Data Pipeline
To Track the Sentiment of
the Twitter Conversation Surrounding
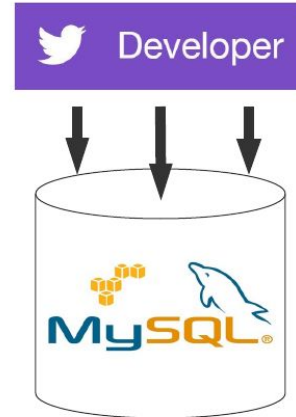# the Covid-19 Vaccine in the United States
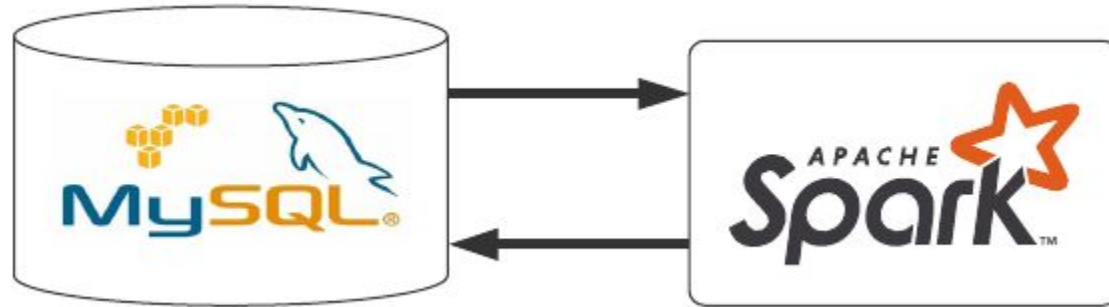
Lee Givhan
Anusha Jangalapalli
Suma Konjeti

Data Pipeline is created with Apache Airflow that automates programs written in Jupyter Notebook.

Relevant tweets are streamed from Twitter's API into an AWS MySQL database.

"Covid-19 Vaccine"

The data gets cleaned with Apache Spark,
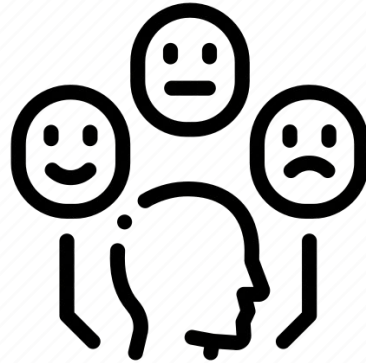then returned to the MySQL database

Natural Language Analysis in Python NLTK scans the words in each tweet, deciding whether it is positive, negative or neutral.

Natural Language Analysis
with Python NLTK

Sentiment information is added to the dataset, then the dataset is divided four ways based on region to generate data visualizations with Matplotlib and Word Cloud and Tableau.

Sentiment Analysis for USA by Region

Negative
Neutral
Positive

USA REGIONS