

Cloudera Runtime 7.1.0

Apache HBase Overview

Date published: 2020-02-29

Date modified:

CLOUdera

<https://docs.cloudera.com/>

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

| | |
|--------------------------------------|----------|
| Overview of Apache HBase..... | 4 |
| Use cases for HBase..... | 4 |
| HBase on CDP..... | 4 |

Overview of Apache HBase

Apache HBase database provides random, persistent access to data. HBase is ideal for scenarios that require real-time analysis and tabular data for end-user applications.

As a natively nonrelational database, Apache HBase can combine data sources that use a wide variety of structures and schemas. HBase is designed for resilient data storage and is intended for hosting very large tables with sparse data. Apache HBase can host big data tables because it scales linearly to handle very large (petabyte scale), column-oriented data sets. The data store is predicated on a key-value model that supports low latency reads, writes, and updates in a distributed environment.

Use cases for HBase

As an operational data store, you can run your applications on top of HBase. You can also integrate your application with HBase. You can use HBase in CDP alongside your on-prem HBase clusters for disaster recovery use cases.

Some of the other use cases of HBase in CDP include:

- Support customer mission-important/mission-critical scale-out applications
- Query data with millisecond latency
- Perform fraud model serving and detection
- Enable serving analytics on mobile and web applications directly to end-customers
- Operationalize Artificial Intelligence/Machine Learning to drive revenue or manage operational cost
- Surfacing useful data in your applications (for example, customer 360 applications for customer support)
- Use as a key-value store for applications
- Bring together data spanning sources, schemas and data types and leverage in your applications
- Use as a small file store. You can use HBase to store logs from various devices into HBase

Related Information

[Getting Started: Operational database cluster](#)

[Importing data into HBase](#)

[Writing data to HBase](#)

[Using Apache HBase Backup and Disaster Recovery](#)

[Using HBase Replication](#)

HBase on CDP

If you are already using a CDH or HDP cluster, and want to start using HBase for new or existing use cases, you must understand the following differences.

How is HBase in CDP Different from HBase in CDH?

If you currently use HBase in CDH, and want to start using HBase in CDP:

- Configure Ranger ACLs in CDP corresponding to the HBase ACLs in your existing CDH cluster.
- Note that Ranger that is present in your Data Lake is used to provide HBase authentication.

How is HBase in CDP Different from HBase in HDP?

If you currently use HBase in HDP, and want to start using HBase in CDP:

- Configure Ranger ACLs in CDP corresponding to the HBase or Ranger ACLs in your existing HDP cluster.
- Migrate your applications to use the new HBase-Spark Connector because the Spark HBase Connector (SHC) is no longer supported in CDP.

Related Information

[Getting Started: Operational database cluster](#)

[HBase authorization using Apache Ranger](#)