# EJERCICIOS

## ECOSISTEMA HADOOP

### EJERCICIO 1

Rellena la siguiente tabla:

|  |  |  |
| --- | --- | --- |
| **NOMBRE** | **DESCRIPCIÓN** | **LOGOTIPO** |
| **Apache Hive** | Data warehouse que facilita la lectura, escritura y gestión de grandes conjuntos de datos almacenados en sistemas distribuidos, utilizando una sintaxis similar a SQL (HiveQL). |  |
| **Apache Pig** | Plataforma que permite analizar grandes conjuntos de datos mediante scripts escritos en Pig Latin, un lenguaje de alto nivel. | Apache Pig - Wikipedia |
| **Apache HBase** | Base de datos distribuida, no relacional, que permite el almacenamiento de grandes volúmenes de datos organizados por columnas, compatible con Apache Hadoop. | Apache HBase – Big Data Dummy |
| **Apache Flume** | Servicio que facilita la recolección, agregación y movimiento de grandes cantidades de datos generados de manera continua, como logs o eventos. | Get to Know Apache Flume from Scratch! - Analytics Vidhya |
| **Apache Sqoop** | Herramienta que permite la transferencia de datos entre bases de datos relacionales y Hadoop. |  |
| **Apache Oozie** | Coordinador de flujos de trabajo y servicios, utilizado para gestionar flujos de trabajo y dependencias en el ecosistema Hadoop. |  |
| **Apache ZooKeeper** | Servicio centralizado para gestionar configuraciones, sincronización y coordinación entre servicios distribuidos. | Apache ZooKeeper - Wikipedia, la enciclopedia libre |
| **Apache Storm** | Sistema para procesamiento de flujos de datos en tiempo real, con baja latencia, distribuido y tolerante a fallos. | Apache Storm - Wikipedia |
| **Apache Spark** | Framework para procesamiento de grandes volúmenes de datos, tanto en batch como en streaming, optimizado para alta velocidad y utilizando procesamiento en memoria. | Apache Spark - Wikipedia |
| **Apache Kafka** | Plataforma de mensajería distribuida que permite procesar flujos de datos en tiempo real, facilitando la publicación, suscripción y almacenamiento de datos de forma distribuida. | Apache Kafka |
| **Apache Atlas** | Herramienta para la gestión y gobernanza de metadatos en entornos de big data, facilitando la clasificación, catalogación y auditoría de datos. | Apache Atlas – Data Governance and Metadata framework for Hadoop |
| **Apache Mahout** | Herramienta para crear algoritmos de machine learning escalables, con un enfoque en análisis colaborativo, clasificación y clustering. | Apache Mahout |
| **Apache Phoenix** | Base de datos SQL sobre Apache HBase que permite consultas SQL nativas y OLTP (Online Transaction Processing). | Overview | Apache Phoenix |
| **Apache Impala** | Motor SQL de código abierto para procesar consultas SQL en tiempo real en Hadoop. Ofrece tiempos de respuesta muy rápidos en grandes volúmenes de datos. | Apache Hive vs Apache Impala: principales diferencias - Geekflare |
| **Apache Nifi** | Herramienta para automatizar el flujo de datos entre sistemas, con capacidades de transformación, enrutamiento y carga de datos en tiempo real. | Apache Nifi. Flujo de datos y ETL/ELT en Streaming. - Inteligencia  Artificial y Big Data |
| **Apache Ambari** | Plataforma de gestión de clústeres Hadoop que simplifica el aprovisionamiento, monitoreo y mantenimiento de clústeres de Apache Hadoop. | Ambari - |
| **Mapreduce** | Modelo de programación que permite el procesamiento distribuido de grandes cantidades de datos utilizando clústeres de computadoras. Los trabajos se dividen en tareas de mapeo y reducción para manejar datos a gran escala. | Hadoop Ecosystem | Hadoop Tools for Crunching Big Data | Edureka |