Tweets realtime nlp analysis

store documents (tweets) and spark computing results

data ETL

twitter

kafka

MONGO DB

Zeppelin

datavisualization (notebook) (real-time analytics)

data source

Spark Streaming + Spark NLP

data streaming processing

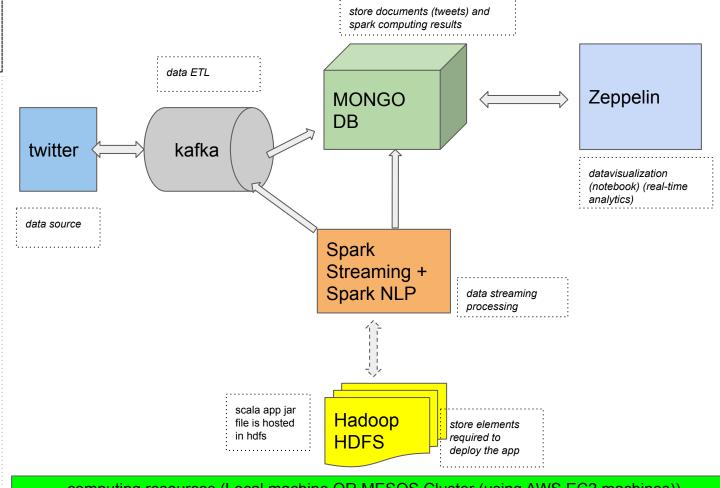scala app jar file is hosted in hdfs

Hadoop HDFS

store elements required to deploy the app

computing resources (Local machine OR MESOS Cluster (using AWS EC2 machines))