
BIG DATA-HADOOP

Productos asociados a
BigData



Productos asociados a BigData

- **Hive**

- Permite acceder a HDFS como si fuera una Base de datos
- Podemos ejecutar comandos muy parecido a SQL para recuperar valores (HiveSQL)
- Esto simplifica enormemente el desarrollo y la gestión con Hadoop

<http://hive.apache.org/index.html>



Productos asociados a BigData

- **HBASE**

- Es el sistema de almacenamiento no relacional para Hadoop. Es de tipo columnar
- HBase es una base de datos de código abierto, distribuida y escalable para el almacenamiento de Big Data.
- Está escrita en Java e implementa y proporciona capacidades similares sobre Hadoop y HDFS.

El objetivo de este proyecto es el de trabajar con grandes tablas, miles de millones de filas de X millones de columnas, sobre un cluster Hadoop.

<http://hbase.apache.org/>



Productos asociados a BigData

- **PIG**

- Pig es un **lenguaje de alto de nivel** para analizar grandes volúmenes de datos.
- Pig trabaja en paralelo lo que permite gestionar gran cantidad de información
- Es un compilador que genera comandos MapReduce.
- Es un lenguaje textual denominado Pig Latin.

<https://pig.apache.org/>



Productos asociados a BigData

- **sqoop**
- Permite transferir gran volumen de datos de manera eficiente entre Hadoop y gestores de datos estructurados, como Bases de datos relacionales
- Sqoop ofrece conectores para integrar Hadoop con otros sistemas, **como por ejemplo Oracle o SqlServer**

<http://sqoop.apache.org/>



Productos asociados a BigData

- **Flume**

- Flume es un servicio distribuido y altamente eficiente para distribuir, agregar y recolectar grandes cantidades de información.
- Útil para cargar y mover en Hadoop información de tipo textuis como ficheros de logs, paquetes de twitter, etc.
- Tiene una arquitectura de tipo streaming con un flujo de datos muy potente y personalizables

<https://flume.apache.org/>



Productos asociados a BigData

- **Zookeeper**

- ZooKeeper es un servicio para mantener la configuración, coordinación y provisionamiento de aplicaciones distribuidas
- No solo vale para Hadoop, pero es muy útil en esa arquitectura
- Elimina la complejidad de la gestión distribuido de la plataforma

<https://zookeeper.apache.org/>



Productos asociados a BigData

- **Spark**

- Es un motor muy eficiente de procesamiento de datos a gran escala
- Implementa procesamiento en tiempo real al contrario que Map Reduce
- Es más rápida que MapReduce
- Trabaja de forma masiva en memoria
- Puede funcionar stand-alone

<http://spark.apache.org/>



Productos asociados a BigData

- **Otros**
- **Avro**: Sistema para serialización de datos
- **Cassandra**: base de datos multi-master muy potente
- **Mohout**: machine learning y Data Mining
- **Tez, Chuwka.....**

