Introduction to data

* Data Sources

1. Machine generated (Server logs, RFID, Sensors, scanner, crawling, mail logs, and many more...)
2. User generated (transactions, social networking, media, searches, emails and many more...)

* Types of data
  1. Structured
  2. Unstructured
  3. Semi-structured
* Data properties

1. Volume
2. Variety
3. Velocity
4. Complexity

Introduction to System

* Types of systems
* Problems with systems

Scaling the Systems

* Types of scaling

1. Scaling up (Vertical)
2. Scaling out (Horizontal)

Data vs Big Data

Big Data => (Transactions + Observations + Operations)

Traditional Systems vs Distributed Systems

Introduction to NoSQL World.

Distributed Systems

* Categories:

1. Data-Intensive 🡪 HADOOP
2. Computation-Intensive 🡪 SETI@HOME

Introduction Hadoop

History

* Comparing Hadoop with Other Distributed Systems
* Comparing Hadoop with SQL Databases and Warehouses

Hadoop Overview

* Hadoop Distributed File System
* Map Reduce Programming Model
* Hadoop Common Utilities
* Hadoop Ecosystem

Sample Use cases

* Telecoms data processing
* Spam filter

Hadoop Adoption in Business Verticals

* Telecommunication
* Finance
* Insurance
* Retail
* Manufacturing
* Media
* E-commerce
* Travel
* Health care
* Natural resources
* Pharmacy
* Geo physics
* Government
* Many other... industries are going to adopt Hadoop for processing vast amount of data

Question and Answers