

Cloudera Runtime 7.0.0

Using Amazon S3 with HBase

Date published: 2019-08-21

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

HBase Object Store Semantics Overview.....	4
HBase configuration to use S3 as a storage layer.....	4

HBase Object Store Semantics Overview

You can use Amazon S3 as a storage layer for HBase in a scenario where HFiles are written to S3, but WALs are written to HDFS.

The HBase Object Store Semantics(HBOSS) adapter bridges the gap between HBase, that assumes some file system operations are atomic, and object-store implementation of S3A filesystem, which does not provide atomic semantics to these operations natively.

HBOSS depends on S3Guard for accessing S3A buckets, so ensure given cluster and target AWS account fulfill all S3Guard requirements. The S3Guard feature guarantees a consistent view of data stored in Amazon S3 by storing additional metadata in a table residing in an Amazon DynamoDB instance. To enable S3Guard, you have to set up an Amazon DynamoDB database from Amazon Web Services. Amazon charges an hourly rate for this service based on the capacity you provision. See [Amazon DynamoDB Pricing](#).

HBOSS relies on Zookeeper locks for emulating atomic access to file system tree nodes. It can be deployed to use the same Zookeeper service that is already accessed by other HBase features.

HBase configuration to use S3 as a storage layer

A list of HBase configuration properties that are set when S3 is used as storage layer.

When an Operational Database cluster where S3 is used as a storage layer is created, the following configuration properties are automatically set to their default value, allowing HBase to use S3:

Property Name in Cloudera Manager	Configuration Property	Default Value
HDFS WAL Directory	hbase.wal.dir	hdfs://[NAME]:8020/hbase-wals
HDFS Root Directory	hbase.rootdir	s3a://[S3bucket]/[opdb-clustern-name]/hbase
HBoss S3A Implementation Class	fs.s3a.impl	org.apache.hadoop.hbase.oss.HBaseObjectStoreSemantics
HBoss Wrapped S3A Implementation Class	fs.hboss.fs.s3a.impl	org.apache.hadoop.fs.s3a.S3AFileSystem
HBoss Lock Implementation	fs.hboss.sync.impl	org.apache.hadoop.hbase.oss.sync.ZKTreeLockManager