



SANTOTO
— Seccional Tunja —

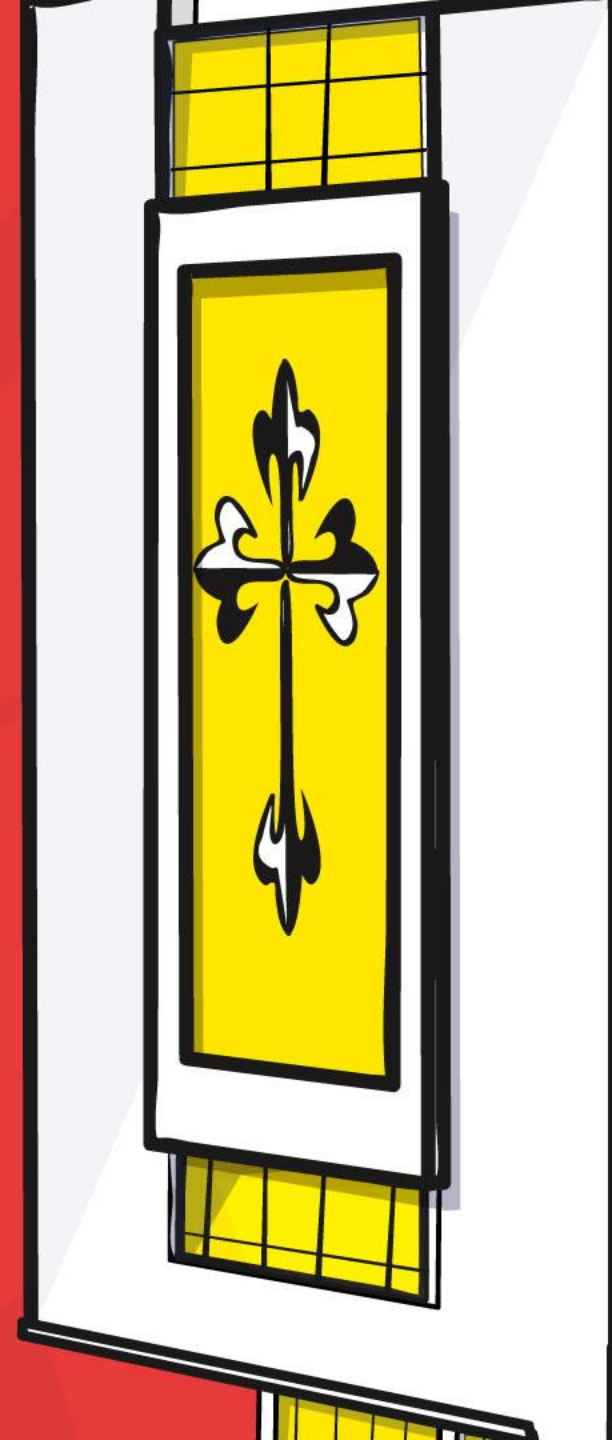
[WWW. USTATUNJA.edu.CO](http://WWW.USTATUNJA.edu.CO)



Apache Hadoop

Brayan David Ahumada Castañada
Juana Valentina Mendoza Santamaría
Luis Felipe Narvaez Gomez

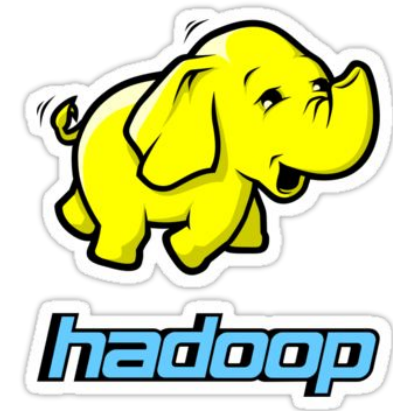
Big Data
Ingeniero Julián Alberto Monsalve Pulido



What is Apache Hadoop?

Hadoop is an open-source framework used to process **enormous datasets**.

- ✓ Hadoop was designed to process **large amounts of data**.
- ✓ Hadoop is optimized to handle massive quantities of data which could be: **structured, unstructured or semi-structured**.
- ✓ Hadoop allows running applications in **clusters**.
- ✓ Hadoop can handle processes and jobs in **parallel**.



In 2002, the **Nutch web** search engine was created by **Doug Cutting** and Mike Cafarella. In 2006, Cutting joined **Yahoo** with the Nutch project, and the project was divided into the web crawler and **distributed processing**. The distributed processing segment was called Hadoop, and in **2008**, Yahoo released Hadoop to the Apache Software Foundation.



How does Hadoop work?

Hadoop has individual components for storing and processing data. The **core** components of Hadoop include:

Hadoop Common

Hadoop Common refers to the common utilities and libraries that support the other Hadoop modules

HDFS

Hadoop Distributed File System (HDFS) stores the data collected from the ingestion and distributes the data across multiple nodes.

MapReduce

MapReduce is used for making Big Data manageable by processing them in clusters.

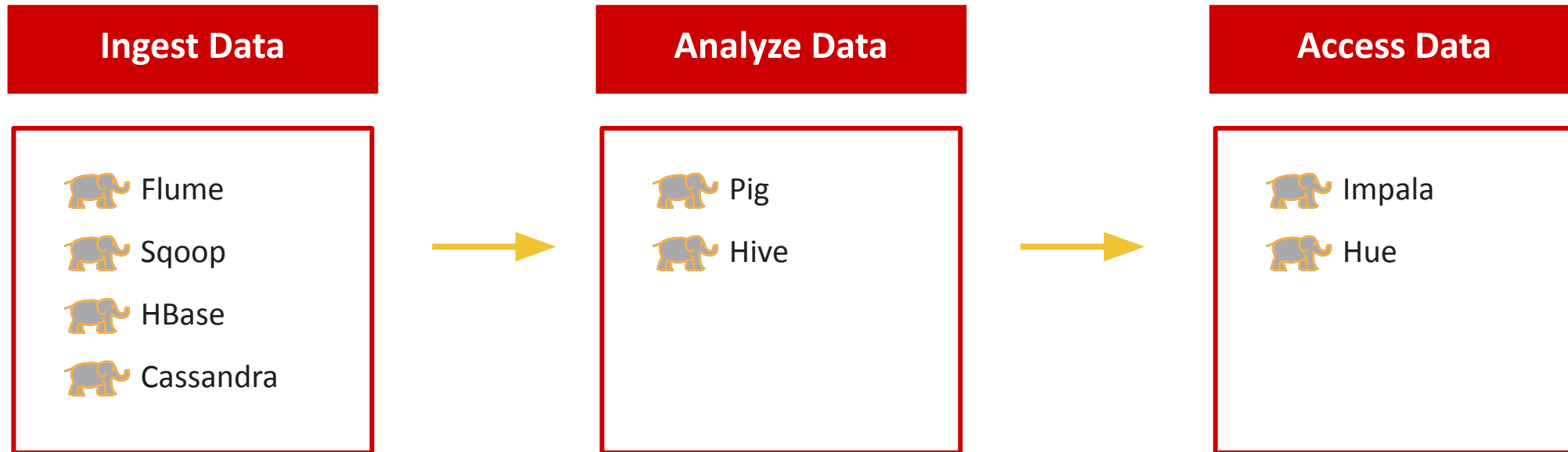
Yarn

Yet Another Resource Negotiator (YARN) is the resource manager across clusters.



How does Hadoop work?

The **extended** Hadoop Ecosystem consists of libraries or software packages that are commonly used with or installed on top of the Hadoop core.



The challenges of Hadoop

✗ Processing transactions
(random access)

✗ When there are
dependencies in the data

✗ Processing lots of
small files.

✗ When work cannot be
parallelized.

✗ Low latency data access

✗ Intensive calculations
with little data.

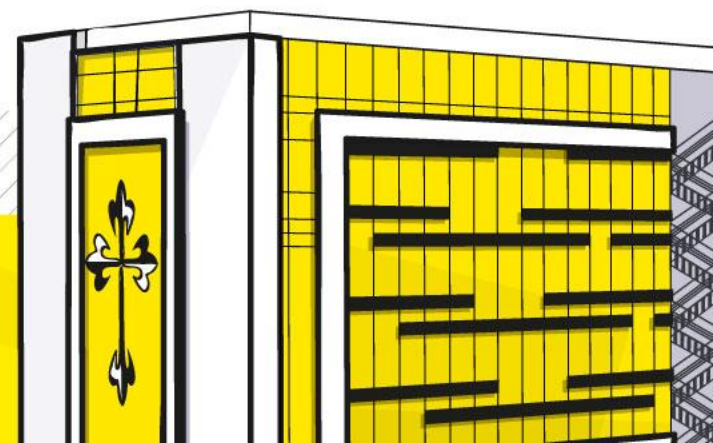


ustatunja.edu.co

-









SANTOTO
— Seccional Tunja —



SC4289-1

www.ustatunja.edu.co

♥ 🔍 📍 @SANTOTOMASTUNJA 📌