Sahara Ambari Plugin Documentation

Release 8.0.0.0rc2.dev2

Sahara team

Sep 13, 2022

CONTENTS

CHAPTER

ONE

USER GUIDE

1.1 Ambari Plugin

The Ambari sahara plugin provides a way to provision clusters with Hortonworks Data Platform on OpenStack using templates in a single click and in an easily repeatable fashion. The sahara controller serves as the glue between Hadoop and OpenStack. The Ambari plugin mediates between the sahara controller and Apache Ambari in order to deploy and configure Hadoop on OpenStack. Core to the HDP Plugin is Apache Ambari which is used as the orchestrator for deploying HDP on OpenStack. The Ambari plugin uses Ambari Blueprints for cluster provisioning.

1.1.1 Apache Ambari Blueprints

Apache Ambari Blueprints is a portable document definition, which provides a complete definition for an Apache Hadoop cluster, including cluster topology, components, services and their configurations. Ambari Blueprints can be consumed by the Ambari plugin to instantiate a Hadoop cluster on OpenStack. The benefits of this approach is that it allows for Hadoop clusters to be configured and deployed using an Ambari native format that can be used with as well as outside of OpenStack allowing for clusters to be re-instantiated in a variety of environments.

1.1.2 Images

For cluster provisioning, prepared images should be used.

Table 1: Support matrix for the *ambari* plugin

Version (image tag)	Distribution	Build method	Version (build parameter)	Notes
2.6	Ubuntu 16.04, CentOS 7	sahara-image-pack	2.6	uses Ambari 2.6
2.5	Ubuntu 16.04, CentOS 7	sahara-image-pack	2.5	uses Ambari 2.6
2.4	Ubuntu 14.04, CentOS 7	sahara-image-pack	2.4	uses Ambari 2.6
2.4	Ubuntu 14.04, CentOS 7	sahara-image-create	2.4	uses Ambari 2.2.1.0
2.3	Ubuntu 14.04, CentOS 7	sahara-image-pack	2.3	uses Ambari 2.4
2.3	Ubuntu 14.04, CentOS 7	sahara-image-create	2.3	uses Ambari 2.2.0.0

For more information about building image, refer to Sahara documentation.

HDP plugin requires an image to be tagged in sahara Image Registry with two tags: 'ambari' and '<plugin version>' (e.g. '2.5').

The image requires a username. For more information, refer to the registering image section of the Sahara documentation.

To speed up provisioning, the HDP packages can be pre-installed on the image used. The packages' versions depend on the HDP version required.

1.1.3 High Availability for HDFS and YARN

High Availability (Using the Quorum Journal Manager) can be deployed automatically with the Ambari plugin. You can deploy High Available cluster through UI by selecting `NameNode HA` and/or `ResourceManager HA` options in general configs of cluster template.

The NameNode High Availability is deployed using 2 NameNodes, one active and one standby. The NameNodes use a set of JournalNodes and Zookepeer Servers to ensure the necessary synchronization. In case of ResourceManager HA 2 ResourceManagers should be enabled in addition.

A typical Highly available Ambari cluster uses 2 separate NameNodes, 2 separate ResourceManagers and at least 3 JournalNodes and at least 3 Zookeeper Servers.

1.1.4 HDP Version Support

The HDP plugin currently supports deployment of HDP 2.3, 2.4 and 2.5.

1.1.5 Cluster Validation

Prior to Hadoop cluster creation, the HDP plugin will perform the following validation checks to ensure a successful Hadoop deployment:

- Ensure the existence of Ambari Server process in the cluster;
- Ensure the existence of a NameNode, Zookeeper, ResourceManagers processes HistoryServer and App TimeLine Server in the cluster

1.1.6 Enabling Kerberos security for cluster

If you want to protect your clusters using MIT Kerberos security you have to complete a few steps below.

- If you would like to create a cluster protected by Kerberos security you just need to enable Kerberos by checkbox in the `General Parameters` section of the cluster configuration. If you prefer to use the OpenStack CLI for cluster creation, you have to put the data below in the `cluster_configs` section:

```
"cluster_configs": {
  "Enable Kerberos Security": true,
}
```

  Sahara in this case will correctly prepare KDC server and will create principals along with keytabs to enable authentication for Hadoop services.

- Ensure that you have the latest hadoop-openstack jar file distributed on your cluster nodes. You can download one at `https://tarballs.openstack.org/sahara-extra/dist/`
- Sahara will create principals along with keytabs for system users like `oozie`, `hdfs` and `spark` so that you will not have to perform additional auth operations to execute your jobs on top of the cluster.

1.1.7 Adjusting Ambari Agent Package Installation timeout Parameter

For a cluster with large number of nodes or slow connectivity to HDP repo server, a Sahara HDP Cluster creation may fail due to ambari agent reaching the timeout threshold while installing the packages in the nodes.

Such failures will occur during the "cluster start" stage which can be monitored from Cluster Events tab of Sahara Dashboard. The timeout error will be visible from the Ambari Dashboard as well.

- To avoid the package installation timeout by ambari agent you need to change the default value of `Ambari Agent Package Install timeout` parameter which can be found in the `General Parameters` section of the cluster template configuration.

CHAPTER

TWO

CONTRIBUTOR GUIDE

2.1 So You Want to Contribute...

For general information on contributing to OpenStack, please check out the contributor guide to get started. It covers all the basics that are common to all OpenStack projects: the accounts you need, the basics of interacting with our Gerrit review system, how we communicate as a community, etc.

sahara-plugin-ambari is maintained by the OpenStack Sahara project. To understand our development process and how you can contribute to it, please look at the Sahara project's general contributor's page: http://docs.openstack.org/sahara/latest/contributor/contributing.html