

Bases de datos: Hadoop

Nubia Suárez Torres

Universidad Central de Colombia
Facultad de Ingeniería
Maestría en analítica de datos

10 de septiembre de 2022



Contenido

- 1 Generalidades
- 2 Requerimientos
- 3 Usos
- 4 Ventajas
- 5 Desventajas
- 6 Diferencias
- 7 Referencias



Generalidades



Generalidades

- Hadoop es un entorno de trabajo para software, bajo licencia libre, para programar aplicaciones distribuidas que manejen grandes volúmenes de datos (big data).
- Apache Software Foundation (ASF) puso Hadoop a disposición del público en noviembre de 2012 como Apache Hadoop.
- Hadoop tiene dos capas centrales, a saber, la capa de procesamiento (MapReduce) y la capa de almacenamiento (sistema de archivos distribuido de Hadoop)



Requerimientos

Hadoop de ser de código abierto es compatible con todas las plataformas ya que está basado en Java.



Usos

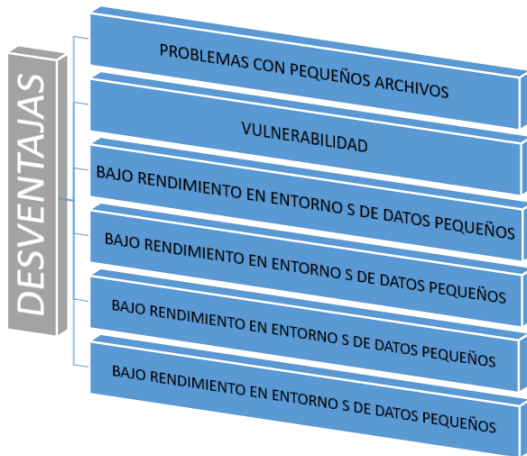
Hadoop está diseñado para trabajar con cualquier tipo de datos, ya sean estructurados, semiestructurados o no estructurados, lo que hace que sea muy flexible a la hora de trabajar con big data.



Ventajas



Desventajas




Diferencias

SQL funciona bien para el esquema estructurado, pero en cuanto a Big Data, no tenía un esquema fijo, sino que eran datos semiestructurados. RDBMS Las 3 V de Big Data: volumen, variedad y velocidad fueron la razón principal que llevó al advenimiento de las bases de datos NoSQL. Hadoop tiene una ventaja sobre SQL en este contexto.



Referencias

 Recuperado el 31 de agosto de
https://es.wikipedia.org/wiki/Apache_Hadoop

 Recuperado el 31 de agosto de
<https://aprenderbigdata.com/hadoop/>

 Recuperado el 31 de agosto de <https://es.acervolima.com/hadoop-caracteristicas-de-hadoop-que-lo-hacen-popular/>

