

Hyper-Scale Streaming Data Processing and Storage

PoC Overview

Impetus conceived and implemented a hyper-scale Stock Ticker application to demonstrate real time processing, storage and visualization of huge volumes of stock-tick data.

Client Overview

Impetus developed a Proof of Concept for financial service companies who face challenges in processing/storing large amounts of streaming data in real-time. The use case was applied to stock ticker data being streamed from over 80 different global exchanges and various sources.

Challenges

Enterprises generally need to compromise with running and maintaining multiple batch processes on the accumulated data due to throughput and management constraints. The overall business process turn-around can be improved if the data can be made available after processing in real-time.

A real-time system of such a large scale requires easy provisioning and monitoring. As a Proof of Concept, and in order to showcase how huge volumes of data can be processed and stored in real-time, Impetus developed a hyper-scale Stock Ticker application with the following goals:

- Receive simulated tick data from 80+ stock exchanges across the globe
- High data ingestion rate of the incoming ticks
- Real time analytics on the high volume of data
- Easy installation, configuration, and monitoring of the required infrastructure
- Real time dashboard showing key statistics of all components and ingestion rates

Highlights and Benefits

- *Real-time data processing on hyper scale and storage*
- *Provisioning, Management, and Monitoring of Cluster*
- *Visualization: Real-time graphs and reports on historical data*

Technologies

StreamAnalytix, Oracle NoSQL Database, Kafka, Intellicus, D3.js, Tomcat

Our Solution

The solution of the Stock Ticker Proof of Concept can be divided into six segments. Impetus team addressed client requirements in the following ways:

- **Data Generator:** A multi-threaded process that can generate mocked-up stock tick data from multiple exchanges around the globe, with the flexibility of increasing/decreasing data velocity.
- **Real-time Data Processing and Storage:** A parallel processing engine, which can process high volume of stock tick data and can publish it to the UI in real-time with high data processing rate.
- **Data Store:** A Data store that stores and makes available huge volume of incoming stock tick data in real-time.
- **Reporting on Historical Data:** A reporting tool that can generate reports on the historical data. The reports can be viewed on the UI for analysis and enabling business decisions.
- **User Interface:** A User Interface that can handle huge real-time data and is capable of showing it as graphs in real-time.
- **Provisioning and Monitoring Tool:** A Cluster Management tool.

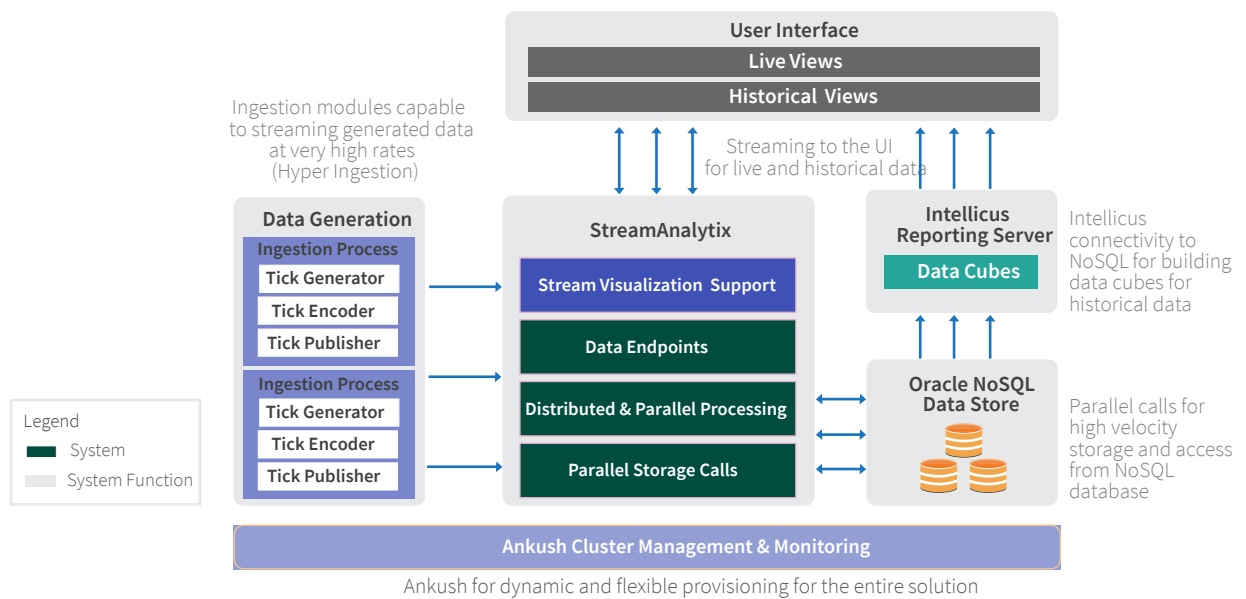


Figure 1: Stock Tick Application - High Level Flow Diagram