کند، که در پروژه‌هایی با حجم بالای داده‌ها بسیار ضروری است.

## معماری Spark

- **Driver**: کنترل کننده اصلی سیستم که نودها و منابع را مدیریت می کند. Driver تصمیم می گیرد که چه داده هایی باید پردازش شوند و به چه نودی ارسال شوند.
- **Executor**: این نودها مسئول انجام پردازش ها و محاسبات واقعی هستند.
- **Cluster Manager**: این بخش برای مدیریت منابع (مثل حافظه و پردازنده ها) در کلاستر Spark استفاده می شود.



## QuestDB

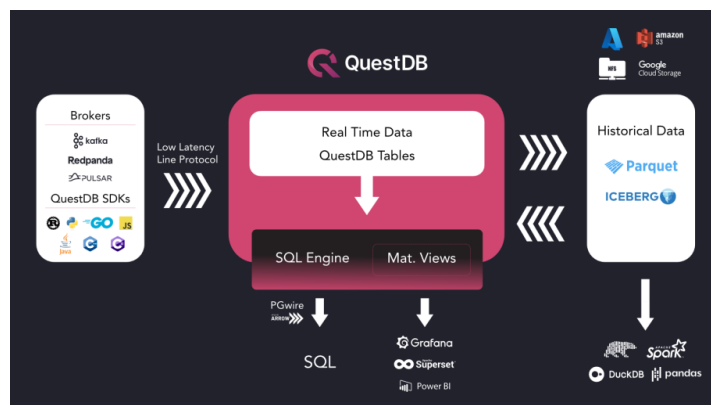
QuestDB یک دیتابیس بهینه شده برای داده های زمانی است. این پایگاه داده به ویژه برای پردازش و ذخیره سازی داده هایی که به ترتیب زمانی ثبت می شوند، مانند داده های مالی و بازارهای رمز ارز، طراحی شده است.

### چرا از QuestDB استفاده شده است؟

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

- **عملکرد بالا**: QuestDB برای پردازش داده های سری زمانی بهینه شده و می تواند حجم عظیمی از داده ها را با سرعت بسیار بالا وارد کوئری کند.
- **کاهش پیچیدگی معماری**: با استفاده از QuestDB دیگر نیازی به ترکیب چندین ابزار نیست.
- **مقیاس پذیری**: QuestDB در مدیریت حجم عظیمی از داده های سری زمانی مقیاس پذیری بالایی دارد، در حالی که Redis برای داده های بسیار بزرگ نیاز به حافظه بیشتر و تنظیمات پیچیده تری دارد.
- **پشتیبانی از ILP**: امکان ارسال داده ها با فرمت ILP باعث ساده تر شدن ارسال داده های بازار سهام می شود.

- **جستجو و انجام کوئری‌های پیشرفته SQL:** قابلیت استفاده از SQL استاندارد برای کوئری‌های پیچیده مانند محاسبه میانگین، محاسبه شاخص‌های معاملاتی (...، SMA, EMA) 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**



## Minikube

ابزاری برای شبیه‌سازی کلاسترهای Kubernetes به صورت محلی طراحی شده است. این ابزار به شما این امکان را می‌دهد که کلاسترهای توزیع شده را در محیط‌های محلی تست و توسعه دهید.

### چرا Minikube انتخاب شده است؟

- **شبیه‌سازی محیط‌های Kubernetes محلی:** Minikube به توسعه‌دهندگان این امکان را می‌دهد که بدون نیاز به محیط‌های ابری، کلاسترهای Kubernetes را در محیط محلی خود شبیه‌سازی کنند.
- **مدیریت کلاسترهای توزیع شده:** Minikube امکان مدیریت و تست آسان کلاسترهای توزیع شده را فراهم می‌آورد که برای تست و توسعه سیستم‌های بزرگ و پیچیده مانند این پروژه ضروری است.
- **صرفه‌جویی در هزینه‌ها:** استفاده از Minikube به جای ایجاد کلاسترهای Kubernetes در محیط‌های ابری می‌تواند به طور قابل توجهی هزینه‌های زیر ساختی را کاهش دهد.

