

Apache Atlas High Availability

Date published: 2020-07-28

Date modified: 2024-09-09



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

About Atlas High Availability.....	4
Prerequisites for setting up Atlas HA.....	4
Installing Atlas in HA using CDP Private Cloud Base cluster.....	4

About Atlas High Availability

Atlas supports multiple instances of the service in an active-passive configuration.

You can deploy multiple instances of the service on different physical hosts in a cluster. One of these instances is automatically selected as an 'active' instance to respond to user requests. All other instances are deemed 'passive'. If the 'active' instance is unavailable either because it is deliberately stopped, or due to an unexpected failure, one of the other instances automatically changes to an 'active' mode and starts servicing requests.

An 'active' instance is the only instance that can respond to user requests. A 'passive' instance redirects requests coming to it directly (using HTTP redirect) to the current 'active' instance. A passive instance does not service any requests. However, all instances (both active and passive), respond to admin status requests that return heartbeat information about that instance.

How Atlas High Availability works

Atlas uses Zookeeper to know about all available Atlas service instances within the cluster. It uses Zookeeper's electoral election to determine the instance that needs to be active. Once the leader is elected, other instances set themselves as passive and do not participate in servicing requests. Atlas thus achieves active-passive configuration for high availability (HA). The active instance is also the one that processes the notification messages from the hook.

When an instance goes down or is turned off, Zookeeper's electoral election starts and elects a new leader. ZooKeeper ensures that one instance is always available for processing requests.

The other server performs the real (active) work only when the "active" server goes down.

Currently, the only way to change the mode from active to passive is to shutdown the active mode; the other server instance receives a message (through ZooKeeper) to make itself active.

**Note:**

You cannot change an instance from active to passive mode using Cloudera Manager.

Prerequisites for setting up Atlas HA

You must ensure that the following prerequisites are met before you set up High Availability (HA) on Atlas:

- You must have a cluster with more than one node.
- You must have a CDP Private Cloud Base on-premise cluster.
- Atlas service must be up and running on that cluster.
- Apache ZooKeeper must be running on a cluster of machines. A minimum of three servers is recommended for the production environment.
- You must select two or more physical nodes for running the Atlas Web Service instances. These nodes define a server ensemble for Atlas.

Installing Atlas in HA using CDP Private Cloud Base cluster

When you are using the CDP Private Cloud Base cluster, you must install Atlas in HA using the Cloudera Manager UI.

Procedure

1. Log into Cloudera Manager and verify that Atlas service is running.

2. Stop the Atlas service.
3. Go to Clusters > Select Atlas > Instances > Add Role Instances > Add Atlas Service role to the required host and proceed with the wizard to complete the process.
4. Restart services for stale configurations.



Note: Cloudera recommends that the selected cluster in the Cloudera Manager must be Kerberos enabled but it is not a mandatory requirement.

5. Using the Cloudera Manager UI, verify that both Atlas services are running (services are green).

The active Atlas instance can be found using a HTTP status request to each Atlas instance:

- `http://<atlas_host_1>:3100/api/atlas/admin/status`
- `http://<atlas_host_2>:3100/api/atlas/admin/status`

The response for ACTIVE instance will be: `{"Status": "ACTIVE"}` The response for PASSIVE instance will be: `{"Status": "PASSIVE"}`

Atlas is currently running in the HA mode (active-passive mode)