**INTRO TO HADOOP AND MAP REDUCE** *by Cloudera*

1. Introduction
2. Data Sources
3. Big Data
4. Big Data
5. Definition of Big Data = a big data that may be hard to process in a single machine
6. Challenges
7. Challenges
   1. Data is created very fast
   2. Data is from different sources in various formats
8. The 3 V’s -- Volume
   1. Volume
   2. Variety
   3. Velocity
9. Worthwhile Data
10. Worthwhile Data
11. Variety
12. Data Formats (hadoop doesn’t care about data format)
13. Using Variety
14. Using Variety (more data is better)
15. Velocity (even a TB ok lang)
    1. Example usage: Shopping usage, or NETFLIX, recommendations etc.
16. Your interests
17. Doug intro
18. Doug cutting the Origins of Hadoop
    1. Billions of pages to be processed at fast speed.
    2. Google published papers (i.e. map reduce), which was the inspiration
    3. Can run on thousands of machines, processors, etc.
    4. Yahoo then funded it
    5. Hadoop became kernel for many big data projects.
19. Hadoop Logo Intro
20. Doug Cutting the Name of Hadoop
    1. Came from Doug’s toy (an elephant)
21. Core hadoop
    1. Store in HDFS (hadoop distributed file system)
    2. Process with mapreduce
    3. May machine cluster sa gitna. Dito iniistore at ito ung kinakausap ng both. Not necessarily high end ung machines sa cluster
22. Hadoop Ecosystem
    1. MapReduce (may be written in java, python, ruby, sql)
    2. Pig, Hive - used para pasimplihin ung scripting para sa mga non-devs
    3. Impala - mas mabilis sa hive
    4. Scoop, flume
    5. Hue, oozie, mahout
    6. Work together
    7. Cloudera created CDH (Cloudera Distribution Hadoop), para madali lang ung installation ng lahat ng installation ng ecosystem
23. Congratulations