

San José State University
CMPE 266 — Big Data Engineering and Analytics



SAN JOSÉ STATE
UNIVERSITY

Project Milestone I

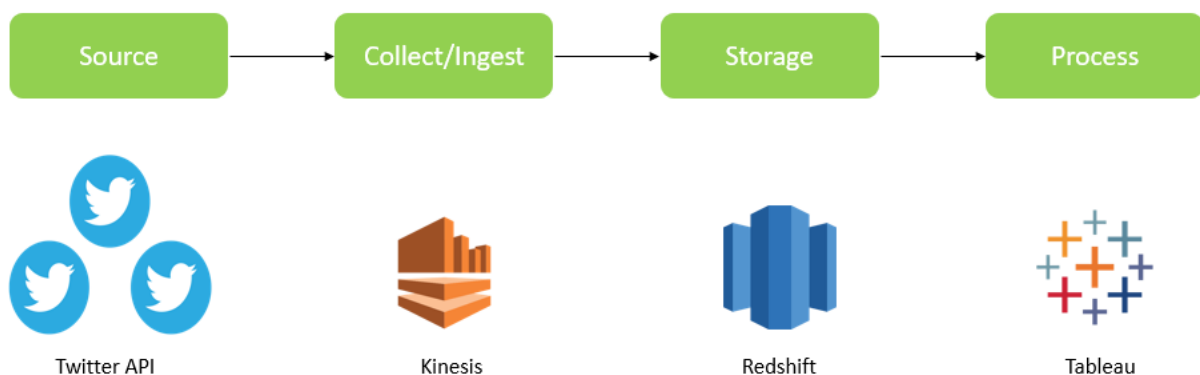
Created By:

| | |
|--------------------------|-----------|
| Aakash Alurkar | 013716729 |
| Aniket Deshpande | 013532519 |
| Madhu Prasanna Kalluraya | 013708071 |
| Shivani Mangal | 012530362 |
| Zainab Khan | 010120812 |

STOCK PRICE PERFORMANCE PERCEPTION USING TWITTER STREAMS

WHAT IS THE PROJECT IDEA /WHAT THE BIG DATA APPLICATION DOES?

As per the random walk theory, “changes in stock prices have the same distribution and are independent of each other, therefore, the past movement or trend of a stock price or market cannot be used to predict its future movement”. However, it is also known that human emotions deeply affect decision making and an individual’s behaviour. This forms one of the prime concepts of behavioural economics. Thus, collective mood swings of humans are bound to affect stock market. This is because our emotions govern how we feel; and how we feel, affects how we behave as an investor. Our emotions help in determining if our decisions are beneficial or harmful. The raw indicators of swinging emotions in today’s world are twitter feeds, blog posts and Reddit feed. Through news articles, one can analyze the emotion of the masses. So can twitter feed be used to support in making decisions for the stock market? Through this project we are aim to analyze the same.



In one of his speeches, Abraham Lincoln once said, “In this age, in this country, public sentiment is everything. With it, nothing can fail; against it, nothing can

succeed. Whoever molds public sentiment goes deeper than he who enacts statutes, or pronounces judicial decisions." Social Media, in the form of twitter and facebook feeds is holding this true in the 21st Century as well. A skilled NLP analyst can mine patterns out of tweets to make conclusions about how the market in general feels about a stock, a company or a brand. John Bollen, in one of his research papers, claimed that Twitter data could predict the Dow Jones Industrial Average with 87.6% accuracy (Bollen, 2010). As per an article by MIT, "an analysis of almost 10 million tweets from 2008 shows how they can be used to predict stock market movements up to 6 days in advance."

In recent times we have seen tweets that turned the market completely upside down, thus cementing the above mentioned theories. When Elon Musk tweeted about taking Tesla private and it's funding being secured, Tesla shares ended at 11%.

Our Big Data application will analyze incoming twitter feed to gauge the market sentiment for particular brand. It will also capture the live stock market feed to check the correlation between market prices and sentiment. The user will be provided with both the information in form of a visualization dashboard to make an informed decision when investing in the market.

WHAT ARE THE TECHNOLOGIES USED?

To ingest the real-time twitter data Twitter Developers API provides the following parameters.

Standard streaming API request parameters

delimited
stall_warnings
filter_level
language
follow
track
locations
count

Any or all of these parameters can be used to ingest the real-time Twitter Streaming Data. After connecting to the API using an HTTP response we can parse the real-time incoming data.

To ensure the real-time performance of the tweet-stream, we will implement ingesting the feed using AWS KDS(Kinesis Data Streams) which provides the infrastructure to handle such a huge volume of big data in the cloud at near real-time speeds which is cost effective and durable.



For storing the data we will use Amazon S3 which will work with Kinesis Data Streams in tandem and provide the appropriate solution to store and query data which is volatile in nature (hot data).

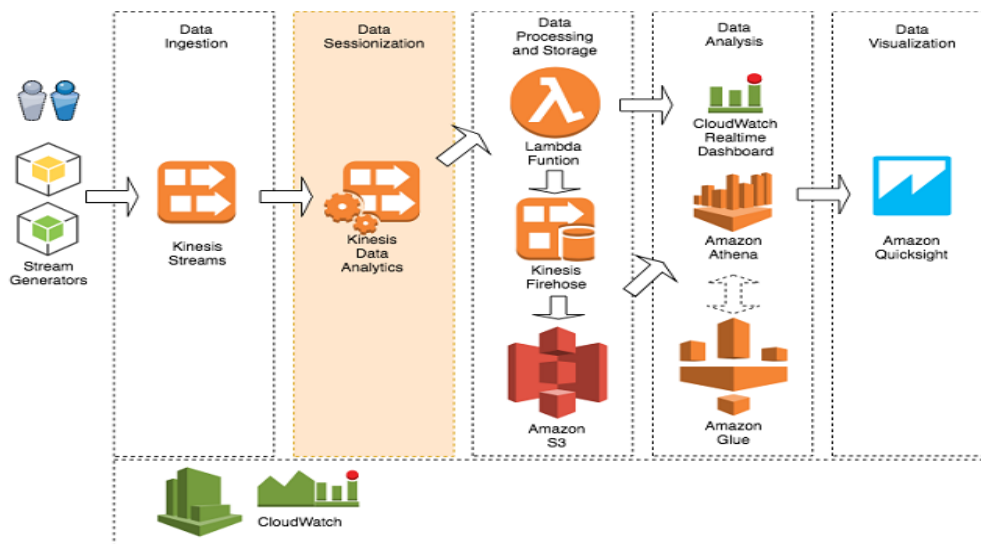
For the Sentiment Analysis module Amazon Translate and Amazon Comprehend will be implemented. They will perform comprehension and NLP on the real time Twitter data.



The Amazon Translate and Amazon Comprehend will work inside Amazon Lambda

Amazon Athena will quickly enable us to leverage and query the volume of data quickly stored in S3 with the help of AWS Glue crawler for optimization.

AWS Quicksight will enable rapid prototyping of dashboards for dynamic visualizations and insights in the real time data.



FEATURES LIST

- **Ingestion of live streaming data**

The streaming data is ingested continuously from Twitter. This allows for near-real-time analytics being displayed to the user. The user can then act based on this data such as selling or buying stock.

- **Storage of data**

The incoming data from the data source will be stored. This is an intermediate dataset because the data needs to be processed. Data transformations and cleaning be applied to this raw intermediately dataset.

- **Machine learning**

Sentiment analysis on the processed data will be done using NLP (Natural Language Processing).

- **Near-Real-Time Analytics**

Since the data being used is live streaming data, it will give the user near-real-time analytics.

- **Dashboard for visualization**

There will be a dashboard provided for visualization. Users can also run queries via RedShift to extract additional insights.

SOLUTION FOR EACH REQUIREMENT PRESENT IN THE PROBLEM STATEMENT

Gauging the market with live Twitter analysis:

The live Twitter stream will be ingested with Amazon Kinesis for stream processing. Further processing on streaming data will be done with Amazon Lambda and Amazon Translate and Amazon Comprehend which will extract valuable insight from the Twitter stream.

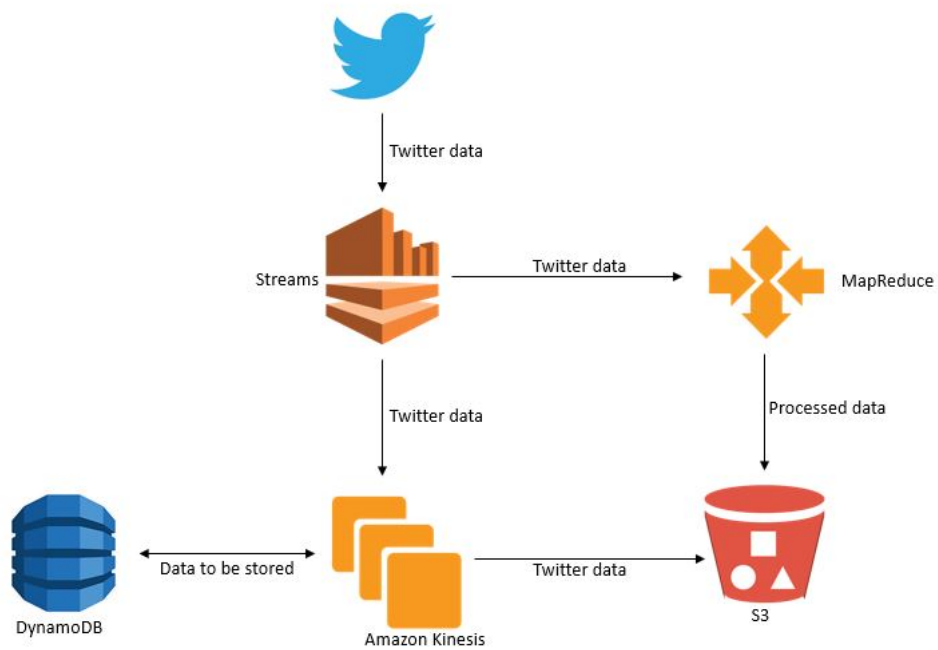
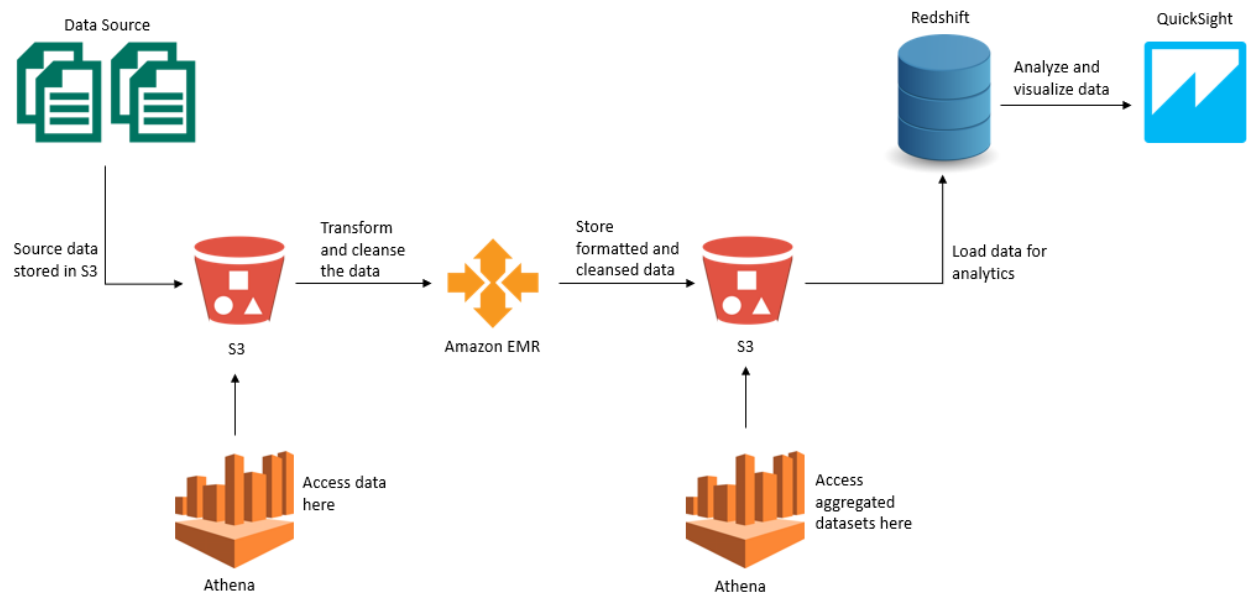
Gauging the market with live stock market analysis:

Near real-time analytics will be performed on the live streaming stock market data with the help of Amazon Kinesis stream ingestion and Amazon NLP features such as Amazon Comprehend and Amazon Translate.

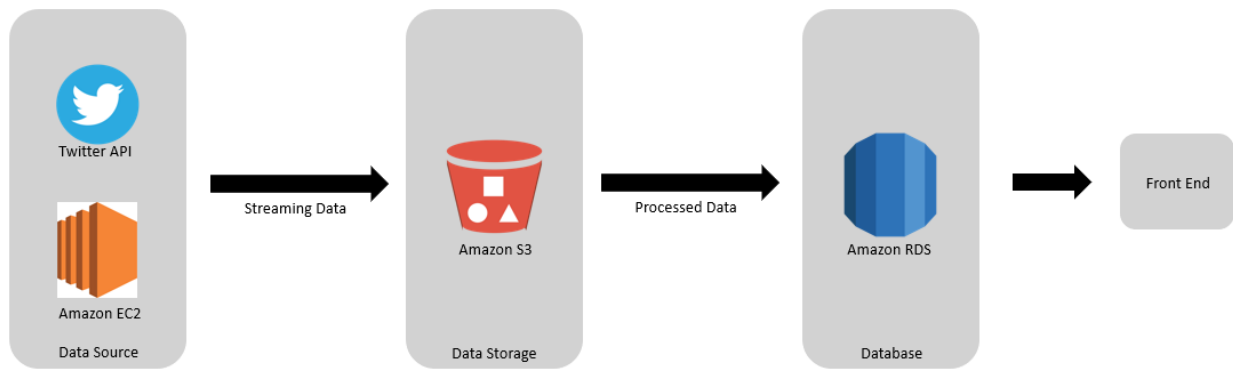
Combined market insight from both Twitter and the stock market

Amazon QuickSight's business intelligence features will be used to derive and visualize meaning from the collective data of Twitter and the stock market, which will help the users to make real-time and well-informed decisions.

SOLUTION ARCHITECTURE DIAGRAM



Using AWS for storing and analysing Twitter streaming data



Twitter Streaming Process

URL TO GITHUB CODE REPO

<https://github.com/aakashalurkar/cmpe266-project>