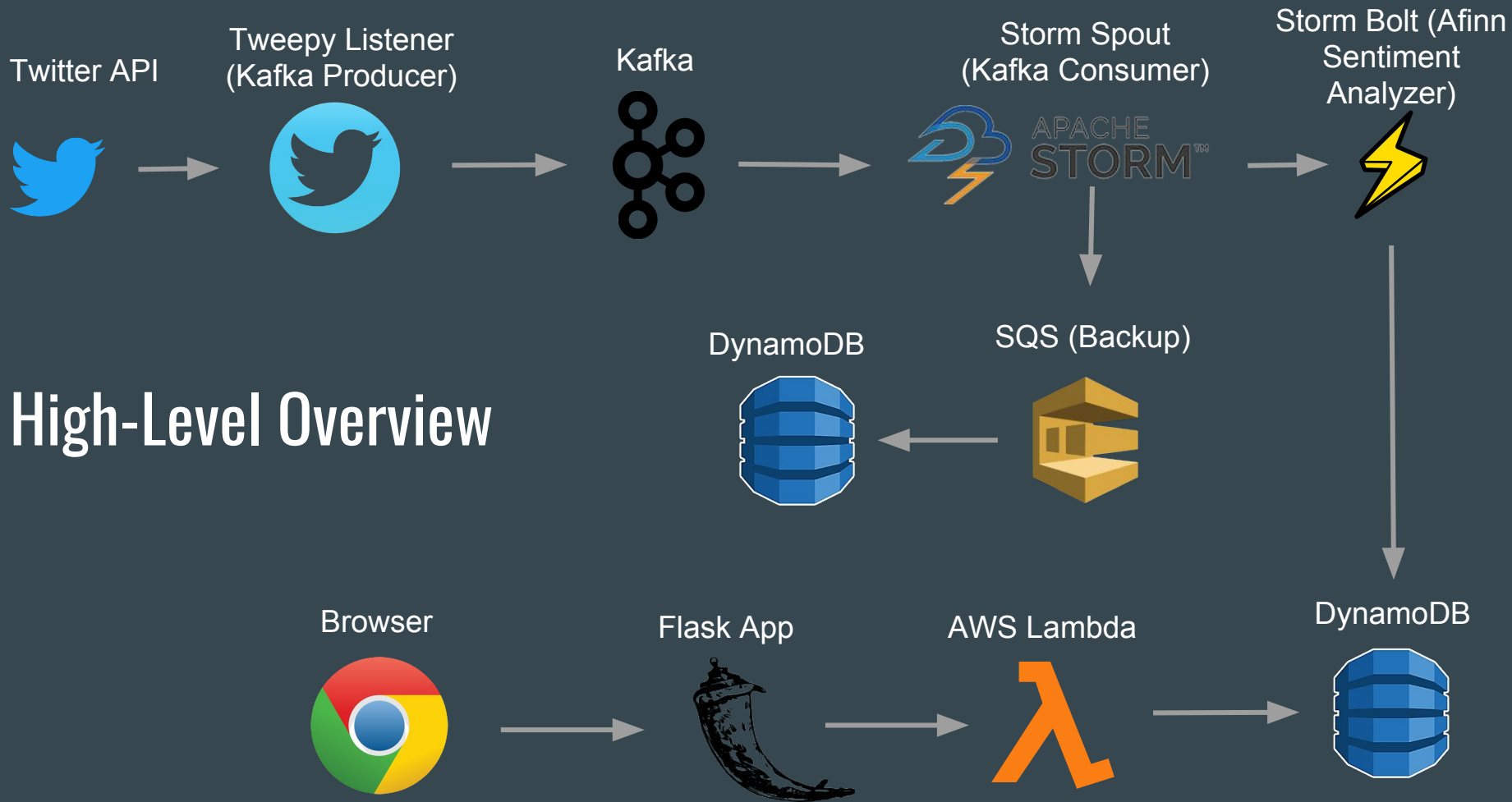


Sentiment Analysis on Donald Trump

...

Removed Names of Students



Purpose & Problem Statement

- Improve Lab 8 framework
- Improved Components
 - Sentiment Analysis - Afinn
 - User Interface - D3 Visualization
 - Rest Interface - Lambda Functionality
 - Backup Queuing - SQS
- Analyze President Trump's sentiment on Twitter
- Use Case: Data presented is post Trump speech on Palestine & Israel conflict
 - Date: Thursday, December 07, 2017

Storm Topology and Ambari Server

Storm UI

Search mytopology-33-1512684622: Search Search Archived Logs:

Topology summary

Name	Id	Owner	Status	Uptime	Num workers	Num executors	Num tasks	Replication count	Assigned Mem (MB)	Scheduler Info
mytopology	mytopology-33-1512684622	storm	ACTIVE	10s	2	17	17	1	0	

Topology actions

[Activate](#) [Deactivate](#) [Rebalance](#) [Kill](#) [Debug](#) [Stop Debug](#) [Change Log Level](#)

Topology stats

Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed
All time			0		

Spouts (All time)

Id	Executors	Tasks	Emitted	Transferred	Complete latency (ms)	Acked	Failed	Error Host	Error Port	Last error	Error Time
spout	5	5	0	0	0.000	0	0				

Showing 1 to 1 of 1 entries

Bolts (All time)

Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed	Error Host	Error Port	Last error	Error Time
----	-----------	-------	---------	-------------	---------------------	----------------------	----------	----------------------	-------	--------	------------	------------	------------	------------

Ambari mycluster

Dashboard Services Hosts Alerts Admin admin

HDFS

- YARN
- MapReduce2
- ZooKeeper
- Storm
- Ambari Metrics
- Kafka
- SmartSense

Summary

No alerts

NameNode Started **No alerts**

SecondaryNameNode Started **No alerts**

DataNode 1/1 Started

JournalNodes 1 live / 0 dead / 0 decommissioning

JournalNodes 0/0 JournalNodes Live

DFSGateway 0/0 Started

NameNode Uptime 1.05 days

NameNode Heap 77.5 MB / 1011.3 MB (7.7% used)

Disk Usage (DFS Used) 254.4 MB / 16.9 GB (1.18%)

Disk Usage (Non DFS Used) 7.3 GB / 16.9 GB (43.12%)

Disk Remaining 9.4 GB / 16.9 GB (55.70%)

Blocks (total) 9

Block Errors 0 corrupt replica / 0 missing / 9 under replicated

Total Files + Directories 40

Upgrade Status No pending upgrade

Safe Mode Status Not in safe mode

Metrics

NameNode GC

NameNode GC time 1 ms

NN Connection Load 2

NameNode Heap 1000 MB

NameNode Host Load 10.1%

NameNode RPC 1.1%

Failed disk volumes 0

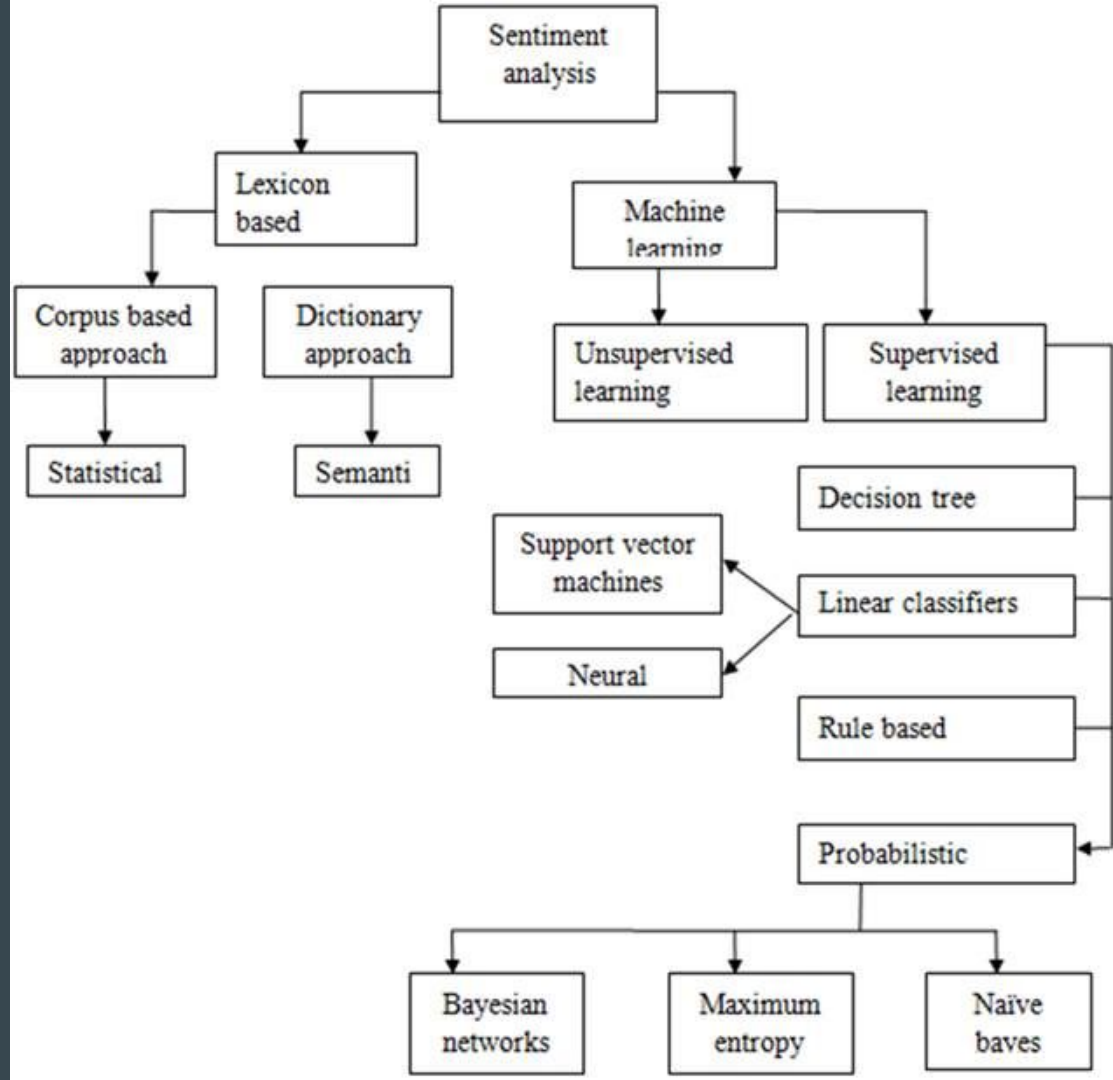
Blocks With Corrupted Replicas 0

Under Replicated Blocks 9

HDFS Space Utilization 1%

Improved Sentiment Analysis

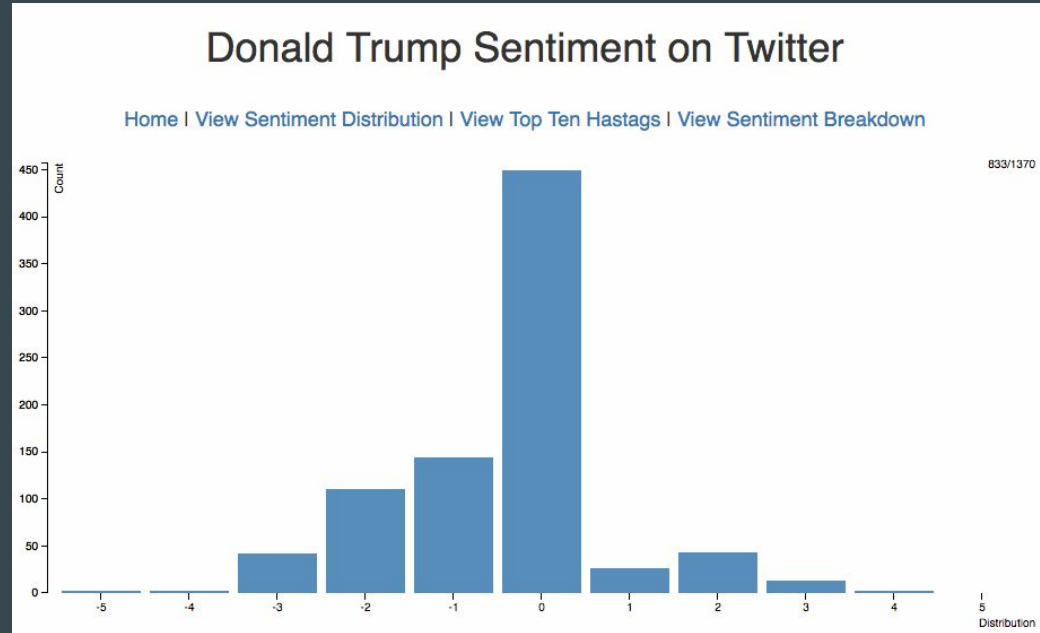
- Lab 8
 - Uses list of 2477 scored words for sentiment analysis
- Final Project
 - Uses Afinn Python library and additional input configurations for advanced sentiment analysis
 - List of 3382 scored words
 - Scores emoticons



D3 Visualizations

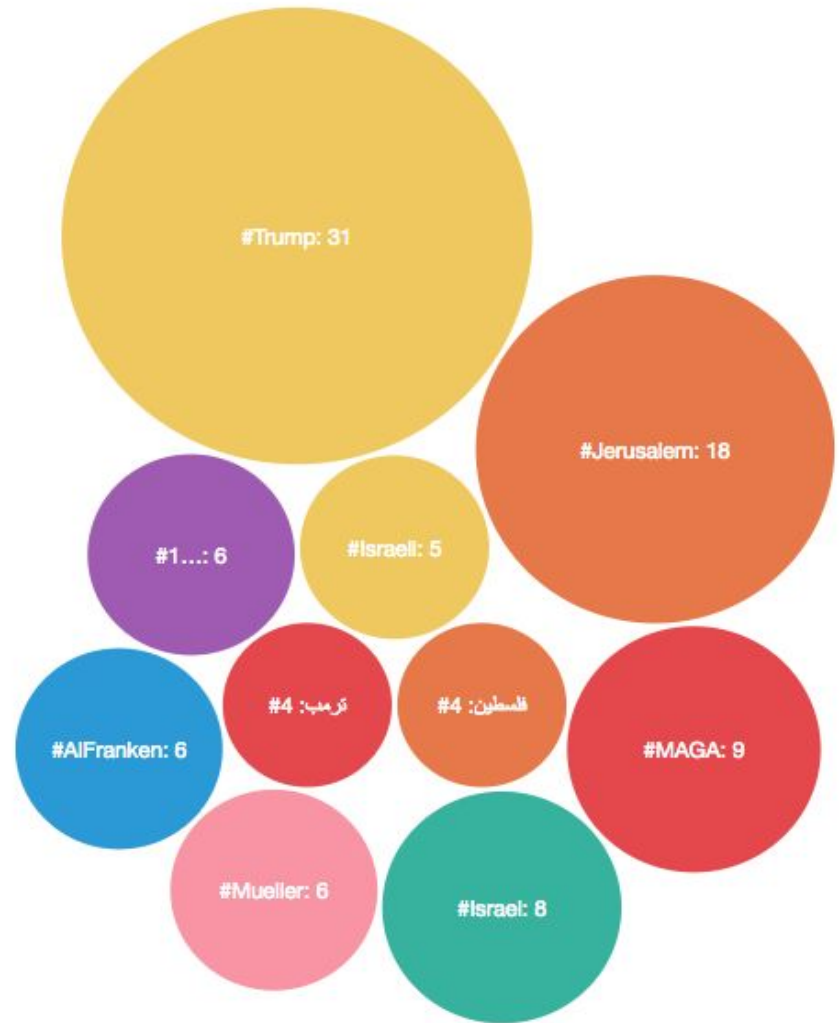
Normalized Distribution Histogram

- Normalization Algorithm
 - Sentiment scores are normalized between -5 and 5
- Utilizes AWS Lambda to obtain information
- Visualizes using D3 histogram and animations
 - Real-time histogram creation



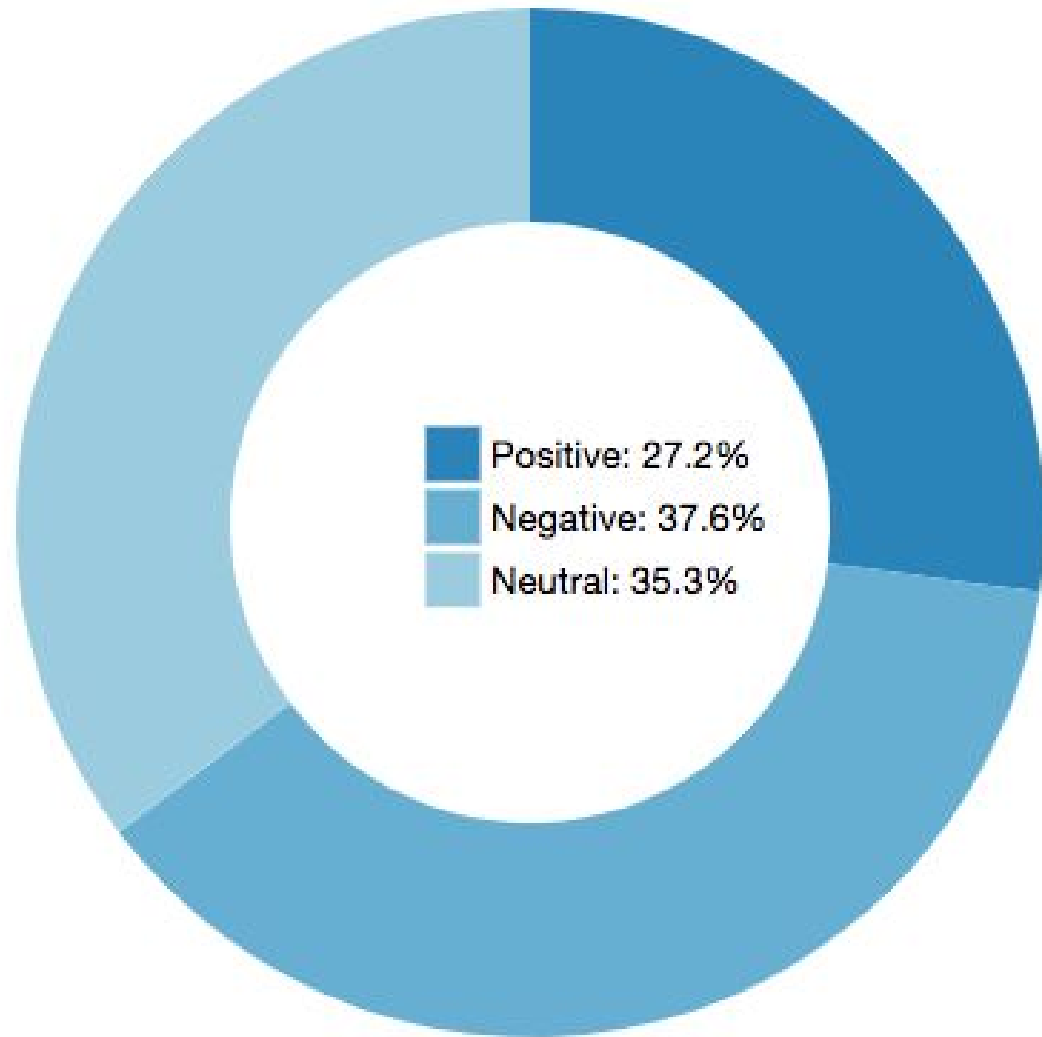
Top Ten Hastags

- Map Reduction for Top Ten Hashtags
- Utilizes Lambda to obtain data
- Visualizes with D3 Bubble Chart



Sentiment Breakdown

- Percentage breakdown of Tweets
 - Positive: Greater than 0.0
 - Neutral: Equal to 0.0
 - Negative: Less than 0.0



Project Evaluation

Test Results - Post Speech

- Sentiment is mostly neutral/negative
- Trending tweets are related to the current Palestine/Israel conflicts

Source Code

- <https://github.com/stripuramallu3/ECE4813FinalProject-TrumpAnalysis>

Obstacles and Tradeoffs

- Sentiment on multiple languages
- Issues with Storm topology
- Initial Machine Learning sentiment analyzer not integrating (NLTK)
- Initial topic was Net Neutrality
 - Did not have depth
 - Many repeated tweets (bots)
 - Inappropriate tweets

Breakdown of Responsibilities

Tweepy Integration	AWS Lambda Functions	AWS DynamoDB
Storm/Kafka Integration	Sentiment Analysis Algorithm	Sentiment Analysis Algorithm
Flask Service	Normalization Algorithm	D3 Visualization
D3 Visualizations	D3 Visualizations	