**Midterm Project**

**Starting Apache Kafka**

**Getting the news from News API**

**Sending news through the producer**

**Receiving the news to the consumer and saving it into HDFS**

**Setting up the Hive database and table**

**Loading the data generated**

**Insights**

**Top 5 sources by news count**

**Top 5 used words in the news**

**Top 5 used words per source**