Cloudera Runtime 7.2.11

Apache Atlas High Availability

Date published: 2020-09-02
Date modified: 2021-09-09

https://docs.cloudera.com/

Legal Notice

Contents

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 or Ambari UI.

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 either have a HDP or 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 HDP cluster

When you are using the HDP cluster, you must install Atlas in HA using the Ambari UI.

Procedure

1. Log into Ambari UI and verify that Atlas service is running.
2. Stop the Atlas service.

3. Navigate to the host page in Ambari where Atlas service is not installed and add an additional Atlas service.

   If the Ambari Infra Solr client is not installed on the host where another instance of Atlas is being installed, Ambari displays a pop-up window to add an Ambari Infra Solr client instance to the cluster.
4. Next, add the Atlas server instance by following step 3.
5. Start the Atlas service.
6. Using the Ambari UI, verify that both Atlas services are running.
7. Verify that both active and passive instances exist. An HTTP request on one of the Atlas instances must display its status as "Active" and the other instance as “Passive”.
   Atlas is currently running in the HA mode (active-passive mode)

Installing Atlas in HA

When you are using the CDP Public Cloud flavor, 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)