HBase

Apache HBase™ is the Hadoop database, a distributed, scalable, big data store.

You can use Apache HBase when you need random, realtime read/write access to your Big Data. This project's goal is the hosting of very large tables – billions of rows X millions of columns – atop clusters of commodity hardware. Apache HBase is an open-source, distributed, versioned, column-oriented store modeled after Google's Bigtable: A Distributed Storage System for Structured Data by Chang et al. Just as Bigtable leverages the distributed data storage provided by the Google File System, Apache HBase provides Bigtable-like capabilities on top of Hadoop and Hadoop-compatible filesystems, such as the MapR-FS.

Installing Apache HBase on a MapR cluster involves storing all HBase components in a single volume mapped to directory /hbase in the cluster. Tables are stored in a flat namespace, not grouped logically with related files. Because all Apache HBase data resides in one volume, only one set of storage policies can be applied to the entire Apache HBase datastore. Mirrors and snapshots of the HBase volume do not provide functional replication of the datastore. Despite this limitation, mirrors can be used to backup HLogs and HFiles in order to provide a recovery point for Apache HBase data.

This section contains documentation on working with HBase on the MapR distribution for Apache Hadoop. You can refer also to documentation available from the Apache HBase project. This section provides all relevant details about using HBase with MapR, but does not duplicate Apache documentation.

> The MapR filesystem provides native storage for table data, compatible with the HBase API. For new applications, consider using MapR tables for increased performance, more versatile table operations, and easier cluster administration. Click here for more information on table storage available in MapR Enterprise Database Edition.

Topics in This Section