

~~HENRY~~



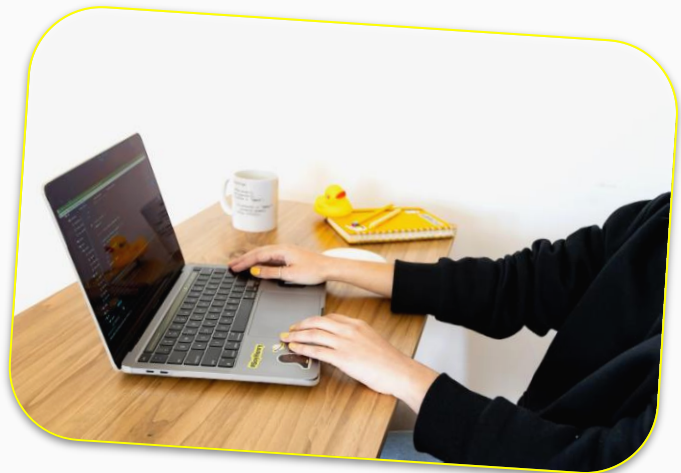
Frameworks Hadoop





OBJETIVOS DE CLASE

- **Entender** qué es un Cluster Hadoop y la diferencia con una arquitectura Cliente-Servidor
- **Identificar** los componentes Core de Hadoop (HDFS, YARN, MapReduce)
- **Comprender** el concepto de Frameworks en Hadoop
- **Conocer** algunos Proveedores de Servicios (Vendors) de grado empresarial



AGENDA

- Hadoop
- Características Hadoop
- Cluster Hadoop
- HDFS (Hadoop Distributed File System)
- YARN (Yet Another Resource Negotiator)
- MapReduce
- Frameworks Hadoop
- Proveedores de Servicios (Vendors) de grado empresarial



Hadoop



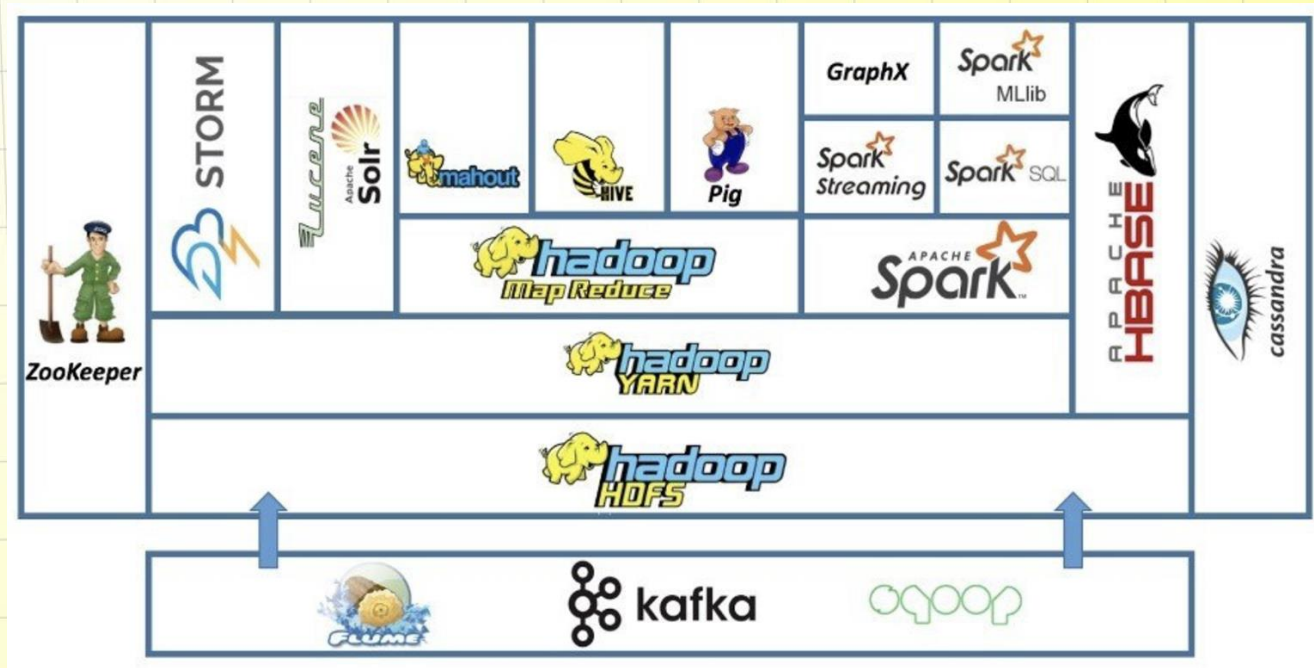
Es un sistema *open-source* diseñado para almacenar y procesar Big Data de forma distribuida utilizando un **clúster de servidores**.

Características:

- Tolerancia a Fallos
- Escalabilidad Horizontal
- Utiliza "Commodity Hardware"
- Desarrollado en lenguaje Java
- Procesamiento en paralelo

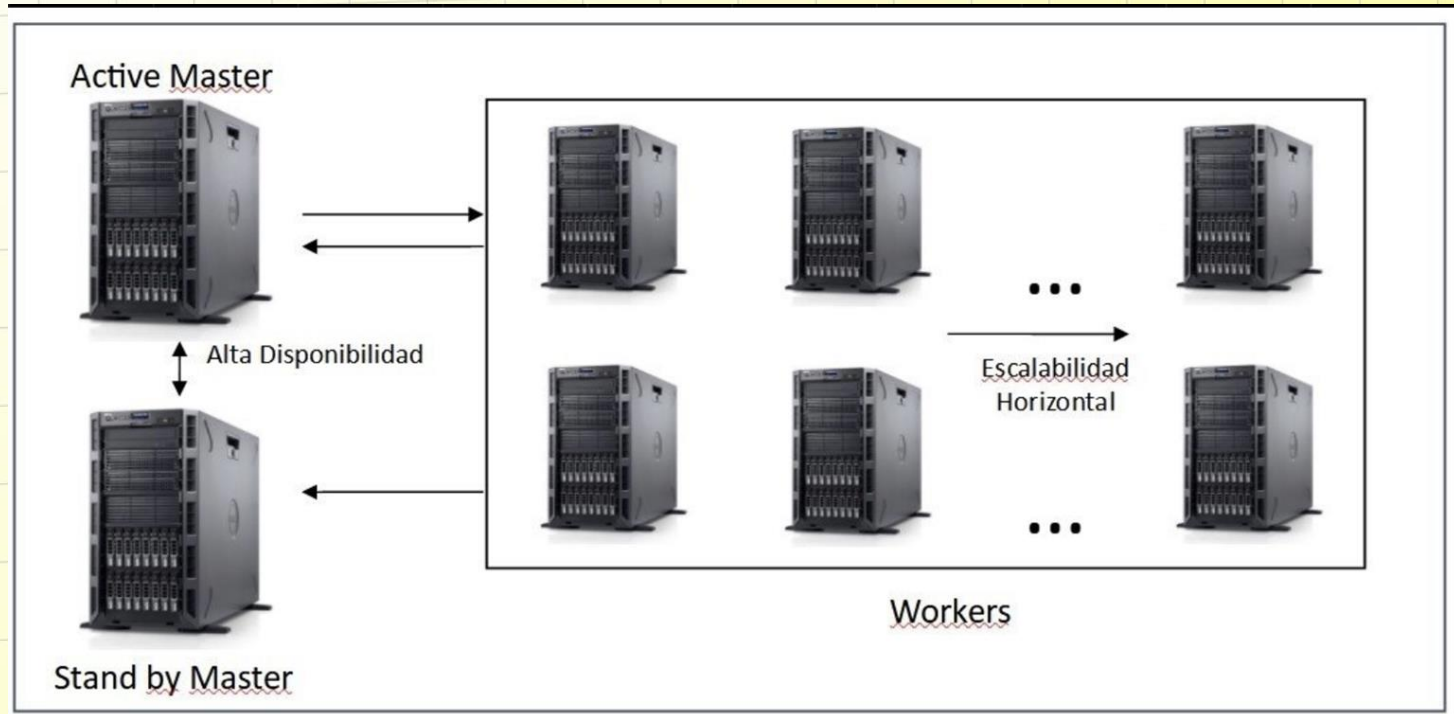


Ecosistema Hadoop



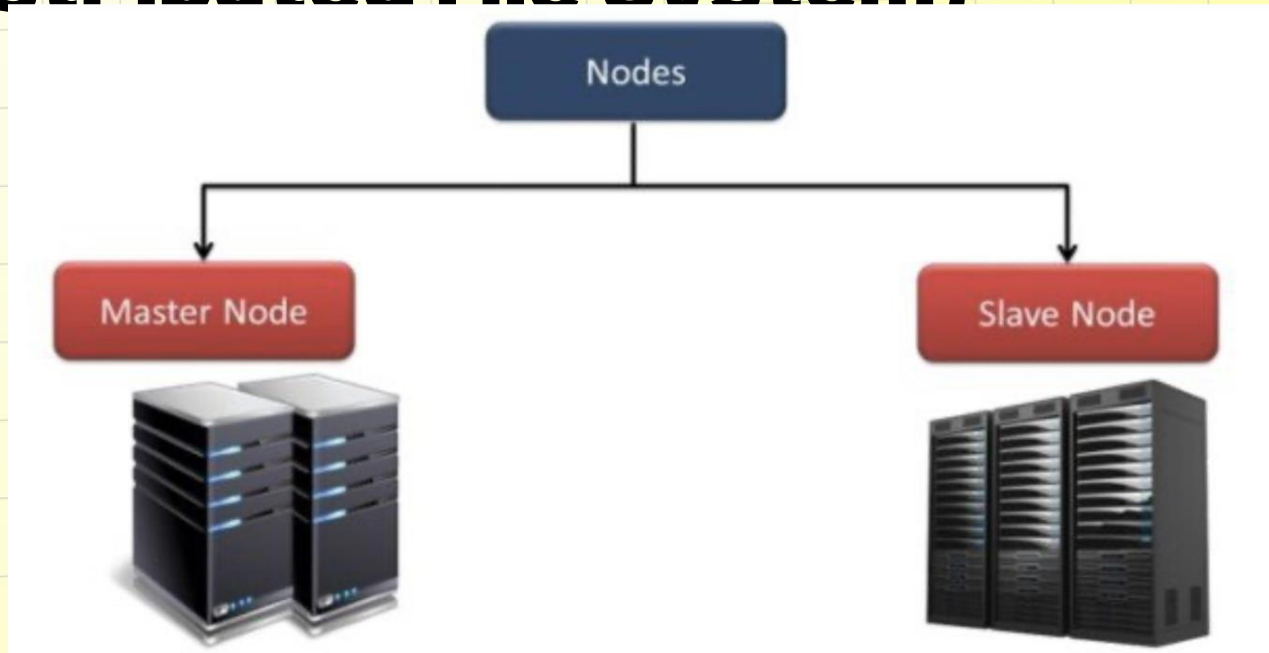


Cluster Hadoop



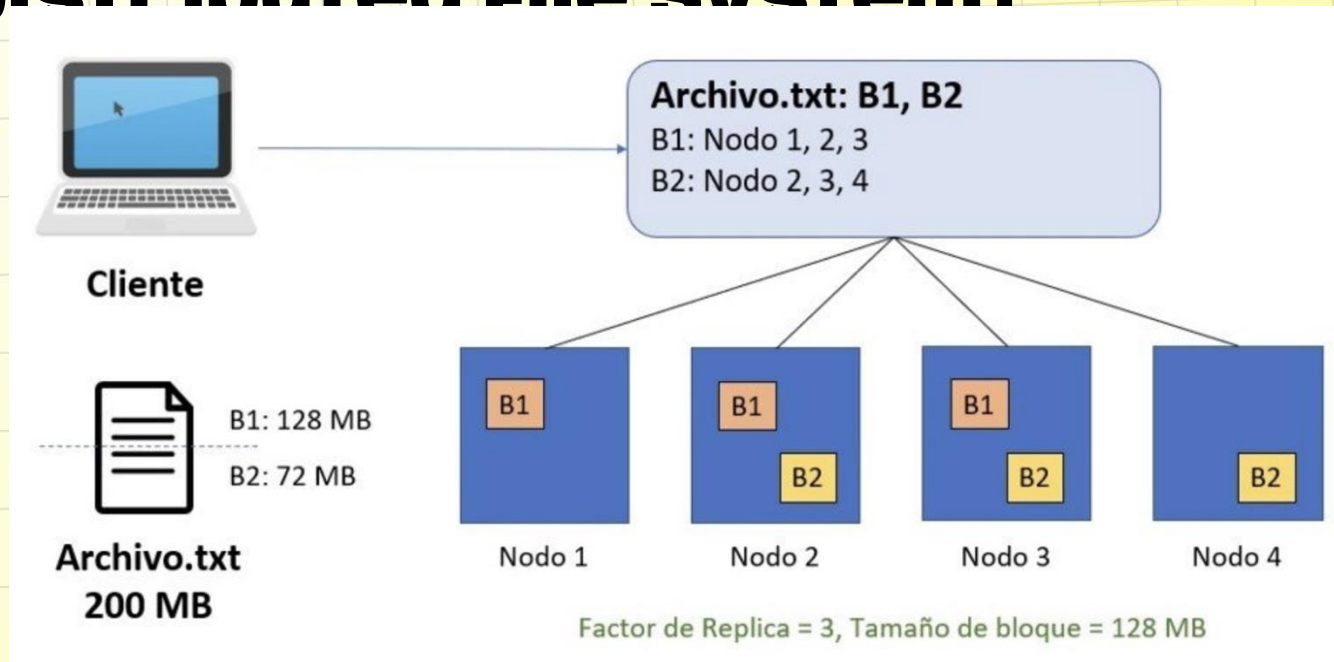


Componentes Core HDFS (Hadoop Distributed File System)



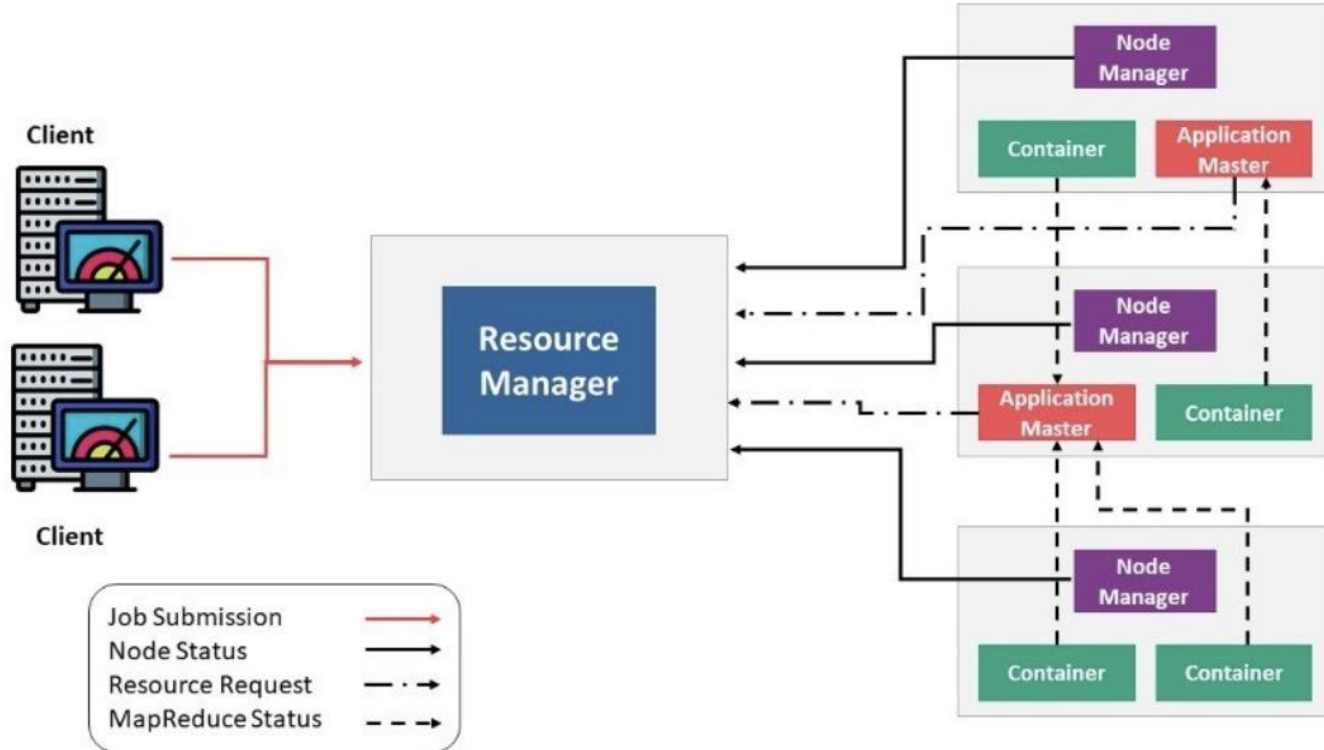


Componentes Core HDFS (Hadoop Distributed File System)



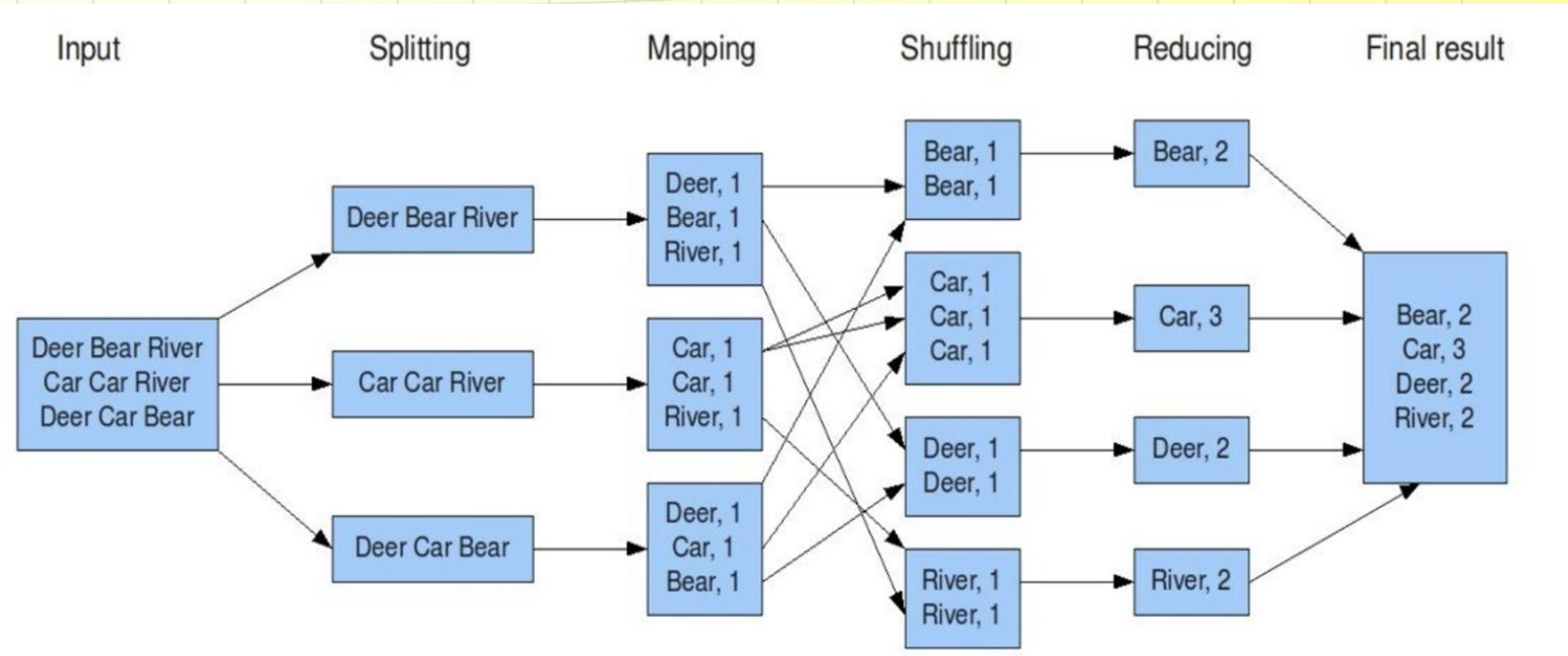


Componentes Core YARN (Yet Another Resource Negotiator)





Componentes Core MapReduce



Frameworks Hadoop





Apache Hive	Consultas SQL sobre Hadoop
Apache Sqoop	Transferencia de datos entre bases relacionadas y Hadoop
Apache Spark	Procesamiento en memoria de ETL's, Streaming, Machine Learning y Grafos
Apache Kafka	Sistema de colas de mensajería que utiliza el patrón productor/consumidor
Apache HBase	Base de datos NoSQL de tipo columnar que se ejecuta sobre HDFS
Apache Ranger	Administración de políticas de seguridad sobre componentes de Hadoop
Apache Atlas	Herramienta que provee funcionalidades de Data Governance sobre Hadoop
Apache Nifi	Orquestación de flujos de datos y desde hacia Hadoop.



Tecnologías de Big Data



Algunas tecnologías de Big Data





¿PREGUNTAS?



**¡Muchas
gracias!**