Designing A Reliable Real Time Feed Listener

Mid Semester Report

Submitted in partial fulfillment of the requirements of
BITSZG628T Thesis

By

Anupam Srivastava
ID No. 2018HT13250

Under the supervision of:

Chetan Dugar
&
Gaurab Basu

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, RAJASTHAN
September 2021

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, RAJASTHAN

Abstract

M.Tech. Software Systems

Designing A Reliable Real Time Feed Listener

by Anupam Srivastava

There is constant need to transfer data between different organizations, as well as the most common way to publish data over distributed systems employing multiple nodes over geographically different locations. It is not always feasible to publish or transfer this data over in batch as many times the data is time-critical. Servers publishing real time feeds are the most common way to solve this problem. Further constraints are put when the nature of data demands as low latency as possible for every client. Here in this dissertation we try to design a real time feed listener that is also able to communicate its data to multiple clients located at geographically different location. The system is highly available, highly consistent and have low latency.

Certificate

This is to certify that the thesis entitled, "*Designing A Reliable Real Time Feed Listener*" and submitted by Anupam Srivastava ID No. 2018HT13250 in partial fulfillment of the requirements of BITSZG628T Thesis embodies the work done by him under my supervision.

Supervisor
Chetan Dugar
Software Engineer III,
Tower Research Capital LLC, Gurgaon
Date:

Co-Supervisor
Gaurab Basu
Senior Software Engineer,
Tower Research Capital LLC, Gurgaon
Date:

Contents

List of Figures

Chapter 1

Introduction to Thesis Topic

Low latency is critical for a high frequency trading firm. Traders build complicated statistical models around the behaviour of market as reflected in ticker prices on exchanges, and execute their plan to buy or sell stocks accordingly. This is called devising a strategy. It is imperative that these strategies get triggered as early as possible to not only be profitable, but also to be ahead of competitive firms.

Refinitiv ™ is one such firm which publishes well known dataset known as Institutional Brokers' Estimate System or *I/B/E/S*. This dataset contains earning estimates of different U.S.A. companies. Even though companies change their names in due course of business, such as when they get acquired, these estimates are not interrupted since they were first published in 1976. Refinitiv publishes this data both as end-of-the-day archived file, as well as a real-time feed.

Tower has a paid subscription to this data and provides it to those traders who have created a strategy around it.

While some traders depend on real time feed, other traders are happy with the archived data. Thus, our design should be able to cater to both kind of users.

Chapter 2

Requirements and Estimates

Following requirements were captured for our full system:

2.1 Requirements

2.1.1 Functional requirements

- System should be able to download the data from Refinitiv's remote servers.
- System should be able to provide this data to different trading teams with equal priority.
- System should be able to store this data for later retrieval.
- System should make it easy to query this data.

2.1.2 Non-functional requirements

- System should be highly reliable. A unreliable system will lose the realtime feed being published by remote server.
- System should be consistent. The real time feed must be delivered in the same order as it was received.
- System should be moniterable.
- System should raise alerts upon failure.

2.2 Estimates

Refinitiv publishes around 3 to 5 MB per seconds of data, but occasionally the rate of delivery may rise upto 50 MB per seconds. Also, since the data is also published in archived format at the end of the day, we were able to estimate average storage requirement, which was around 1 GB per day.

A always-on system downloading 1GB per day will require 365 GB for a year's worth of data. Since we also backfilled the data for last 3 years, 1 TB of storage was estimated to be required. To make sure our system remains stable for next 3 years also, 2 TB of storage will be required. Furthermore, we estimate that a 60% full disk should raise an alert, a minimum of 3.5 TB of disk will be required.

Chapter 3

Design

3.1 Modules

Following modules were identified for development.

- Thin client that runs 24x6 and continuously listens and downloads the data from remote server.
- Kafka cluster to reliably distribute the data to multiple clients. Each trader gets a unique *group*, which ensures same data is delivered to everyone with equal priority.
- A consumer that subscribes to kafka topic and stores the data in persistent storage.
- An uploader program that reads that periodically uploads this data to a sql server.
- A monitor program that continously checks the health of the thin client and raises alerts
- A program called *Enricher* that can be used to read the raw data and republish it with any required changes.

3.1.1 Thin client

This client listens to Refinitiv's TIBCO-EMS queue, and immediately publishes the same data to a pre-defined kafka topic. Since Refinitiv expects the queue to be continuously emptied, it is important that the client remains highly available. To maintain low latency, C++ was used

3.1.2 Kafka topics

A pre-defined kafka topic is used to consume the raw data.

Another topic is used by the uploader.

Another topic is used by the enricher.

3.1.3 Enricher

This is used to read the raw data and convert to some other format, such as protobuf etc.

3.1.4 SQL

SQL database was used to maintain historical consistency. The database was partitioned based on quarters, as we found that most queries were made only for previous 30 days.

3.2 HLD

A high level architecture diagram of the proposition is shown in Figure 3.1

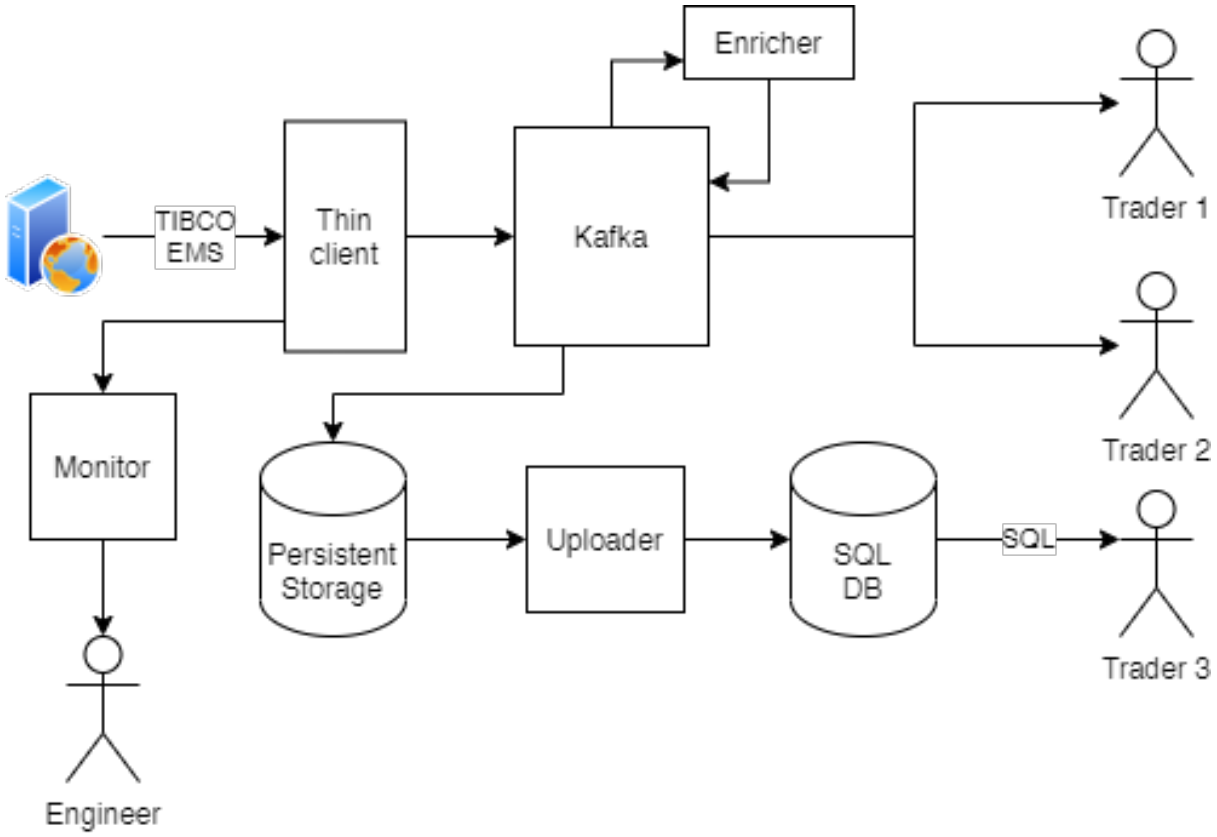

FIGURE 3.1: Architecture Diagram of proposed system

Appendix A

Institutional Brokers' Estimate System

The Institutional Brokers' Estimate System (I/B/E/S) is a service founded by the New York brokerage firm Lynch, Jones and Ryan and Technimetrics, Inc. I/B/E/S began collecting earnings estimates for U.S. companies around 1976 and used the raw data to calculate statistical time series for each company. The data subsequently was used as the basis for articles in academic finance journals attempting to demonstrate that changes in consensus earnings estimates could identify opportunities to capture excess returns in subsequent periods. After starting with annual earnings estimates and estimates of "Long Term Growth, the database later was expanded to include quarterly earnings estimates. This allowed for the analysis of "Quarterly Earnings Surprises." Other innovations made possible by the I/B/E/S data included estimates for various equity indexes on a "top down" basis (made by strategists and economists) and estimates made on a "bottom up" basis (by individual analysts) for those same indexes. In the mid-1980s I/B/E/S began to expand its dataset to include companies trading in international markets. Lynch, Jones was sold to Citigroup in 1986. Barra bought I/B/E/S in 1993, selling it to Primark two years later. Thomson Financial purchased Primark in 2000. Successor companyThomson Reuters spun off its financial division under the name Refinitiv in 2018.

The I/B/E/S database currently covers over 40,000 companies in 70 markets. It provides to a client base of 50,000 institutional money managers. More than 900 firms contribute data to I/B/E/S, from the largest global houses to regional and local brokers, with US data back to 1976 and international data back to 1987.