

+++



APACHE HBASE

+++

# ¿Que es HBASE?

- HBase es una base de datos NoSQL distribuida y escalable que se ejecuta sobre el sistema de archivos HDFS.  
(Los archivos HDFS (Hadoop Distributed File System) son parte integral del ecosistema Hadoop y se utilizan para almacenar y gestionar grandes conjuntos de datos de manera distribuida. )
- Es ideal para gestionar grandes cantidades de datos distribuidos en clústeres de servidores.



# Historia y Desarrollo

HBase se originó como un proyecto dentro de la comunidad Apache y ha experimentado varias versiones y mejoras a lo largo del tiempo.

2005

Inicio del Proyecto Apache Hadoop (2005):

El proyecto Apache Hadoop se inició en 2005 como un esfuerzo para crear una plataforma de software de código abierto para el procesamiento distribuido de grandes conjuntos de datos.

2007

Inicios de HBase (2007):

En 2007, durante el desarrollo de Hadoop, se identificó la necesidad de una base de datos distribuida que pudiera aprovechar la infraestructura de HDFS para proporcionar un acceso eficiente a grandes conjuntos de datos.

2008  
y  
2009

**Hitos Principales:**

En 2008, HBase se convirtió en un subproyecto de Apache Hadoop, lo que marcó el reconocimiento oficial de la importancia de esta tecnología dentro del ecosistema Hadoop.

2010

Liberación de Versiones y Mejoras

HBase ha continuado su evolución, con lanzamientos regulares que introducen nuevas características y mejoras. Se ha utilizado en una variedad de casos de uso, desde análisis en tiempo real hasta almacenamiento de datos a gran escala.





## Arquitectura Interna:



### División en Regiones:

Piensa en tus datos como un libro grande. Este libro se divide en secciones llamadas "regiones". Cada región contiene un rango específico de filas. Esta división ayuda a distribuir eficientemente los datos en diferentes partes del sistema.



### Nodos RegionServer:

Cada región se almacena y gestiona en un "nodo RegionServer". Puedes visualizar un RegionServer como una estantería en tu biblioteca que contiene varias secciones (regiones) del libro (datos).





## HMaster - Gestión de Regiones:

Ahora, imagina que el HMaster es como el bibliotecario principal. Este "bibliotecario" se encarga de asignar las diferentes secciones del libro (regiones) a los distintos estantes (RegionServers) para asegurarse de que todo esté organizado y accesible.



## ZooKeeper - Coordinación:

Piensa en ZooKeeper como el asistente personal del bibliotecario. Ayuda en la coordinación general y asegura que todos los estantes y libros estén en su lugar. Si algo cambia (como agregar o quitar un estante), ZooKeeper se asegura de que el bibliotecario esté al tanto.



## Características Clave

1

### Escalabilidad:

HBase escala horizontalmente al agregar más servidores al clúster, lo que permite manejar grandes volúmenes de datos.

2

### Consistencia y tolerancia

- La consistencia se mantiene mediante la replicación de datos en varios nodos.
- La tolerancia a fallos se logra mediante la replicación y la reasignación automática de regiones.

3

### Operaciones en tiempo real

HBase permite operaciones en tiempo real, como consultas y actualizaciones, haciendo que sea adecuado para aplicaciones que requieren respuestas rápidas.

## Ejemplo de uso de HBase:

- **Empresas de Tecnología:**

- a. Grandes empresas de tecnología, como Facebook y Twitter, han utilizado HBase para gestionar grandes cantidades de datos generados por usuarios, como mensajes, tweets y datos sociales.

- **Industria de Juegos en Línea:**

- a. Empresas en la industria de juegos en línea, como Tencent, han utilizado HBase para gestionar datos de juego en tiempo real y para proporcionar experiencias de juego más personalizadas.

- **Empresas Financieras:**

- a. Instituciones financieras, como Goldman Sachs, han implementado HBase para gestionar grandes conjuntos de datos financieros y realizar análisis de riesgos en tiempo real.

- **Investigación Científica:**

- a. En entornos de investigación científica, HBase se ha utilizado para almacenar y analizar datos generados por experimentos, observaciones y simulaciones.

- **Empresas de Publicidad en Línea:**

- a. Plataformas de publicidad en línea, como Criteo, han adoptado HBase para gestionar datos de clics, impresiones y realizar análisis en tiempo real para mejorar la orientación de anuncios.

- **Organizaciones Gubernamentales:**

- a. Algunas organizaciones gubernamentales han utilizado HBase para gestionar grandes volúmenes de datos gubernamentales, realizar análisis y mejorar la eficiencia en la toma de decisiones.

## Conclusiones y Futuro de HBase:



- HBase ha demostrado ser una solución valiosa para gestionar grandes volúmenes de datos de manera distribuida y escalable.
- Su flexibilidad en el modelo de datos y su capacidad para escalar horizontalmente lo hacen atractivo para una variedad de aplicaciones en tiempo real.





**Gracias.**