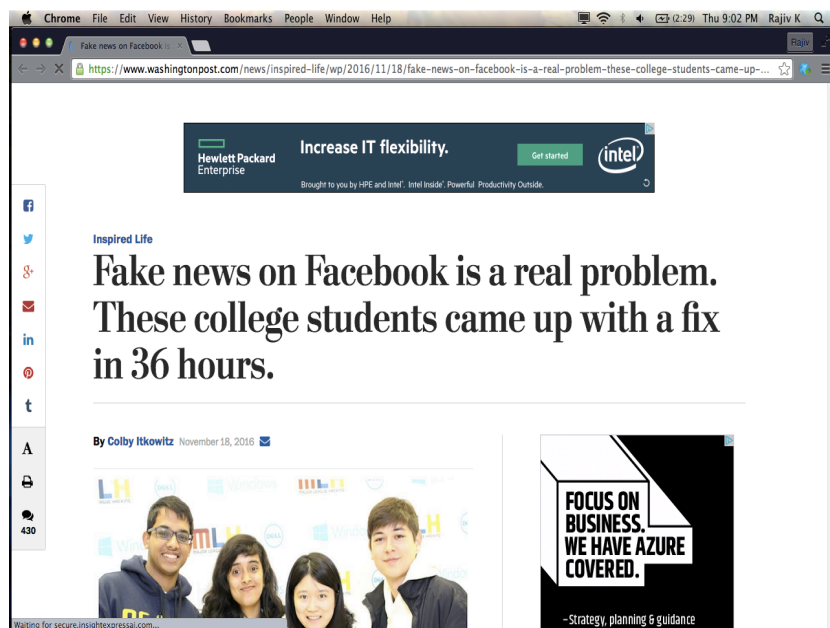




Fake News Detection: Using Apache Spark and AWS

Rajiv Kurapati
Cloudwick

Inspiration




Source : https://www.washingtonpost.com/news/inspired-life/wp/2016/11/18/fake-news-on-facebook-is-a-real-problem-these-college-students-came-up-with-a-fix/?utm_term=.ee0bcec715fa

- What can I do?
- Can I use Big Data ?

Clouwick

What is Fake News?

dis·in·for·ma·tion

/ˌdɪsənˈfərˈmɑːʃ(ə)n/ 

noun

false information that is intended to mislead, especially propaganda issued by a government organization to a rival power or the media.

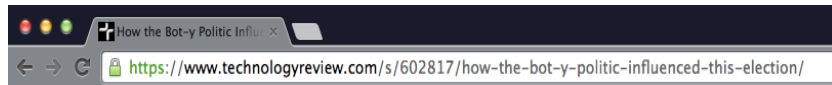
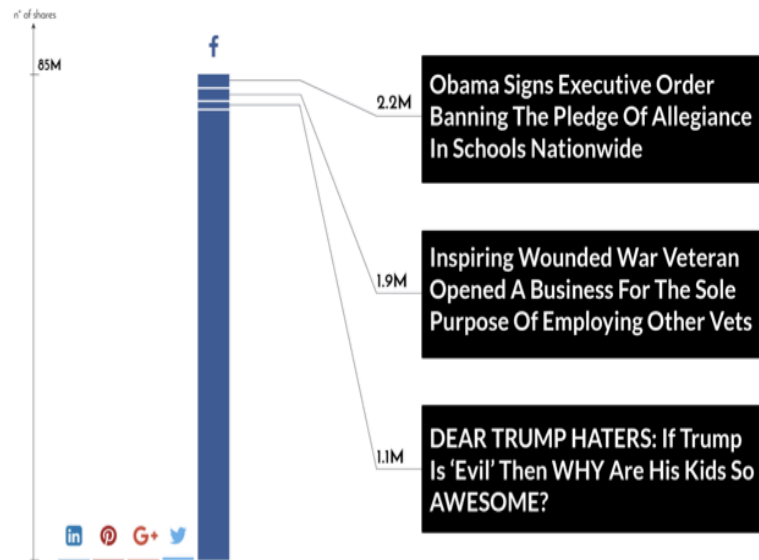


Translations, word origin, and more definitions

- Platforms Spreading Fake News
- Who is responsible for this

Impact of Disinformation

Most shared fake articles - English



MIT
Technology
Review

Topics+ Top Stories

Business Impact

How the Bot-y Politic Influenced This Election

Nearly 20 percent of all election-related tweets come from an army of influential robots.

by Nanette Byrnes November 8, 2016

Clouduick

Impact of Disinformation



5 Top Fake Political News Stories On Facebook

"Obama Signs Executive Order Banning The Pledge Of Allegiance In Schools Nationwide" [ABCNews.com.co](#)

2,177,000 Facebook shares, comments, and reactions

"Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement" [Ending the Fed](#)

961,000

"Trump Offering Free One-Way Tickets to Africa & Mexico for Those Who Wanna Leave America" [tmzhiphop.com](#)

802,000

"FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide" [Denver Guardian](#)

567,000

"RAGE AGAINST THE MACHINE To Reunite And Release Anti Donald Trump Album" [heaviermetal.net](#)

560,000

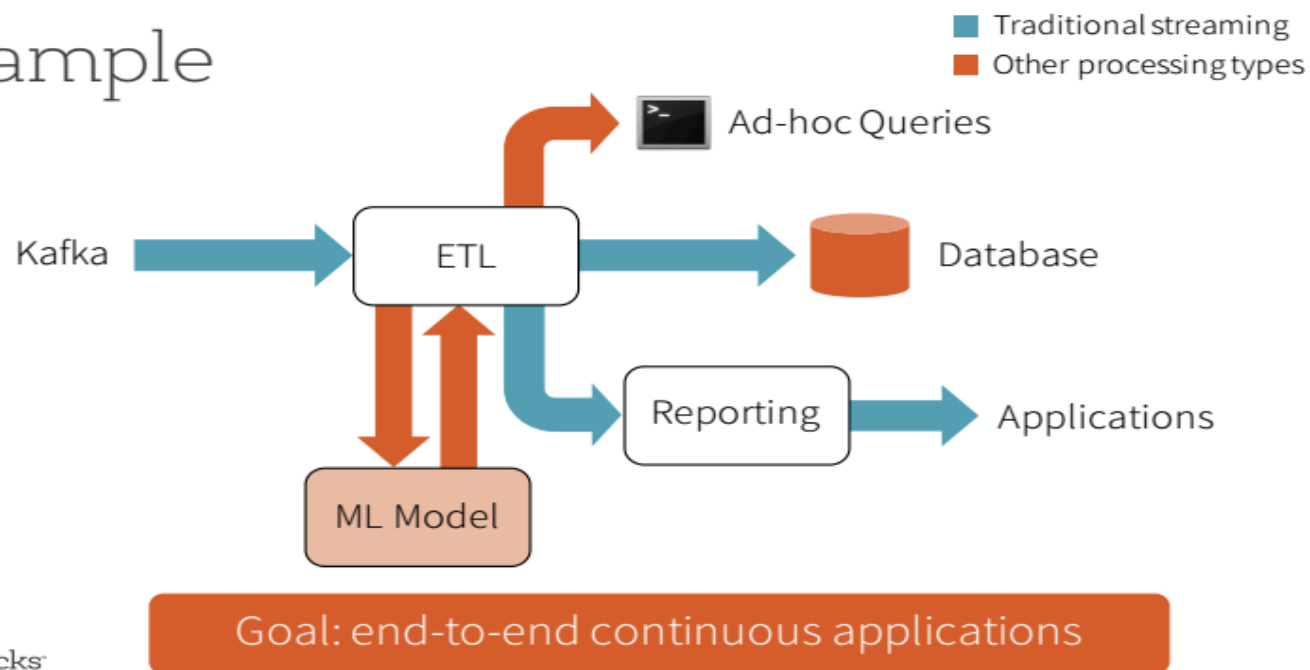
0 500,000 1M 1.5M 2M 2.5M

ENGAGEMENT REFERS TO THE TOTAL NUMBER OF SHARES, REACTIONS, AND COMMENTS FOR A PIECE OF CONTENT ON FACEBOOK
SOURCE: BUZZSUMO

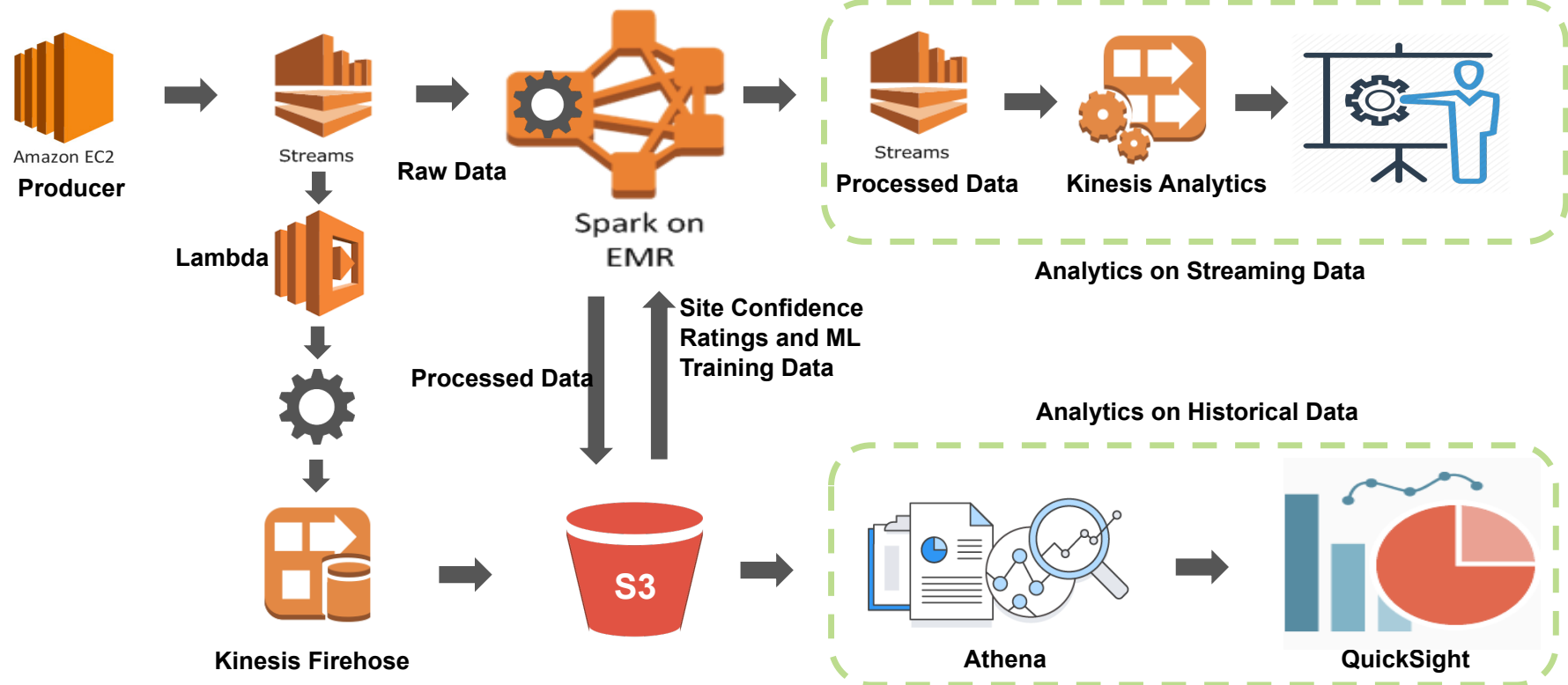
Cloudwick

Programming Model

Example



Architecture Diagram



Demo: “There will be War and Code will Flow”

- Producer Publishing Information to the Stream
- Spark Client Reading the stream and processing the information
- Streaming data Analytics on Processed data using
- Historical data Analytics on Processed data using Athena
- Visualization using QuickSight

References

➤ Databricks Mastering Advance Analytics with Apache Spark

➤ Persisting Streaming data to S3 using Firehose

<https://aws.amazon.com/blogs/big-data/persist-streaming-data-to-amazon-s3-using-amazon-kinesis-firehose-and-aws-lambda/>

➤ Kinesis and Apache Spark Integration

<https://spark.apache.org/docs/2.1.0/streaming-kinesis-integration.html>

Questions



Cloudwick



Thank You



Cloudwick