Kafka and Spark Streaming for Analyzing Stocks, Currencies, and Crypto Data

Julián Benítez Gutiérrez

Escuela Colombiana de Ingeniería Julio Garavito
julian.benitez@mail.escuelaing.edu.co

October 20, 2022

Abstract

This article is a proposal for the Functional Programming in Distributed Systems course final project. Using a free-tier API for financial data, which provides information on Stocks, Currencies and Crypto, a flow of data using Kafka and Spark Streaming will be used to analyze the information providing useful insight and services for the user's needs.

I. Introduction

The planet has become data-driven and the finance world is no exception. With many open source projects and free to use tools, analyzing the financial market has never been easier. The globe is currently changing at a fast pace, it will be a great opportunity for using these tools that will enable us to have a better understanding of the continuous changing financial market.

II. Proposal

The project will use Finnhub as its main data source (free-tier), providing real-time RESTful APIs and Websockets[Finnhub, 2022].

- The producer component will use the Websocket and REST endpoints to obtain the data and send it through a Kafka topic.
- The consumer will receive the information from the Kafka topic and using Spark Streaming it will perform the pertinent data analysis.
- The consumer could provide an API for any external components to use, example a database service, email service or even a CLI service.

The next image is an architecture proposal for the previous explained solution.

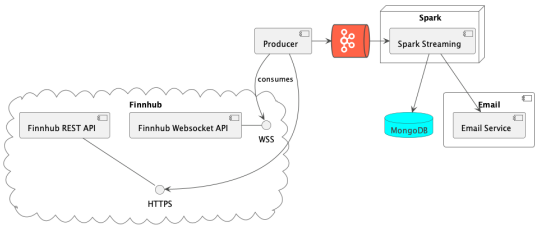

Figure 1: *Architecture Proposal*

III. Expected Results

The project should be able to:

- Be easily deployed in the preferred cloud solution.
- Provide information to a Kafka topic at a relatively fast pace using the Finnhub API.
- Use Spark Streaming for consuming the data from the Kafka topic.
- Provide insightful analysis of the financial market.

- Provide an useful API so that any service can use it at will.
- The service in question will be useful for the user's needs.

References

[Finnhub, 2022] Finnhub Stock API