Sunbeam Infotech

Big Data Analytics
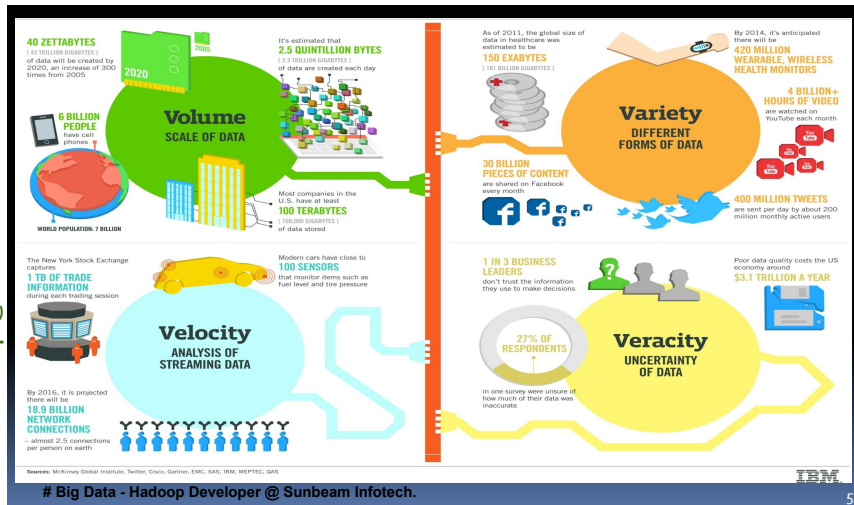

More than A CD per day?

Big Data - Hadoop Developer @ Sunbeam Infotech.

2

KB -> MB -> GB -> TB -> PB -> XB -> ZB -> YB -> BB -> JB

Big Data - Hadoop Developer @ Sunbeam Infotech.

3

Big Data - Hadoop Developer @ Sunbeam Infotech.

4

IoT & Big Data

- Home Automation
  - ON/OFF appliance
  - Report state of appliance periodically.
- Huge Storage
  - High velocity, High volume & High variety.
- Process data
  - Analyze Time/Day of max/min usage.
  - Decide state of appliance - minimize electricity.
  - Invent new products, new marketing schemes.

Data Engg vs Data Science

- Databases
  - RDBMS & NoSQL
  - Data Warehousing
- Big Data
  - Hadoop, Spark, HDInsight, BigTable, ...
- Infrastructure
  - Linux, Cloud Computing
- Parallel computing

- Data Science
  - Statistics and Machine learning
  - R Programming, Python
- Data Analytics
  - Data Visualizations & Business Decisions
  - Tableau, QlikView, ...

Big Data

- It is all about :- Think, Collect, Manage, Analyze, Summarize, Visualize, Discover Knowledge and Take Decisions.

- Domains: Health-care, Retails, Trading/Share market, Finance, Security, Fraud, Search engines, Log Analysis, Telecom, Traffic Control, Manufacturing and lot more.
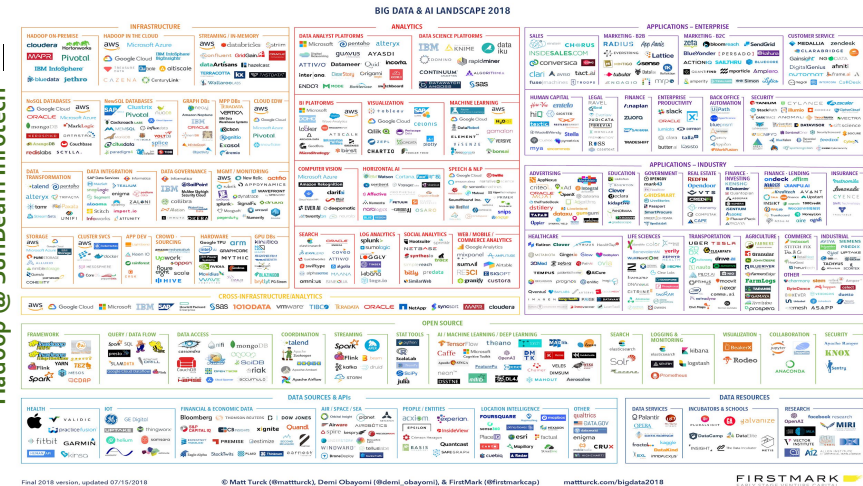
Scope for Big Data

- By 2020, more than 60 billion dollar industry.
- Job profiles:
  - Big Data Developer, IT operations
  - Data analyst, Data Scientist, Business Intelligence
  - Big Data Architect
- The sexiest job in the 21st century require a mixture of multi-disciplinary abilities and suitable candidates must be prepared to learn and develop constantly.
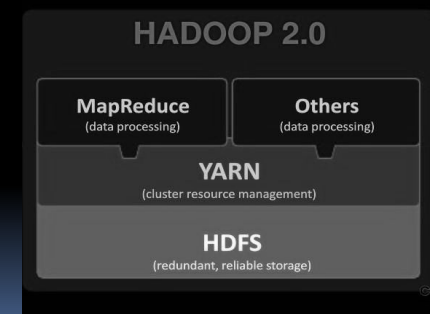

Sunbeam Infotech

Apache Hadoop

Hadoop 2.x

- Open source implementation of Big Data.
  - Doug Cutting
- Apache Software Foundation
  - Hadoop - Core/Kernel/Platform
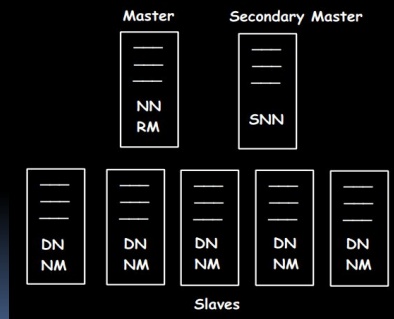
Hadoop 2.x Daemons

- Hadoop framework is implemented in Java.
- It is made up of few java processes running in background.
- HDFS Daemons
  - Name Node, Secondary Name Node, Data Node (s)
- MapReduce/YARN Daemons
  - Resource Manager, Node Manager (s)

Master | Secondary Master

NN RM | SNN

DN NM | DN NM | DN NM | DN NM | DN NM

Slaves

Big Data - Hadoop Developer @ Sunbeam Infotech

Hadoop Eco Systems

Big Data - Hadoop Developer @ Sunbeam Infotech.

Hadoop Distributions

- Installing, monitoring, troubleshooting and maintaining Hadoop cluster along with Eco-systems is tedious.
- Hadoop distributions provide installations & tech-support for Hadoop eco-systems commercially.
- Some of the services are available on-premise and most of them are cloud based.
- Most popular are: Cloudera, HortonWorks, MapR, IBM, Intel, Microsoft, AWS and Pivotal Software.
- These services are preferred in production due to their stability, reliability and cost effectiveness.

Big Data - Hadoop Developer @ Sunbeam Infotech.

Nilesh Ghule <nilesh@sunbeaminfo.com>
Technical Head, Sunbeam Infotech.

THANK YOU!