Cloudera Runtime 7.0.0

Using Amazon S3 with HBase

Date published: 2019-08-21
Date modified:

CLOUDERA

Legal Notice

Contents

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