
Sahara Vanilla Plugin Documentation

Release 10.0.1.dev1

Sahara team

Oct 19, 2023

CONTENTS

- 1 User Guide** **1**
- 1.1 Vanilla Plugin 1
- 2 Contributor Guide** **5**
- 2.1 So You Want to Contribute... 5

1.1 Vanilla Plugin

The vanilla plugin is a reference implementation which allows users to operate a cluster with Apache Hadoop.

Since the Newton release Spark is integrated into the Vanilla plugin so you can launch Spark jobs on a Vanilla cluster.

1.1.1 Images

For cluster provisioning, prepared images should be used.

Table 1: Support matrix for the *vanilla* plugin

Version (image tag)	Distribution	Build method	Version (build parameter)	Notes
2.8.2	Ubuntu 16.04, CentOS 7	sahara-image-create	2.8.2	Hive 2.3.2, Oozie 4.3.0
2.7.5	Ubuntu 16.04, CentOS 7	sahara-image-create	2.7.5	Hive 2.3.2, Oozie 4.3.0
2.7.1	Ubuntu 16.04, CentOS 7	sahara-image-create	2.7.1	Hive 0.11.0, Oozie 4.2.0

For more information about building image, refer to [Sahara documentation](#).

Vanilla plugin requires an image to be tagged in Sahara Image Registry with two tags: 'vanilla' and '<hadoop version>' (e.g. '2.7.1').

The image requires a username. For more information, refer to the [registering image](#) section of the Sahara documentation.

Build settings

When `sahara-image-create` is used, you can override few settings by exporting the corresponding environment variables before starting the build command:

- `DIB_HADOOP_VERSION` - version of Hadoop to install
- `HIVE_VERSION` - version of Hive to install
- `OOZIE_DOWNLOAD_URL` - download link for Oozie (we have built Oozie libs here: <https://tarballs.openstack.org/sahara-extra/dist/oozie/>)
- `SPARK_DOWNLOAD_URL` - download link for Spark

1.1.2 Vanilla Plugin Requirements

The image building tools described in [Building guest images](#) add the required software to the image and their usage is strongly suggested. Nevertheless, here are listed the software that should be pre-loaded on the guest image so that it can be used to create Vanilla clusters:

- ssh-client installed
- Java (version ≥ 7)
- Apache Hadoop installed
- 'hadoop' user created

See [Swift Integration](#) for information on using Swift with your sahara cluster (for EDP support Swift integration is currently required).

To support EDP, the following components must also be installed on the guest:

- Oozie version 4 or higher
- mysql/mariadb
- hive

1.1.3 Cluster Validation

When user creates or scales a Hadoop cluster using a Vanilla plugin, the cluster topology requested by user is verified for consistency.

Currently there are the following limitations in cluster topology for Vanilla plugin:

For Vanilla Hadoop version 2.x.x:

- Cluster must contain exactly one namenode
- Cluster can contain at most one resourcemanager
- Cluster can contain at most one secondary namenode
- Cluster can contain at most one historyserver
- Cluster can contain at most one oozie and this process is also required for EDP
- Cluster can't contain oozie without resourcemanager and without historyserver
- Cluster can't have nodemanager nodes if it doesn't have resourcemanager

- Cluster can have at most one hiveserver node.
- Cluster can have at most one spark history server and this process is also required for Spark EDP (Spark is available since the Newton release).

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-vanilla 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>