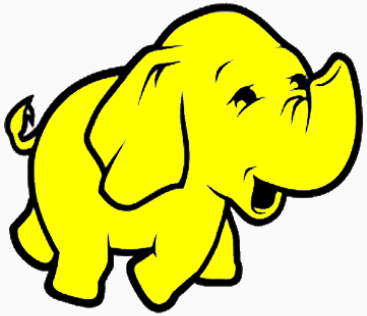


# HADOOP
## Open Source Framework

Leonardo Enrique Alvarado Alvarez
Jesús Alfredo Sanabria Mejía

0060 Bases de Datos

September 10, 2022


# Content

What is Hadoop?
A bit of History
How it Works?
advantage
Who uses it?


# What is Hadoop?

Hadoop is an open source framework that is used to efficiently store and process large datasets ranging in size from gigabytes to petabytes of data.


# Hardware Requirements

1) Intel Core 2 Duo/Quad/hex/Octa or higher end 64 bit processor PC or Laptop (Minimum operating frequency of 2.5GHz)

2) Hard Disk capacity of 1- 4TB.

3) 64-512 GB RAM

4) 10 Gigabit Ethernet or Bonded Gigabit Ethernet


# A bit history i

Hadoop was started with Doug Cutting and Mike Cafarella in the year 2002 when they both started to work on Apache Nutch project, that could index 1 billion pages. That system will cost around half a million dollars in hardware, and along with a monthly running cost of $30, 000 approximately, which is very expensive.
in 2006, Doug Cutting joined Yahoo along with Nutch project and formed a new project Hadoop (He gave name Hadoop it was the name of a yellow toy elephant which was owned by the Doug Cutting's son.)
In 2009, Hadoop was successfully tested to sort a PB (PetaByte) of data in less than 17 hours for handling billions of searches and indexing millions of web pages.

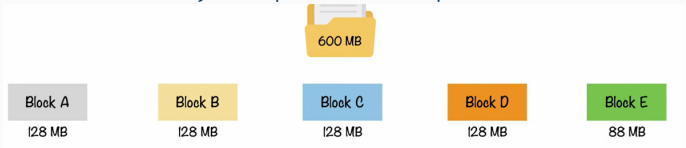

# A bit history ii

Doug Cutting left the Yahoo and joined Cloudera to fulfill the challenge of spreading Hadoop to other industries.
Currently, we have Apache Hadoop version 3.0 which released in December 2017.

# How it works?

# How it works? 1. HDFS

1. **HDFS** Data is stored in many computers and split in blocks

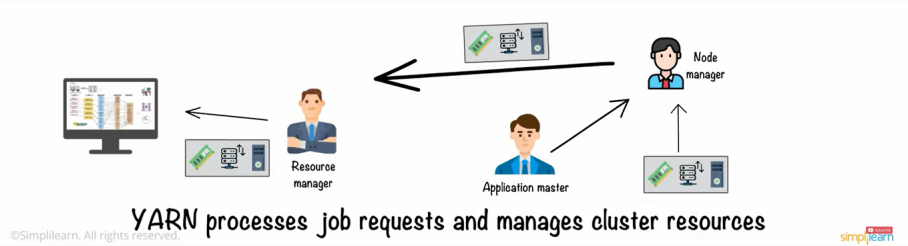


# How it works? 2. YARN

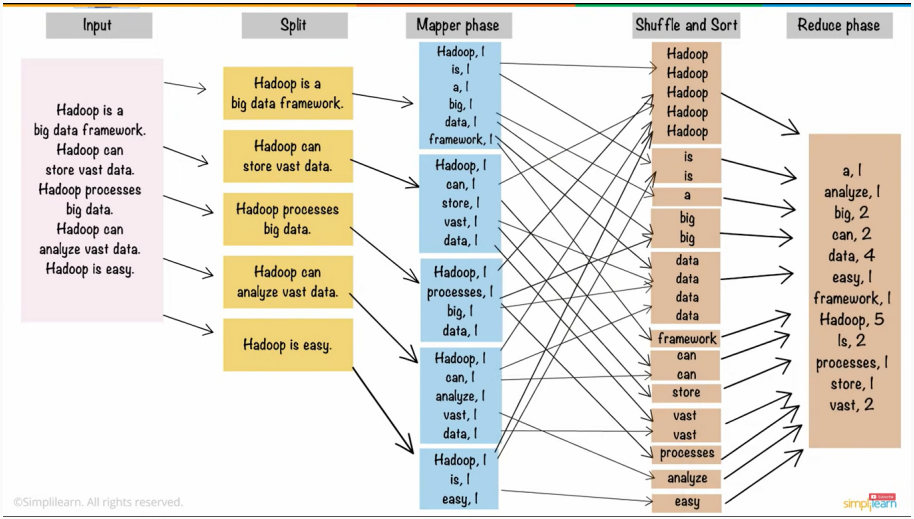
**2.YARN** Yet Another Resource Negotiator, It manages the resources

# How it works? 3. MapReduce

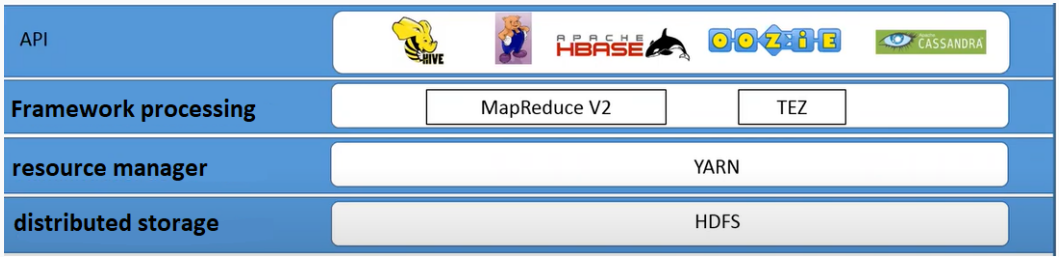
**3.MapReduce** Insted of using a single machine and processor, MapReduce split data into parts and processes each of them separately on different data nodes



# How it works? 3. MapReduce

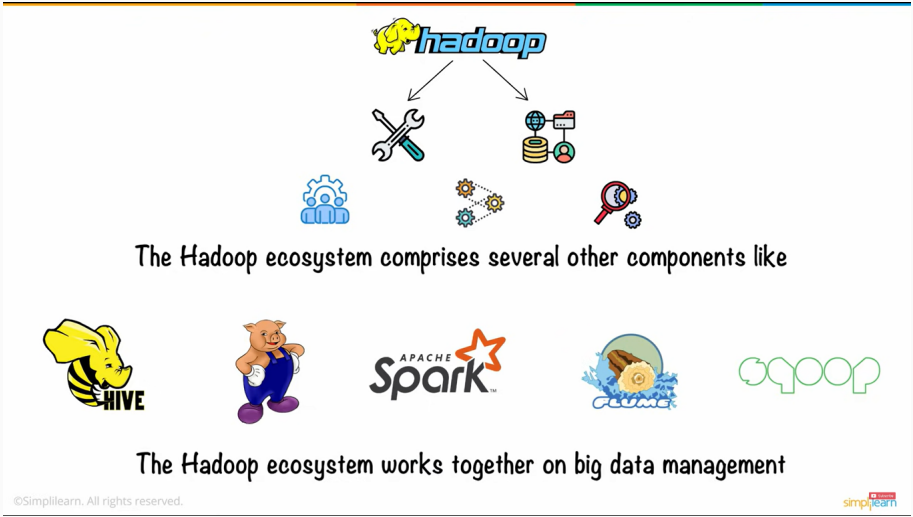


# How it works? resume



# How it works? ....Tools

**Tools** It also have tools to dedicated to managing, processing and analyzing data




# advantage

**storage** Handles large volumes of data, structured, unstructured and semi-structured
**prosecution** processes in a distributed and parallel manner with high performance
**flexibility** handling large volumes of data is done in batches
**Fault tolerance** It has replication, in several computers so that if one fails another replaces it
**scalability** growing horizontally, adding nodes

# Finally! it is used by!

Warehousing, Recomendation system, Fraud Detection

# Material to investigate!


**(a)** short ebook

[inc, ]


**(b)** taller online

[Bayona, ]fig:taller


**(c)** Course / 10 hours

[simplilearn, ]

# Recomendations


Bayona, I. R.
**Hadoop en la practica.**
https://www.youtube.com/watch?v=bBGU0GP5$_K$s.

inc, P.
**¿qué significa hadoop en el mundo del big data? un contenido para perfiles técnicos.**
https://f.hubspotusercontent30.net/hubfs/239039/Ebook

simplilearn.
**Getting started with hadoop, free introductory course to hadoop.**
https://www.simplilearn.com/introduction-to-hadoop-free-course-skillup.

Thanks!

# Referencias i


- Lam, Chuck (28 July 2010). Hadoop in Action (1st ed.). Manning Publications. p. 325. ISBN 978-1-935-18219-1.
- Venner, Jason (22 June 2009). Pro Hadoop (1st ed.). Apress. p. 440. ISBN 978-1-430-21942-2. Archived from the original on 5 December 2010. Retrieved 3 July 2009.
- White, Tom (16 June 2009). Hadoop: The Definitive Guide (1st ed.). O'Reilly Media. p. 524. ISBN 978-0-596-52197-4.
- Vohra, Deepak (October 2016). Practical Hadoop Ecosystem: A Definitive Guide to Hadoop-Related Frameworks and Tools (1st ed.). Apress.p. 429. ISBN 978-1-4842-2199-0.

# Referencias ii


- Wiktorski, Tomasz (January 2019). Data-intensive Systems. Cham, Switzerland: Springer. ISBN 978-3-030-04603-3.