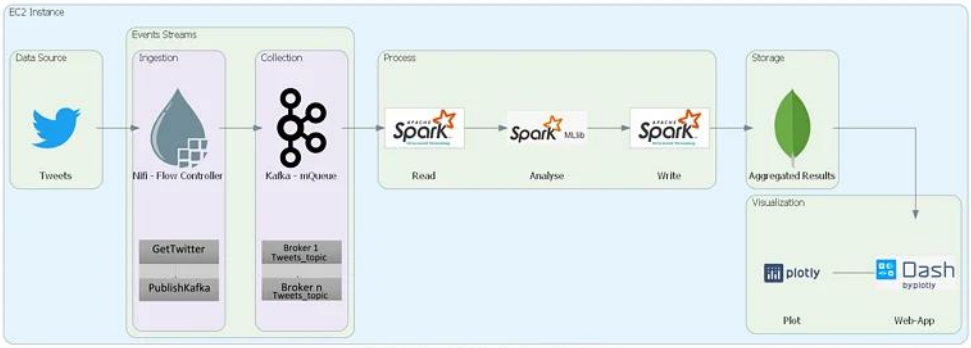
I have completed the analysis as per below:

1) Live Twitter Sentiment Analysis with Spark

When it comes to influencing purchase decisions or finding people's sentiment for a political party, people's opinion is more important than traditional media. That means there is a significant opportunity for brands on Twitter. Twitter sentiment is a term used to define the analysis of sentiments in the tweets posted by the users. Generally, Twitter sentiment is analyzed in most big data projects using parsing. Analyzing users' sentiments on Twitter is fruitful to companies for their product that is mostly focused on social media trends, users sentiments, and future views of the online community.

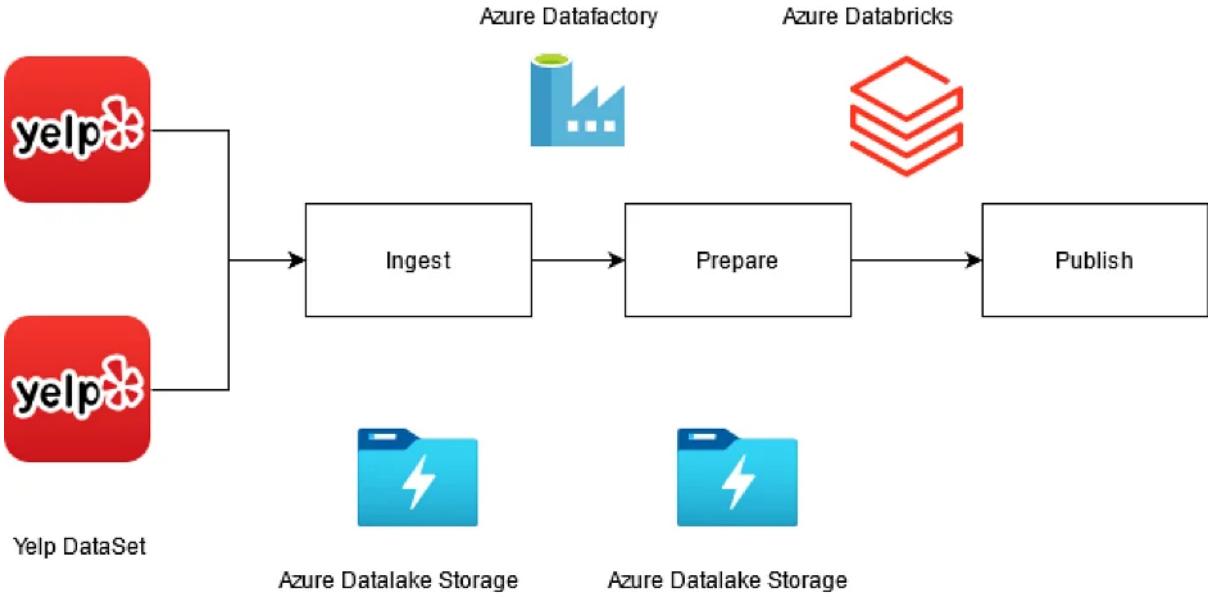


Live DashBoard - Twitter Sentiment Analysis

The data pipeline for this data engineering project has five stages - data ingestion, NiFi GetTwitter processor that gets real-time tweets from Twitter and ingests them into a messaging queue. Collection happens in the Kafka topic. The real-time data will be processed using Spark structured streaming API and analyzed using Spark MLib to get the sentiment of every tweet. MongoDB stores the processed and aggregated results. These results are then visualized in interactive dashboards using Python's Plotly and Dash libraries.

2.Yelp Data Analysis using Azure Databricks

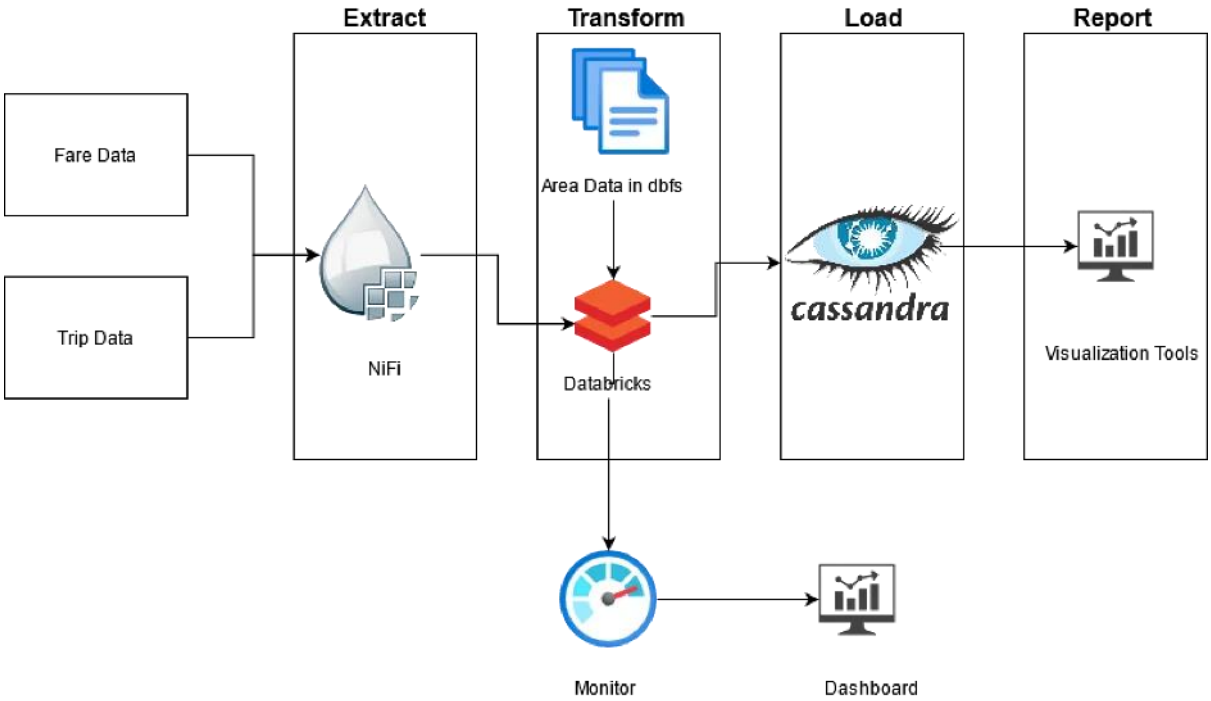
Yelp dataset consists of data about Yelp's businesses, user reviews, and other data made publicly available for personal, educational, and academic purposes. Available as JSON files, use it to learn NLP for sample production data. This dataset contains 6,685,900 reviews, 192,609 businesses, 200,000 pictures in 10 metropolitan areas. This Azure project helps you understand the ETL process i.e, how to ingest the dataset, clean it and transform it to get business insights. Also, you get a chance to explore Azure Databricks, Data Factory, and Storage services.



There are three stages in this real world data engineering project. Data ingestion: In this stage, you get data from Yelp and push the data to Azure Data lake using DataFactory. The second stage is data preparation. Here data cleaning and analysis happens using Databricks. The final step is Publish. In this stage, whatever insights we drew from the raw Yelp data will be visualized using Databricks.

3.Realtime Data Analytics with Databricks - Olber Cab Service

A Cab service company called Olber collects data about each cab trip. Per trip, two different devices generate additional data. The Cab meter sends information about each trip — the duration, distance, and pick-up and drop-off locations. A mobile application accepts payments from customers and sends data about fares. The Cab company wants to calculate the average tip per KM driven, in real-time, for each area to spot passenger trends.



This architecture diagram demonstrates an end-to-end stream processing pipeline. This type of pipeline has four stages: extract, transform, load, and report. In this reference architecture, the pipeline extracts data from two sources, performs a join on related records from each stream, enriches the result, and calculates an average in real-time. The results are stored for further analysis.