sibedge

Delivering Resilient Real-Time Data Streaming, at 600,000 Messages per Second

About Our Client

A leading financial data management company

Our client is a global enterprise crypto asset software and data provider, serving over 100 hedge funds. They work with traditional institutions, digital banks, exchanges, trading desks, wallet providers, and more. All of our clients' products adhere to institutional standards, such as AICPA Service and Organization Controls (SOC).

Challenge

Real-time data streaming has become indispensable in many business domains like banking, payment, customer service, manufacturing or even inventory management. Real-time data can help boost business agility, increase fraud detection, improve campaign performance, increase operational efficiency, improve supply tracking and enhance customer understanding. In our case, our client wanted to own real-time data without 3rd party platforms and create a solution which can stay resilient in case the predicted growth of all connected to the system exchange users on the horizon of 5 years.

The client already worked with 100+ exchange companies and required real-time aggregated data feeds, which enabled traders to respond to market changes as they occur and safeguard our client from unscrupulous behavior of fraudsters.

Company NDA

About company

A global enterprise crypto asset software and data provider, serving over 100 hedge funds.

Industry

Fintech

Team

Delivery Manager	1
Project Manager	1
Solution Architect	1
Business Analyst	1
DevOps Engineer	1
Backend Engineer	5
QA Automator	2

Duration

1 year 2 months

Technologies

- Spring
- Java

Challenge

Key Requirements from Our Client:

- Design and develop a high-performance real-time data streaming solution capable of handling 600,000 messages per second, with a guaranteed processing speed of less than 0.5 seconds.
- Build an architecture capable of integrating with 100+ trading and foreign exchange platforms like Binance.
- Ensure seamless integration with any exchange platform, including plans to integrate 100 of them in 2023 and assure cost-efficiency.

Solution

First stage: data transfer between exchanges and kafka

When discussing real-time data processing, Apache Kafka is the de facto standard. It is supported by numerous vendors. However, it's important to note that Kafka is not merely a messaging platform or an ingestion layer, as commonly perceived. While it facilitates real-time messaging at any scale, its storage component is equally crucial. Kafka allows storing all events within the event streaming platform, achieving true decoupling between producers and consumers. This capability enables automatic handling of backpressure, as consumers often struggle to handle the load from the producer side. Hence, our team decided to power the new solution with Apache Kafka.

The solution consists of two parts: data transfer between exchanges and Kafka, and aggregation, business logic, and the user interface (UI).

For each exchange, a WebSocket stream is established. The data is collected by a component called "Gate," which converts it into a unified format desired by our client for all exchanges. Subsequently, the data is transferred to Kafka.

During this stage, various problems may arise. Exchanges of different maturity levels may terminate socket connections without explanation or encounter other technical issues, potentially resulting in data loss. To ensure data integrity, our team implemented horizontal scaling of the Gate components.

Solution

Additionally, based on our knowledge base, we created duplicate components to receive two copies of the data simultaneously, minimizing the risk of loss. To combine these duplicates into a single stream, we introduced the Deduplicator component, which passes the data through itself and removes duplicates. The Gate component also incorporates a Data Uploader.

By September 2022, our team completed the proof-of-concept (POC). They developed custom low-level serializers, which improved system performance by more than 40%. Testing on hardware allowed processing up to 2.5+ million operations. The solution successfully passed load stress testing, leading our client to deem these performance indicators satisfactory.

The second stage: connect effortlessly to any exchange platforms like Binance

Our team created a code generator to streamline the editing of parsers for different exchanges, expediting the connection process. They also developed an emulator capable of simulating any exchange and wrote a custom data parser for the exchange, resulting in a performance boost of 40%.
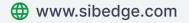
Furthermore, our solution implemented support for WebSocket (WS) and REST API to receive data from exchanges, as well as other interfaces such as FIX and Ethereum blockchain.

Result

Our team designed and developed a universal system that allows our client to receive data with maximum performance, eliminating data loss.

→ We achieved the target of processing 600,000 messages per second with a guaranteed processing speed of less than 0.5 seconds.

→ The Sibedge team assisted our client in effortlessly connecting to 18 exchanges while minimizing costs for adding new exchanges.

→ On average, each new exchange can be connected once every two weeks with the help of DevOps and Backend engineers, a significant cost drop.

This solution can also be applied to specific use cases in other industries.

🌐 www.sibedge.com   in www.linkedin.com/company/sibedge/

✉ contacts@sibedge.com