Big data e inteligencias de las organizaciones

Ramon de la Rosa Falguera
PUE IT Manager
ramon.delarosa@pue.es

¿A qué nos dedicamos?

- PUE es una entidad privada creada en 1998 con la misión de **llevar la realidad del mercado laboral** al ámbito educativo.
- Gestionamos las **iniciativas académicas de las compañías y organizaciones de referencia en el mundo TIC**, diseñadas para acercar la formación y certificación oficial a las instituciones educativas de nuestro país.
- **Más de 1.000 centros adscritos a nuestro proyecto** y 100.000 estudiantes al año acreditan nuestro

TRAINING | **EDUCATION** | **CONSULTING** | **CERTIFICATION**

cloudera ACADEMIC PARTNER

hadoop

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

Iniciativas académicas

Linux Professional Institute

Microsoft Imagine Academy

Linux Professional Institute LINUX ESSENTIALS

RED HAT ACADEMY redhat.

cisco Networking Academy

CompTIA Authorized Academy

cloudera ACADEMIC PARTNER

vmware IT ACADEMY

ORACLE ACADEMY

Premios y reconocimientos obtenidos

cloudera ACADEMIC PARTNER

hadoop

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

¿Qué es Big Data?

cloudera ACADEMIC PARTNER

hadoop

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

¿Qué es Big Data?

Big Data nació con el objetivo de cubrir unas necesidades no satisfechas por las tecnologías existentes, como es el almacenamiento y tratamiento de grandes volúmenes de datos que poseen unas características muy concretas definidas como las tres **V's** (en la actualidad puede haber más).

cloudera ACADEMIC PARTNER

hadoop

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

¿Qué es Hadoop?

El 28 de enero de 2006 **Doug Cutting** creó el código considerado como la "***génesis***" de Hadoop, un ecosistema de código abierto que cambió, principalmente, la forma en la que las empresas almacenan, procesan y analizan los datos.

La procedencia del nombre es mucho menos técnica de lo que se podía esperar. El hijo de tres años de Cutting llamaba a su peluche Hadoop y así bautizó su inventor a la plataforma, que también tomaría de ahí su logo, un elefante amarillo.

¿Qué es Hadoop

- La principal diferencia con el resto de sistemas tradicionales es que Apache Hadoop permite que se ejecuten múltiples tipos de tareas analíticas con los mismos datos y en el mismo momento.
- **Hadoop** se inspiró en los documentos Google para **MapReduce** y **Google File System** (GFS).
- Enlaces a la documentación publicada por Google:
  - **The Google File System** Octubre 2003
    - **http://research.google.com/archive/gfs.html**
  - **MapReduce: Simplified Data Processing on Large Clusters** Diciembre 2004
    - **http://research.google.com/archive/mapreduce.html**

Cloudera Apache Hadoop

Apache Hadoop is an open-source software framework for distributed storage and distributed processing of very large data

Cloudera was the first commercial software vendor to release a Hadoop Distribution with enterprise features security and governance

Packages included are:
Apache Pig, Apache Hive, Apache HBase, Apache Spark, Apache ZooKeeper, Cloudera Impala, Apache Flume, Apache Sqoop, Apache Oozie, Solr

El Valor de Hadoop

Un lugar para almacenar datos de forma ilimitada

- Cualquier tipo de datos
- Más fuentes de datos
- Ingesta rápida y de grandes tamaños

Acceso unificado y multi-framework

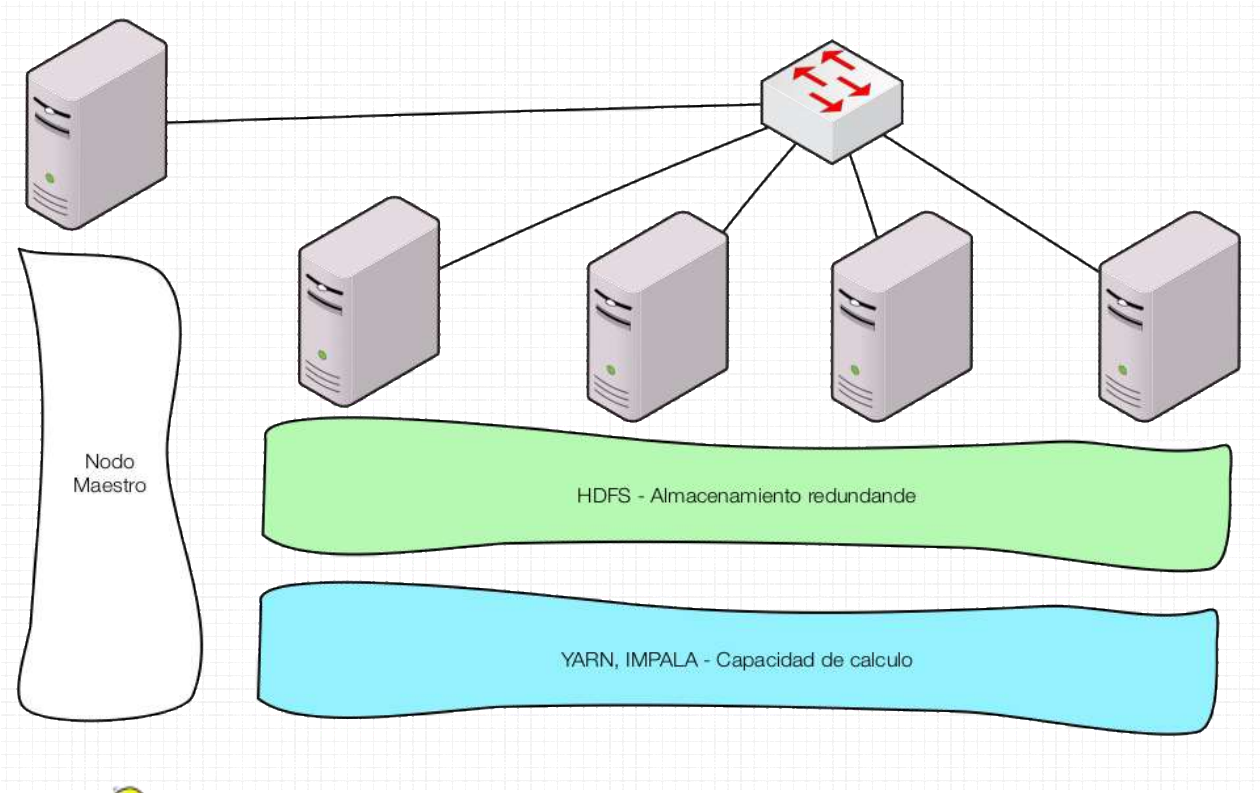
- Más usuarios
- Más herramientas
- Cambios rápidos

cloudera ACADEMIC PARTNER

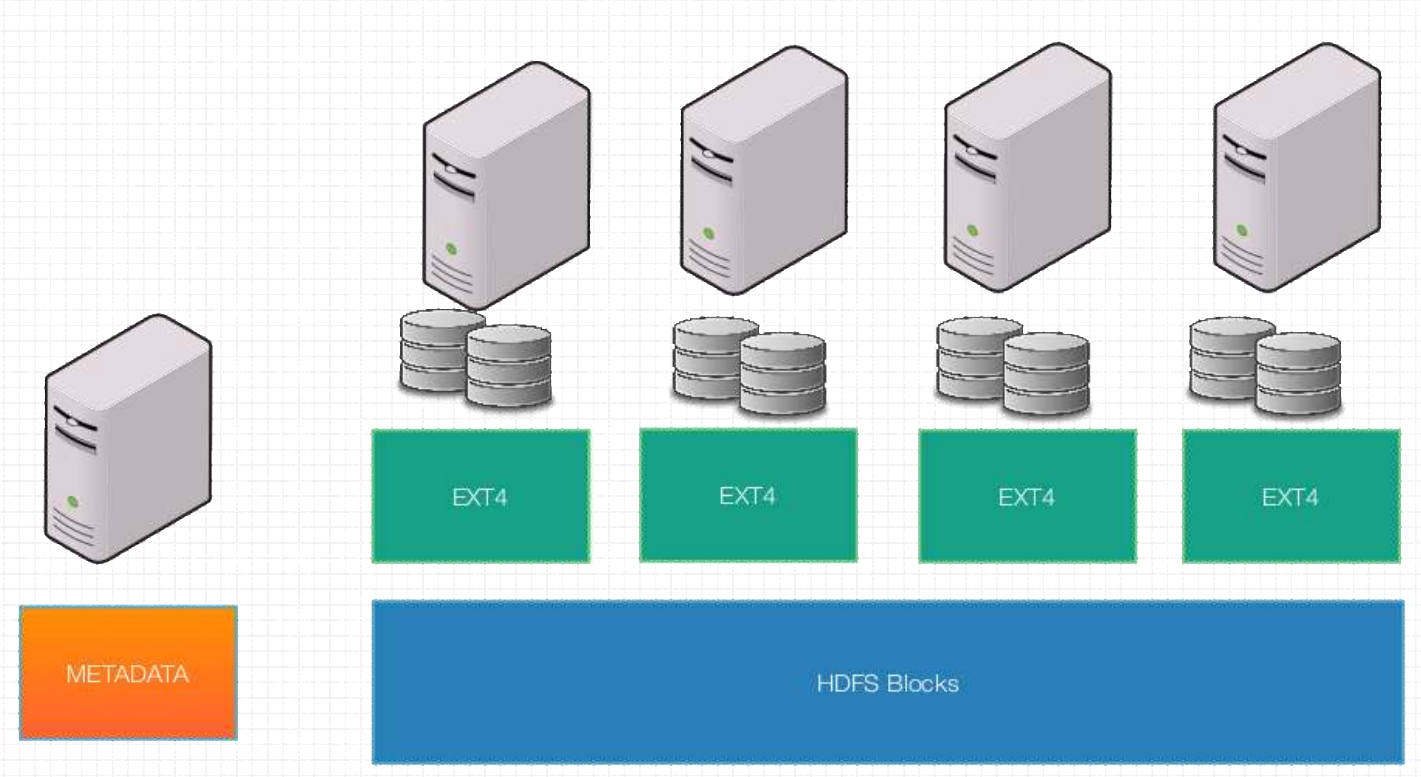
hadoop

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

¿Qué es un cluster hadoop?



HDFS data lake



Welcome to the data age

Instrumentación

Todo lo que pueda ser medido se medirá. Todos nosotros generamos datos pero las máquinas aún más.

Consumo

Los datos son la aplicación. Esperamos que los datos hagan mejor nuestras vidas, pero no con el coste de nuestra privacidad

Experiencia

Saberse impulsar con los datos es un imperativo para los negocios. Las organizaciones competitivas adoptan métodos ágiles.

cloudera ACADEMIC PARTNER

hadoop

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

Casos de uso por industria

Finanzas	Telco / Media	Manufactura	Retail / Transportes	Gobierno
• Cliente 360 • Fraude / Cyber • Compliance • Cálculo de riesgos • Store de datos operacioneles • Datos de mercado	• Cliente 360 • Reducir perdida de clientes • Optimización de la red • Monetización de datos • EDW • Streaming de media	• Dispositivos conectados: Coches, equipamiento … • Agilizar la cadena de suministro • Mantenimiento predictivo • IoT Data enabled “Smart Services” • Diagnostico	• Envíos a tienda • Agilizar la cadena de suministro • Next Best Offer • Connected Store • Cestas completas • IoT – Stores • Buque inteligente • Fidelidad del cliente	• Control de fronteras • Riesgo / inteligencia • Contribuyente 360 • Optimización de impuestos • Prevención de riesgos • Ciudadano 360

Nuevas profesiones en Big Data

- Arquitecto Big Data
- Desarrollador Big Data
- Data Analyst
- Científico de datos
- Administrador de Hadoop

SQL | Java | Python

Linux | Scala | Ansible

Kudu | Spark | Hbase

Hive | Impala | Hadoop

Kafka | NiFi

cloudera ACADEMIC PARTNER

hadoop

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

Cloudera Academy Program CAP

- Cursos
  - Introduction to Hadoop and Big Data
  - Developer Training for Spark and Hadoop
- Máquinas virtuales
  - 1 máquina virtual por curso simulando un cluster
  - Cloudera Quick Start Virtual Machine
- Licencia Cloudera Enterprise
- Más información: www.pue.es/cloudera-academy

cloudera ACADEMIC PARTNER hadoop

pue IMPULSANDO EL CONOCIMIENTO TIC CUALIFICADO

Demo - Spark

- Cargar un fichero de logs de Apache y analizar
- Ejecutado en local
- Ejecutado en un cluster


```
Last login: Thu Mar 21 06:37:36 on console
MacBook-Pro-de-Ramon:~ ramon$ cd demo_fp/
MacBook-Pro-de-Ramon:demo_fp ramon$
```


```
[ramon@puemaster1 ~]$
```

PUE ACADEMY Day

6a edición - Madrid 2019

Your learning experience

8 de mayo de 2019 - Madrid

https://www.pue.es/academy-day/2019/