September 20 2017 An intro to Big Data

September 20th, 2017 An intro to big data

References:

https://research.google.com/archive/gfs.html "Google FIle System" paper

https://research.google.com/archive/mapreduce.html Mapreduce paper

Public datasets on AWS:https://aws.amazon.com/public-datasets/

Erasure Codes are coming to HDFS (like RAID) https://code.facebook.com/posts/536638663113101/saving-capacity-with-hdfs-raid/

Cloud Providers

https://aws.amazon.com/emr/ The Amazon Elastic Map Reduce

https://aws.amazon.com/marketplace/pp/B01M193KGR Hortonworks cloud image inside amazon

https://azure.microsoft.com/en-us/services/hdinsight/ Azure HDInsight

Hortonworks Sandbox (VM for demo)

https://hortonworks.com/products/sandbox/

Data for Demo:

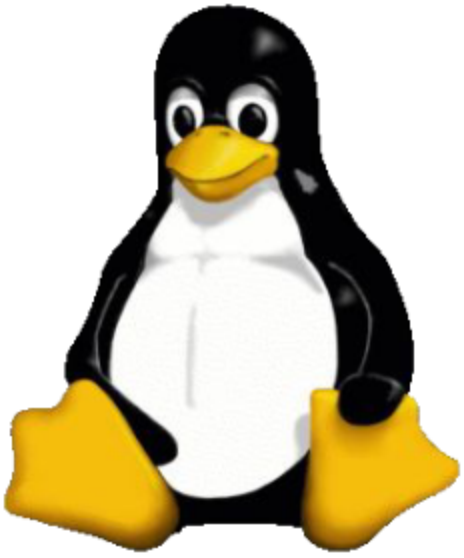
http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

http://www.andresmh.com/nyctaxitrips/

Slide Deck

Big Data.pptx

Big Data.pdf 1 of 24 pages •



Welcome To CIALUG

Big Data.odp

Video: https://youtu.be/a9DTISzhBmA