Overview
BIG DATA Computation

What is Big Data?

- It's a Buzz Phrase. No Single Definition

- Big data is a term for data sets that are so large or complex that traditional data processing application software is inadequate to deal with them.

- Challenges include capture, storage, analysis, data curation, search, sharing, transfer, visualization, querying, updating and information privacy.

- Now, refer to Big Data Analytics

The Rise of Big Data

- Technology Growth
- Internet Adoption
- People Behaviour
- Digitize Everything
- Competition

Big Data and AI Landscape

Matt Turck
VC at FirstMark Capital
mattturck.com/bigdata2019

Infrastructure

Hadoop on Prem, Hadoop on Cloud, Streaming/In Memory, NoSQL Databases, NewSQL Databases, Graph DBs, MPP DBs, Cloud EDW, Data Transformation, Data Integration, Data Governance, Mgmt/Monitoring, Storage, Cluster Svcs, App Dev, Crowd Sourcing, Hardware, GPU DBs

ANALYTICS

Analytics

Data Analytics Platforms, Data Science Platforms, BI Platforms, Visualization, Machine Learning, Computer Vision, Horizontal AI, Speech & NLP, Search, Log Analytics, Social Analytics, Web/Mobile/Commerce Analytics

APPLICATIONS – ENTERPRISE

Application Enterprise

Sales - Marketing B2B - Marketing B2C - Customer Service - Human Capital - Legal - Finance - Enterprise Productivity - Back Office Automation - Security

APPLICATIONS – INDUSTRY

Application Industry

Advertising - Education - Government Finance - Real Estate - Insurance - Healthcare Life Science - Transportation - Agriculture Commerce - etc

CROSS-INFRASTRUCTURE/ANALYTICS

Cross Infrastructure/Analytics

OPEN SOURCE

Open Source

Frameworks, Query/Data Flow, Data Access, Coordination, Streaming, Stat Tools, AI/ML/DL, Search, Logging & Monitoring, Visualization, Collaboration, Security

DATA SOURCES & APIS

Data Source & APIS

Health, IoT, Financial & Economic Data, Air/Space/Sea, People, Location Intelligence, Other

DATA RESOURCES

Data Source

Open Source Technology

- Most of big data component is open source
- We can download the code, use and modify freely
- Require adequate human resources
- Lots of choices

What is Hadoop?

- Open Source Platform for data management
- Combination of distributed storage and distributed processing
- Computer cluster built from commodity hardware
- Framework written in java programming
- Offering scalability and high performance
- The name Hadoop is not an acronym; Doug Cutting named it after his son's toy elephant

History of Hadoop

- Mike Cafarella and Doug Cutting started the Nutch project in 2002

- In 2003, Google published Google File System paper, that described the architecture of Google's distributed file system

- By adopting GFS, Nutch Distributed File System (NDFS) began to be implemented on the Nutch project in 2004

- In 2004, Google published the paper that introduced MapReduce to the world

- Early in 2005, the Nutch developers had a working MapReduce implementation in Nutch

- In February 2006 they moved out of Nutch to form an independent subproject of Lucene called Hadoop

- April 2006 – Hadoop 0.1.0 was released

Disruptive Technology

- Open source – zero license
- Proven by big internet company
- Active community
- Fast adoption

YAVA Data Management Platform

Kafka	HGrid247 Visual workflow designer	Indexing Solr
Flume	Machine Learning Mahout, MLib, SparkR	SQL Hive, SparkSQL, Phoenix
Sqoop	Processing Frameworks MapReduce, Spark, Storm, Tez	NoSQL HBase, Accumulo

Resource Management
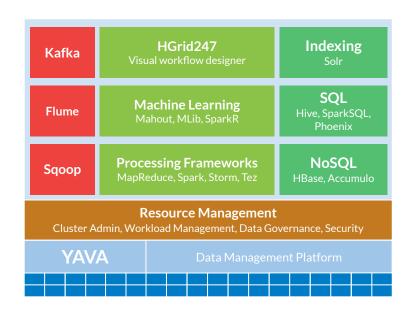Cluster Admin, Workload Management, Data Governance, Security

YAVA | Data Management Platform

All in one data management platform
Programming/Scripting :
- Java, Python, Scala, R
- SQL
- HGrid247 - Visual Designer

For further info : **yava.labs247.id**

Big data and artificial intelligence platform based on open source component. It is designed to make organization easier to implement big data.

BIG DATA Computation

Use Case

Archival and Storage

- Retain years of data
- Retain intermediate format

Transformation

- Map inputs and outputs where needed
- Turn unstructured data into structured at runtime

Analysis

- Explore data in-place
- Execute arbitrary code

Use Case - Media Analytics

Use Case - Media Analytics

Use Case - Media Analytics

Use Case - Media Analytics

Use Case - Media Analytics