**Real-Time Presidential Election Processing**

\*tried to run the project on the centos 6.5 machine but streamlit wouldn’t work with it so worked on RHEL 9.4 and installed the tools on it

**Project Workflow**

\*Starting Hadoop HDFS Kafka and Zookeeper.

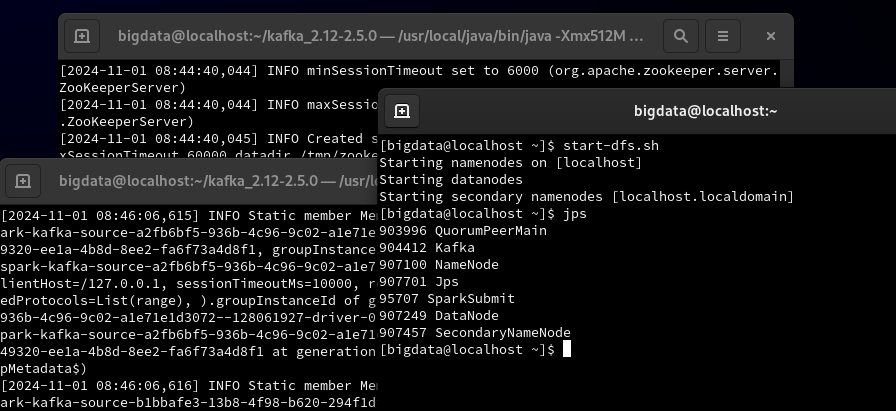
Start-dfs.sh # to start HDFS

cd $KAFKA\_HOME # entering kafka path

bin/zookeeper-server-start.sh config/zookeeper.properties # starting zookeeper

cd $KAFKA\_HOME # entering kafka path

bin/kafka-server-start.sh config/server.properties # starting kafka

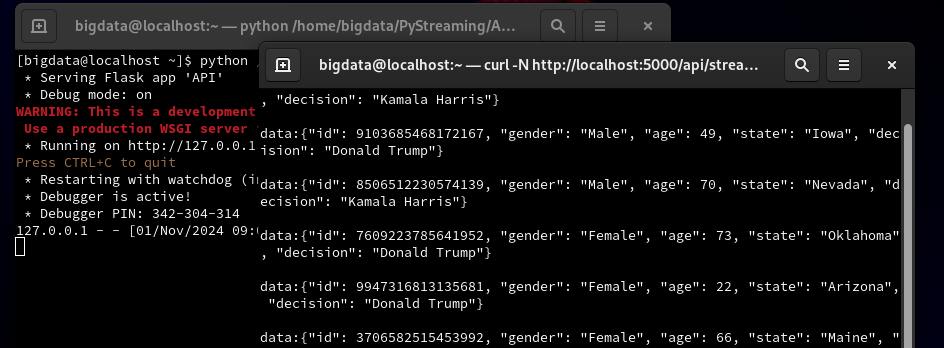


\***Set up the API** to begin data generation.

included the API code in the project files as API.py

python /home/bigdata/PyStreaming/API.py # run the API code

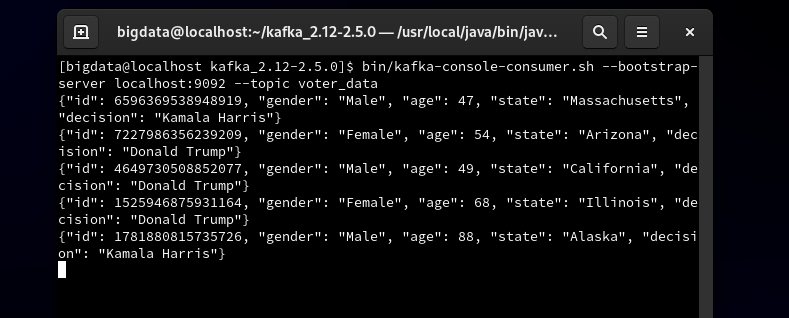
curl -N <http://localhost:5000/api/stream_votes> # start the API to continually generate data



\*setup kafka consumer

cd $KAFKA\_HOME

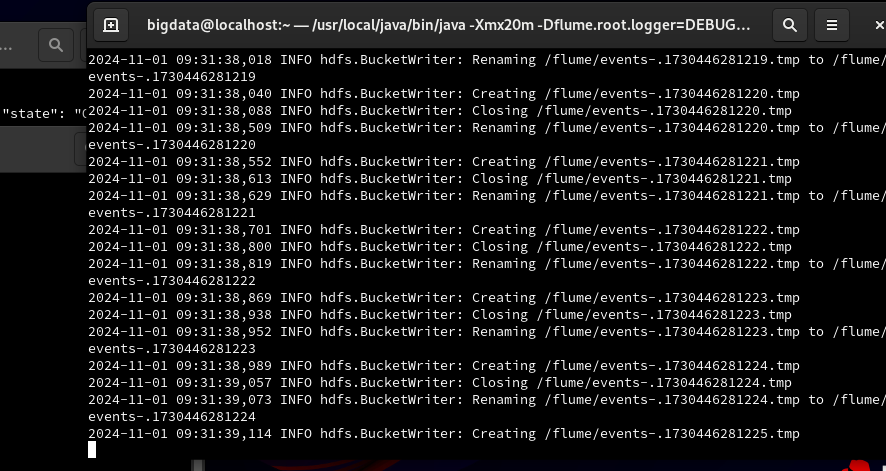
bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic voter\_data

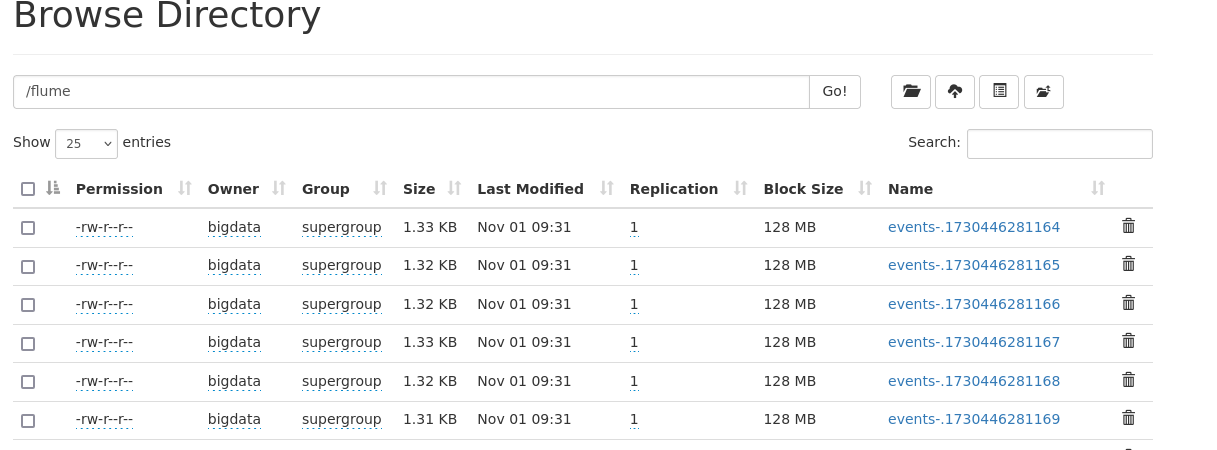


\*Runining the flume agent to take the data from kafka to HDFS

The flume agent conf file can be found in the project files as HdfsSink.conf

$FLUME\_HOME/bin/flume-ng agent --conf conf --conf-file $FLUME\_HOME/conf/HdfsSink.conf --name agent -Dflume.root.logger=DEBUG,console

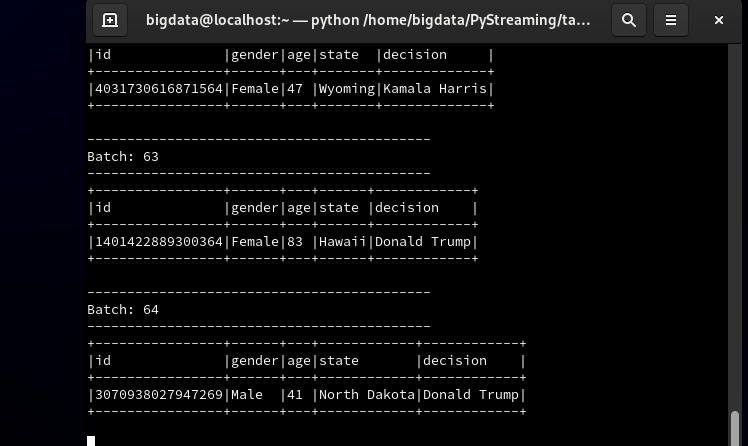




\*Processing the data with PySpark

This code can be found in the project files as table.py

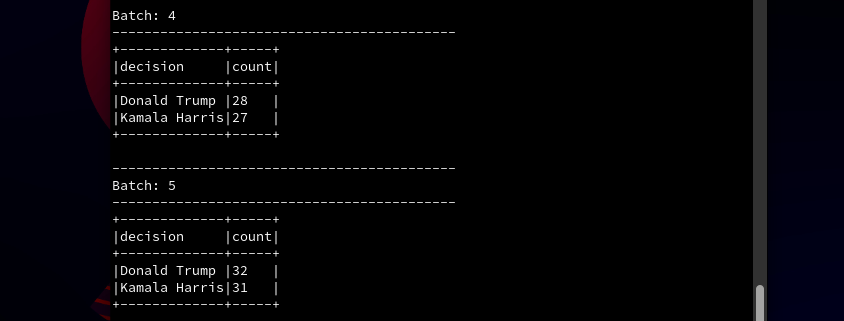
python /home/bigdata/PyStreaming/table.py



The total number of votes for each candidate

This code can be found in the project files as winning.py

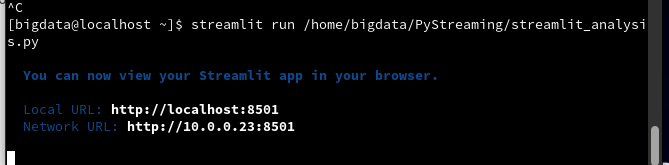
python /home/bigdata/PyStreaming/winning.py

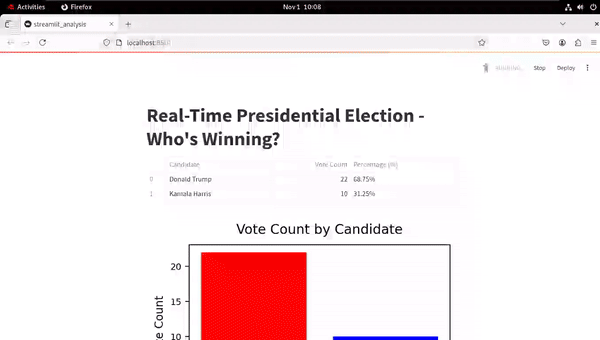


\*streaming the data to streamlit and creating visualization

This code can be found in the project files as streamlit\_analysis.py

streamlit run /home/bigdata/PyStreaming/streamlit\_analysis.py





\*Commands helped me to run the project

spark-submit --packages org.apache.spark:spark-sql-kafka-0-10\_2.12:3.2.0

pip install kafka-python

pip install setuptools wheelspark

pip install --upgrade pip setuptools

pip install pyarrow --only-binary=:all:

pip install streamlit