My View on Big Data

Presented by Ethan Moore

Outline

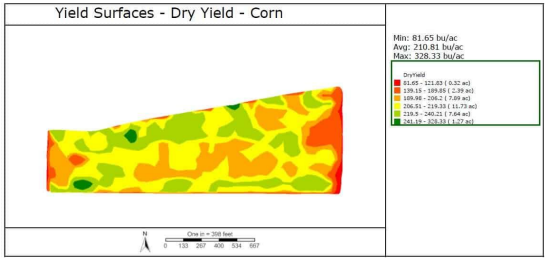
1. Introduction
    a. Brief explanation of Big Data
2. Body
    a. Who uses Big Data and Why?
    b. Data Collection Approaches
    c. Distributed File Systems and their Models
    d. Results
3. Conclusion
    a. Concluding Thoughts
4. Sources

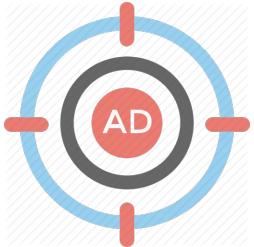
Brief Introduction to Big Data

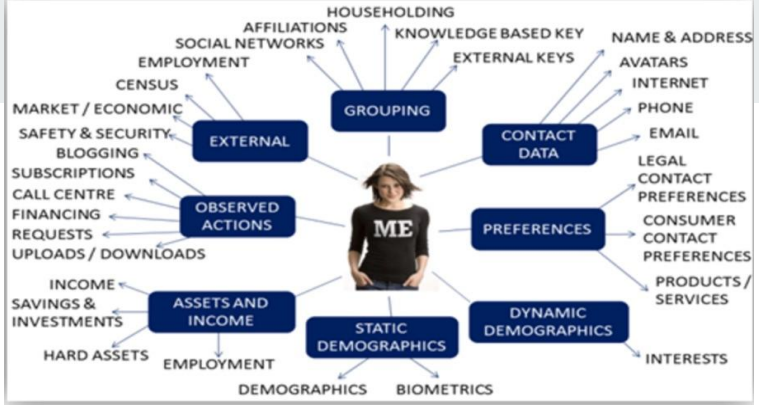
- The U.S department of Commerce defined Big Data as "extensive datasets-primarily in the characteristics of volume, variety, velocity, and/or variability-that require a scalable architecture for efficient storage, manipulation, and analysis."
- Using Big Data to provide a service
- Data collection techniques
- Storing datasets using a DFS
- Analyze patterns within datasets

Who uses Big Data and Why?

- The public and private sector use Big Data to gain a greater insight into either customers or situations that need to be monitored
  - Ex. U.S. Govt., Facebook, Walmart etc.
- Many job roles encapsulate themselves in Big Data
  - North Eastern University compiled a list of relevant Big Data careers
    - Ex. Big Data Engineer, Data Architect, Data Warehouse Manager, Database Manager, etc.
- Utilizing collected information from a subject can illuminate previously unseen correlations
  - Ex. From farm yields to targeted ads on your choice social media platform the applications are endless







Data Collection Approaches

- Edward Snowden leaked the previously unknown United States mass surveillance program on the public.
- Walmart saving your shopping preferences and patterns.
- Facebook targets ads by correlating a person's demographic and browsing history with a third parties products.
- Ethical Questions should be raised
    - Is Governmental Mass surveillance acceptable if the only people that would be affected is the most egregious and dangerous domestic and international criminals/terrorists?
    - Should I be notified that my data has been collected?
    - Should collected data be able to be sold to a third party?

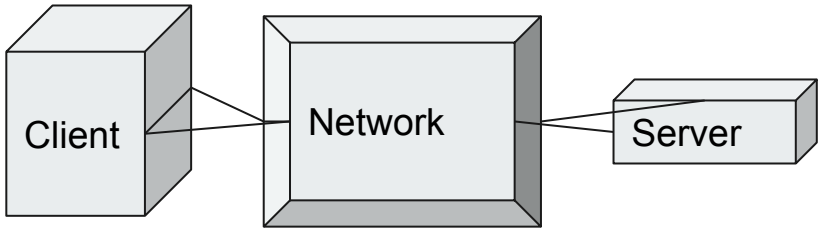
Distributed File Systems and their Models

A Distributed File System or also known as DFS is defined in OS concepts as “a file system whose clients, servers, and storage devices are dispersed among the machines of a distributed system. Accordingly, service activity has to be carried out across the network. Instead of a single centralized data repository, the system frequently has multiple and independent storage devices.”

The framework of a distributed file system can be expressed in several ways. The most prominent being the Client Server Model and the Cluster Based Model.

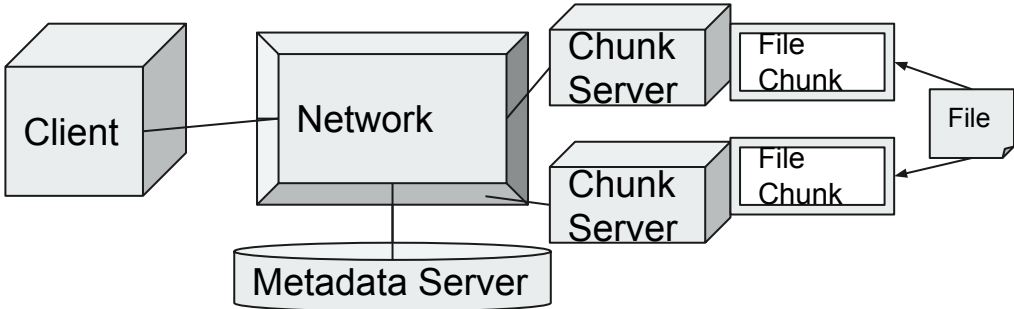
Client Server Model

Client(s) communicates to the Server(s) over the Network to deliver file(s)



Cluster Based Model

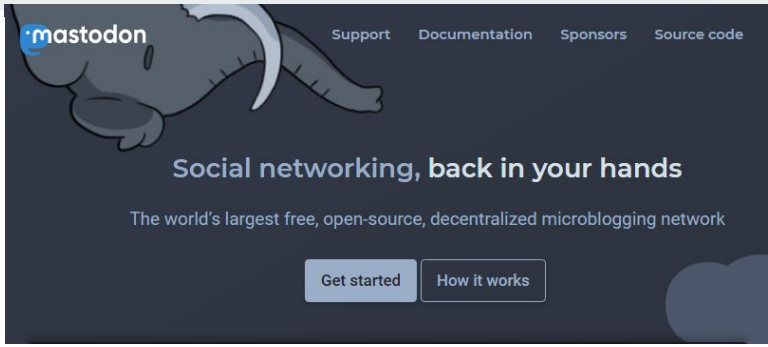
Client(s) communicates to the Metadata Server which communicates to the Chunk Server(s) to deliver the file(s)



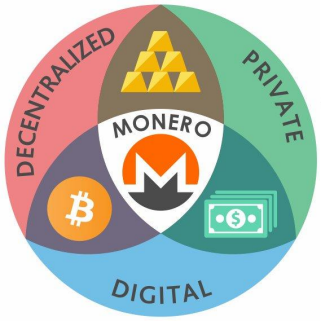
Results

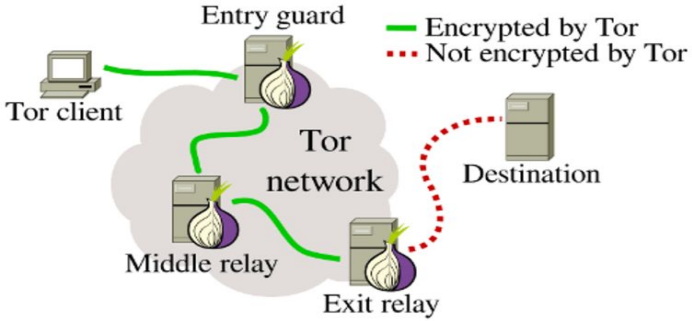
- The book Big Data in practice stated that the U.S. government “prevented $820 million in fraudulent payments”
- Walmart through data analytics provides a better customer experience
- Facebook has become one of the largest platforms for targeted advertising due to their user base.

Concluding Thoughts



- Data science in the future will be even more important and demonstrated in its infinite implementations due to incredible financial and tracking results.
- With so much being collected and the frequency of applications of Big Data appearing. Privacy and circumventing these surveillance traps is a important consideration.
- Privacy related solutions to Big Data Dominance and increasing public awareness will become increasingly critical and important.





Sources

Slide 1 background image :
https://blogs.oracle.com/content/published/api/v1.1/assets/CONTF90B029FF1994CA7A4432BFDE0B96DC7/native/oracle_big_data_blog_banner_copy.png?channelToken=94cb2007ef2e49d0916bb6718de06990

Outline and Sources image:

https://www.pixelstalk.net/wp-content/uploads/2016/11/Ultra-HD-Data-Wallpaper.jpg

https://cdn3.iconfinder.com/data/icons/seo-and-digital-marketing-5-3/48/227-512.png

https://andersonscanada.com/wp-content/uploads/2015/03/2015-05-25-15_13_19-Yield-Data.jpg

https://www.northeastern.edu/graduate/blog/highest-paying-big-data-careers/

https://neilpatel.com/wp-content/uploads/2016/07/customer-demographics-chart.jpg

U.S.DepartmentofCommerce, 2015] U.S.DepartmentofCommerce(2015).Nist big data interoperability framework: Volume 1, definitions. page 13.

[Marr, 2016] Marr, B. (2016). The us government using big data to run acountry. InBig Data in Practice: How 45 Successful Companies Used BigData Analytics to Deliver Extraordinary Results, pages 229–235.