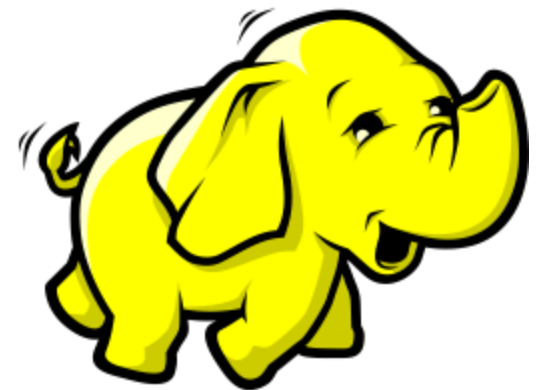




Introduction to Hadoop

Welcome

The motivation for Hadoop



The motivation for Hadoop

We're
generating
more data
than ever

- Sensor networks
- Server log files
- E-Mail
- Social Media



The motivation for Hadoop

2021 *This Is What Happens In An Internet Minute*



The motivation for Hadoop

Pre-smartphone area

Few companies are generating data, all others are consuming data

Now

All of us are generating data, and all of us are consuming data

The motivation for Hadoop

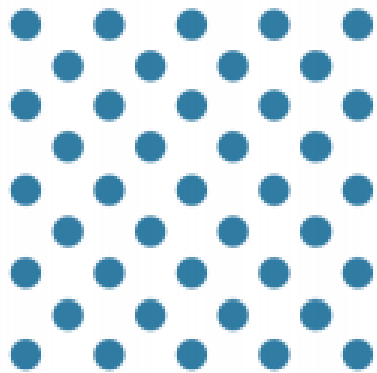
- **Data is getting bigger**
- **Data Storage is cheap**
 - Less than a \$0.10 per GB
- **Problem however is not the storage**
 - It's getting the data off the disk and to the processor
 - Producing value out of the data becomes is challenging
 - Storing and handling of data becomes an issue

The motivation for Hadoop

- **Data is getting bigger**
- **Data Storage is cheap**
 - Less than a \$0.10 per GB
- **Problem however is not the storage**
 - It's getting the data off the disk and to the processor
 - Producing value out of the data becomes is challenging
 - Storing and handling of data becomes an issue

Volume

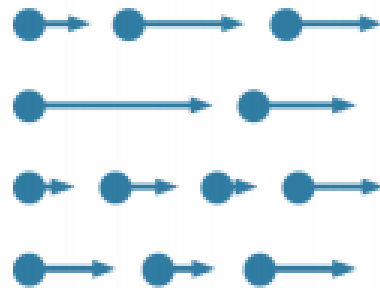
Data at rest



- Tera- to Exabytes to process
- Sensor and social data
- New data stores

Velocity

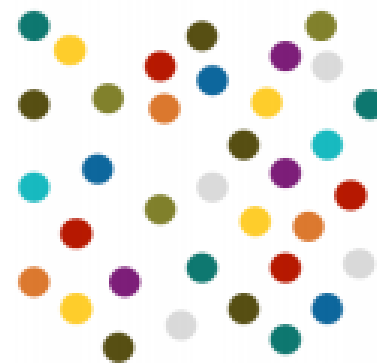
Data in motion



- Streaming data
- (Milli)seconds to minutes to respond

Variety

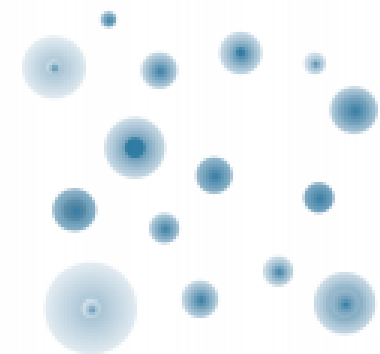
Data in many forms



- Structured and unstructured data
- Text, numbers and multimedia

Veracity

Data in doubt



- Uncertainty due to data inconsistency and incompleteness, ambiguity, latency, deception and model approximations

The motivation for Hadoop

- **BigData is big**
 - Technological infrastructure solutions needed
- **BigData is complicated**
 - Developers to handle the data
 - Operations for Cluster management
 - Data analysts and scientists to produce value

The motivation for Hadoop

Infrastructure
needs

Distributed and scalable

Job isolation

Efficient and cost-effective storage

Quick Network infrastructure