Enterprise Steam Installation and Setup

Release 1.4.9.120

H2O.ai

May 21, 2019
## CONTENTS

1 Installing and Upgrading Enterprise Steam 3
   1.1 Before You Begin the Installation 3
   1.2 Pre-Installation Worksheet 3
   1.3 Obtaining the License Key 5
   1.4 Ubuntu Installation 6
   1.5 RHEL Installation 10
   1.6 Upgrading Enterprise Steam 13

2 Starting Enterprise Steam 15
   2.1 Uploading a License File 16

3 Configurations 19
   3.1 Authentication 20
   3.2 Profiles 24
   3.3 Users 27
   3.4 Roles 31
   3.5 YARN 34
   3.6 Engines 36
   3.7 Spark 37
   3.8 Token 37

4 Clusters and Notebooks 39
   4.1 Launching a New Cluster 39
   4.2 Deleting Clusters 43
   4.3 Notebooks 44
Enterprise Steam is a service for securely starting and connecting to H2O YARN jobs in a Hadoop environment. Enterprise Steam offers security, resource control, and resource monitoring out of the box in a multi-tenant architecture so that organizations can focus on the core of their data science practice. Enterprise Steam enables streamlined H2O adoption in a secure manner that complies with company policies.

For data scientists, Enterprise Steam provides easy R/Python APIs and a Web UI for starting H2O YARN jobs and allows data scientists to practice data science in their own H2O cluster.

For Admins, Enterprise Steam provides control over which H2O versions are available and the YARN queues to use. Admins can also cap resources that data scientists can use.

This document applies to Administrators and describes how to install and start Enterprise Steam in a Hadoop environment and make it accessible to a set of users. The process includes uploading an H2O driver and adding users through either the built-in SQLite database or through an LDAP config file.

**Note:** Before you begin using Enterprise Steam, be sure that your minimum version of H2O is 3.10.4.1. If necessary, follow the instructions on the [H2O Download page](#) for your platform to upgrade H2O. For current customers with enterprise support, earlier versions can be supported. Contact H2O.ai if you require support for an earlier version.
CHAPTER ONE

INSTALLING AND UPGRADING ENTERPRISE STEAM

Enterprise Steam is supported on Ubuntu and Red Hat Enterprise Linux. Be sure to follow the instructions for your platform:

1.1 Before You Begin the Installation

1.1.1 General Information

Please review the following information before you begin installing Enterprise Steam.

- A Hadoop admin is needed for the install in order to change the core-site.xml configuration.
- Admins should verify whether their Hadoop environment requires sudo. If it does, then users must have a root password/root access.
- This installation creates a SQLite database.
- The installation creates a new system user called “steam”.
- Enterprise Steam requires any two ports to be open before installation. Ports used by Enterprise Steam are not automatically opened via firewall. Follow your regular process for opening ports for Enterprise Steam.
- If configuring LDAP, be sure that you have a good understanding of LDAP groups in order to get the LDAP connection strings to be configured in the Steam UI.
- Keep the LDAP authentication certificate, if that exists in your enterprise.

1.2 Pre-Installation Worksheet

The questions below are meant to assist you with the Enterprise Steam installation process. For a smoother installation process, please print and fill this out so that you have this information ready when installing Enterprise Steam.

1.2.1 General Information

1. Who is your Hadoop administrator?

2. Who is your IT firewall administrator?

3. What is the Hadoop distribution that H2O is running on (for example, HDP2.4 or CDH5.5 or MAPR 5.1)?
4. What are the IP address and hostname of the server where Enterprise Steam will run?

5. What TCP port will Enterprise Steam use (recommended 443)?

6. Does the Enterprise Steam port need to be opened through a firewall?

7. What OS version is the above server running (for example, RHEL 6.7 or Ubuntu 12.04) ["cat /etc/redhat-release" or "cat /etc/lsb-release"]?

8. Do you have root login access to the server named above (required for installation of the package using rpm or dpkg)?

9. What is the service ID that will be used to run Enterprise Steam?

10. Will you provide a certificate and private key or generate a self-signed certificate?

11. Will user passwords be authenticated by LDAP/AD or an internal Steam database?

12. Does your environment use Kerberos authentication?

1.2.2 For LDAP/AD Users

1. Who is your LDAP/AD administrator?

2. What is the LDAP/AD host IP and port?

3. Does your LDAP/AD setup use SSL?

4. What is the Bind Distinguished Name (DN) used by the LDAP/AD server if extended access is required (for example, cn=admin,dc=0xdata,dc=loc)? Note that you can also use anonymous bind.

5. What is the User Base DN or the location of the LDAP/AD users, specified by the DN of your user subtree (for example, ou=users,dc=0xdata,dc=loc)?

6. What will be the LDAP/AD search filter used to filter users (for example, department=IT)?
7. What is the User Attribute that contains the user name (for example, uid)?

8. What is the Group DN or the Distinguished Name used for group synchronization (for example cn=jettygroup,ou=groups,dc=0xdata,dc=loc)?

9. What is the Group Base DN or the location of your LDAP/AD groups specified by the DN of your subtree (for example, ou=groups,dc=0xdata,dc=loc)?

10. What is the Group Attribute that contains the user name (for example, cn)?

11. What is the attribute for static group entries (for example, memberUid)?

12. What is the desired search request size limit? Note that you can also specify 0 for unlimited size.

13. What is the desired search request time limit? Note that you can also specify 0 for unlimited time.

1.2.3 For Kerberos Users

1. Who is your Kerberos administrator?

2. What is the Enterprise Steam Principal?

3. Does your Kerberos environment use multiple realms? If so, what is the Realm the Enterprise Steam Principal will authenticate to?

4. Do you have a keytab file for the Enterprise Steam Principal?

5. Does the Enterprise Steam service ID have Hadoop core-site.xml settings to run as a proxyuser (required)?

1.3 Obtaining the License Key

Contact H2O Sales to obtain a license key. The Enterprise Steam Admin should save this license file on his/her local machine. The Enterprise Steam Admin will be prompted to enter this license key the first time that Enterprise Steam is started. (See Uploading a License File section.)
1.4 Ubuntu Installation

This section describes how to install Enterprise Steam on Ubuntu.

1.4.1 Requirements for Enterprise Steam with Ubuntu

- Ubuntu 12.04 or greater
- Enterprise Steam .deb file. This is available on the Steam download page.
- Chrome version 50+ browser with an Internet connection. Note that Chrome is currently the only supported browser, and the minimum version is 50.
- H2O driver for your version of Hadoop. This is available from the H2O Download page. Click the Install on Hadoop tab, and select the correct version for your environment.
- HAProxy 1.5 or greater. For Ubuntu, this is available from haproxy.debian.net.

Optional

- Sparkling Water package (available from https://www.h2o.ai/download/#sparkling-water). Note that the minimum supported versions for enabling Sparkling Water in Enterprise Steam are:
  - 2.1.41
  - 2.2.27
  - 2.3.16
  - 2.4.1
- Enterprise Steam R Package (available from the Enterprise Steam Download page on the STEAM API tab)
- Enterprise Steam Python Package (available from the Enterprise Steam Download page on the STEAM API tab)
- PEM certificate file
- PEM privatekey file
- Your own Enterprise Steam Principal Keytab file, if Keberos is enabled.

1.4.2 Install HAProxy for Ubuntu

This section describes how to install HAProxy 1.5. You can skip this section if your environment already has HAProxy 1.5 or greater.

1. In your browser, go to https://haproxy.debian.net.
2. Select the system and version that you are running, then select an HAProxy version of 1.5-stable or greater.
3. Open a Terminal window and run the commands that are listed (using sudo if required). The example below shows the commands to use with Ubuntu version Trusty (14.04 LTS) and HAProxy version 1.7-stable.
1.4.3 Install Enterprise Steam on Ubuntu

1. On your local machine, download the Enterprise Steam .deb from the Steam download page.
2. Review and accept the terms of the EULA.
3. Open a terminal window and ssh to your Hadoop edge node.

   `ssh <user>@<hadoop_edge_node>`

4. Copy the Enterprise Steam .deb file to your edge node.

   `scp <user>@<hadoop_edge_node>:./esteam_1.0.0_amd64.deb .`

5. Install the Enterprise Steam .deb file.

   `sudo dpkg -i esteam_1.0.0_amd64.deb`

6. Set the administrator username and password.

   `sudo service steam set-admin
   username: administrator
   password: ***********`
7. The Enterprise Steam installation requires the following updates to the Hadoop coresite.xml. These changes provide the hosts that proxyuser can be a superuser on. These changes also provide for the case where superuser is someone who can run H2O on behalf of another user. Note that this step is typically performed by a Hadoop engineer.

```xml
<property>
  <name>hadoop.proxyuser.steam.hosts</name>
  <value>host1,host2</value>
</property>
<property>
  <name>hadoop.proxyuser.steam.groups</name>
  <value>group1,group2</value>
</property>
<property>
  <name>hadoop.proxyuser.steam.users</name>
  <value>user1,user2</value>
</property>
```

where:
- • host1, host2 are the hostnames of the machines. Separate multiple hostnames with commas.
- • group1, group2 are the group IDs. Separate multiple group IDs with commas.
- • user1, user2 are the user IDs. Separate multiple user IDs with commas.

Note: In most cases, you will set either the proxyuser groups or proxyuser users. You are not required to set both.

Additional information about these changes is available here: https://hadoop.apache.org/docs/r2.7.3/hadoop-project-dist/hadoop-common/Superusers.html.

8. (Optional) Install the certificate and private key for the Enterprise Steam server by adding these in /etc/steam/private_key.pem, /etc/steam/cert.pem.

9. Optionally make changes in the /etc/steam/steam.yaml file. Below is a sample steam.yaml file showing the available configuration options.

```yaml
# Working directory points to directory for Steam assets
STEAM_WORKING_DIRECTORY: /opt/h2oai/steam/var/master

# Directory of Steam and Hadoop temporary files. Defaults to your OS's temp directory.
STEAM_TMP_DIR: /tmp

# Certificate and private key PEM files used by both the Steam process and Steam's haproxy process.
# If not present a self-signed certificate will be autogenerated by Steam.
STEAM_WEB_TLS_CERT_PATH: /etc/steam/cert.pem
STEAM_WEB_TLS_PRIVATE_KEY_PATH: /etc/steam/private_key.pem

# Choose minimal crypto protocol: ss13, tls10, tls11, tls12
STEAM_WEB_MIN_CRYPTO_PROTOCOL: tls11

# Steam itself uses this port.
STEAM_HTTPS_PORT: 9000

# Steam will save application logs into this directory
STEAM_LOG_DIR: /var/log/steam
```

(continues on next page)
# Unix permission of the log files
STEAM_LOG_PERMISSIONS: 0644

# Uncomment the following security related values to enable kerberized/
→ mapraccess ticket access to hadoop
# STEAM_MAPR_TICKETS_ENABLED: FALSE
# STEAM_MAPR_TICKET_DIR: /opt/h2oai/steam/mapr
# STEAM_MAPR_SERVICE_NAME: steam
#
# STEAM_KERBEROS_ENABLED: FALSE
# STEAM_KERBEROS_PRINCIPAL: pcpl@REALM
# STEAM_KERBEROS_KEYTAB_PATH: /home

# These configuration options modify Yarn/Hadoop implementations
# Coerce flags change the casing of usernames in regards to impersonation
→ for case sensitive implementations.
# STEAM_COERCE_USERNAMES_LOWER: FALSE

# Steam log level can be set to (0 - Panic level, 1 - Fatal level, 2 - Error
→ level, 3 - Warning level, 4 - Info level, 5 - Debug level)
STEAM_LOG_LEVEL: 4

# The HTTP Strict-Transport-Security response header is a security feature
→ that lets a web site tell browsers that it should only be communicated
→ with using HTTPS, instead of using HTTP.
# Value is in Seconds, the default value is equivalent to 20 years.
# Set to empty to disable.
SERVER_STRICT_TRANSPORT: max-age=631138519

# The HTTP X-XSS-Protection response header is a feature of Internet
→ Explorer, Chrome and Safari that stops pages from loading when they detect
→ reflected cross-site scripting (XSS) attacks.
# When value is set to 1 and a cross-site scripting attack is detected, the
→ browser will sanitize the page (remove the unsafe parts)
SERVER_X_XSS_PROTECTION: 0

# Content Security Policy (CSP) is an added layer of security that helps to
→ detect and mitigate certain types of attacks, including Cross Site
→ Scripting (XSS) and data injection attacks.
SERVER_CONTENT_SECURITY_POLICY: style-src 'self' 'unsafe-inline' https://
→ fonts.googleapis.com; font-src 'self' https://fonts.gstatic.com data;;

# You might need to set it to true if Steam is being too strict about your
→ SAML response
SAML_INSECURE_ALLOWED: FALSE

# Set to true to disable the build-in admin user from logging-in
SAML_DISABLE_ADMIN: FALSE

# Set how long to wait before timing out idle web session
STEAM_WEB_UI_TIMEOUT_MIN: 480

# Used to turn off embedded Jupyterhub
STEAM_SW_DISABLE_JUPYTER: FALSE

10. (Optional) If your environment uses Kerberos authentication, then uncomment the Kerberos related values in
At this point, you are ready to Start Enterprise Steam.

### 1.5 RHEL Installation

This section describes how to install Enterprise Steam on Red Hat Enterprise Linux.

#### 1.5.1 Requirements for Enterprise Steam with RHEL

- RHEL 6.7 or greater. Note that HAProxy is already included with this version of Red Hat.
- Enterprise Steam .rpm file. This is available from the Steam download page.
- Chrome version 50+ browser with an Internet connection. Note that Chrome is currently the only supported browser, and the minimum version is 50.
- H2O driver for your version of Hadoop. This is available from the H2O Download page. Click the Install on Hadoop tab, and select the correct version for your environment.

**Optional**

- Sparkling Water package (available from [https://www.h2o.ai/download/#sparkling-water](https://www.h2o.ai/download/#sparkling-water)). Note that the minimum supported versions for enabling Sparking Water in Enterprise Steam are:
  - 2.1.41
  - 2.2.27
  - 2.3.16
  - 2.4.1
- Enterprise Steam R Package (available from the Enterprise Steam Download page on the STEAM API tab)
- Enterprise Steam Python Package (available from the Enterprise Steam Download page on the STEAM API tab)
- PEM certificate file
- PEM privatekey file
- Your own Enterprise Steam Principal Keytab file, if Keberos is enabled.

#### 1.5.2 Install HAProxy on RHEL

RHEL 6.7 or greater includes HAProxy. Run the following command if you have not already installed HAProxy. Note that SSL must be enabled before you run this command.

```
sudo yum install haproxy
```
1.5.3 Install Enterprise Steam on RHEL

1. On your local machine, download the Enterprise Steam .rpm file from the Steam download page.
2. Review and accept the terms of the EULA.
3. Open a terminal window and ssh to your Hadoop edge node.
   ```
   ssh <user>@<hadoop_edge_node>
   ```
4. Copy the Enterprise Steam .rpm file to your edge node.
   ```
   scp <user>@<hadoop_edge_node>:./esteam_1.0.0_amd64.rpm .
   ```
5. Install the Enterprise Steam .rpm file.
   ```
   sudo rpm -i <esteam_rpm_package>
   ```
6. Set the administrator username and password.
   - On RHEL 6: `sudo /etc/init.d/steam set-admin`
   - On RHEL 7: `sudo su -s /bin/bash -c "/opt/h2oai/steam/steam set admin" steam`
7. The Enterprise Steam installation requires the following updates to the Hadoop coresite.xml. These changes provide the hosts that proxyuser can be a superuser on. These changes also provide for the case where superuser is someone who can run H2O on behalf of another user. Note that this step is typically performed by a Hadoop engineer.
   ```xml
   <property>
     <name>hadoop.proxyuser.steam.hosts</name>
     <value>host1,host2</value>
   </property>
   <property>
     <name>hadoop.proxyuser.steam.groups</name>
     <value>group1,group2</value>
   </property>
   <property>
     <name>hadoop.proxyuser.steam.users</name>
     <value>user1,user2</value>
   </property>
   ```
   where:
   - `host1,host2` are the hostnames of the machines. Separate multiple hostnames with commas.
   - `group1,group2` are the group IDs. Separate multiple group IDs with commas.
   - `user1,user2` are the user IDs. Separate multiple user IDs with commas.
   **Note:** In most cases, you will set either the proxyuser groups or proxyuser users. You are not required to set both.
   Additional information about these changes is available here: [https://hadoop.apache.org/docs/r2.7.3/hadoop-project-dist/hadoop-common/Superusers.html](https://hadoop.apache.org/docs/r2.7.3/hadoop-project-dist/hadoop-common/Superusers.html).
8. (Optional) Install the certificate and private key for the Enterprise Steam server by adding these in `/etc/steam/private_key.pem`, `/etc/steam/cert.pem`.
9. Optionally make changes in the `/etc/steam/steam.yaml` file. Below is a sample steam.yaml file showing the available configuration options.

---

1.5. RHEL Installation
# Working directory points to directory for Steam assets
STEAM_WORKING_DIRECTORY: /opt/h2oai/steam/var/master

# Directory of Steam and Hadoop temporary files. Defaults to your OS's temp directory.
# STEAM_TMP_DIR: /tmp

# Certificate and private key PEM files used by both the Steam process and Steam's haproxy process.
# If not present a self-signed certificate will be autogenerated by Steam.
STEAM_WEB_TLS_CERT_PATH: /etc/steam/cert.pem
STEAM_WEB_TLS_PRIVATE_KEY_PATH: /etc/steam/private_key.pem

# Choose minimal crypto protocol: ssl3, tls10, tls11, tls12
STEAM_WEB_MIN_CRYPTO_PROTOCOL: tls11

# Steam itself uses this port.
STEAM_HTTPS_PORT: 9000

# Steam will save application logs into this directory
STEAM_LOG_DIR: /var/log/steam

# Unix permission of the log files
STEAM_LOG_PERMISSIONS: 0644

# Uncomment the following security related values to enable kerberized/maprticket access to hadoop
# STEAM_MAPR_TICKETS_ENABLED: FALSE
# STEAM_MAPR_TICKET_DIR: /opt/h2oai/steam/mapr
# STEAM_MAPR_SERVICE_NAME: steam
#
# STEAM_KERBEROS_ENABLED: FALSE
# STEAM_KERBEROS_PRINCIPAL: pcpl@REALM
# STEAM_KERBEROS_KEYTAB_PATH: /home

# These configuration options modify Yarn/Hadoop implementations
# Coerce flags change the casing of usernames in regards to impersonation for case sensitive implementations.
# STEAM_COERCE_USERNAMES_LOWER: FALSE

# Steam log level can be set to (0 - Panic level, 1 - Fatal level, 2 - Error level, 3 - Warning level, 4 - Info level, 5 - Debug level)
STEAM_LOG_LEVEL: 4

# The HTTP Strict-Transport-Security response header is a security feature that lets a web site tell browsers that it should only be communicated with using HTTPS, instead of using HTTP.
# Value is in Seconds, the default value is equivalent to 20 years.
# Set to empty to disable.
SERVER_STRICT_TRANSPORT: max-age=631138519

# The HTTP X-XSS-Protection response header is a feature of Internet Explorer, Chrome and Safari that stops pages from loading when they detect reflected cross-site scripting (XSS) attacks.
# When value is set to 1 and a cross-site scripting attack is detected, the browser will sanitize the page (remove the unsafe parts)
SERVER_X_XSS_PROTECTION: 0

(continues on next page)
# Content Security Policy (CSP) is an added layer of security that helps to detect and mitigate certain types of attacks, including Cross Site Scripting (XSS) and data injection attacks.


# You might need to set it to true if Steam is being too strict about your SAML response
SAML_INSECURE_ALLOWED: FALSE

# Set to true to disable the build-in admin user from logging-in
SAML_DISABLE_ADMIN: FALSE

# Set how long to wait before timing out idle web session
STEAM_WEB_UI_TIMEOUT_MIN: 480

# Used to turn off embedded Jupyterhub
STEAM_SW_DISABLE_JUPYTER: FALSE

10. (Optional) If your environment uses Kerberos authentication, then uncomment the Kerberos related values in /etc/steam/steam.yaml. Be sure to also specify the correct Kerberos principal and path to the keytab file. Note that you may also be required to add another proxyuser configuration. Refer to https://hadoop.apache.org/docs/current/hadoop-kms/index.html#KMS_Proxyuser_Configuration for more information.

At this point, you are ready to Start Enterprise Steam.

### 1.6 Upgrading Enterprise Steam

#### 1.6.1 Upgrade from version 0.9.2.14 or Older

**Server**

1. Backup /opt/h2oai/steam and /etc/steam
2. Install the Enterprise Steam package
   - RPM: yum localinstall esteam-1.1.0.3....rpm
   - DEB: dpkg -i esteam-1.1.0.3....deb
3. Configure Enterprise Steam by editing /etc/steam/steam.yaml
   - Verify that the paths to TLS certificate and key are correct
   - (optional) Set kerberized/maprticket access to Hadoop.
   - (optional) Set desired log properties.
   - Note that the steam.conf file is no longer used.
4. Run sudo service steam restart.
   - For troubleshooting, see the logs at /var/log/steam/steam.log
Admin Client

1. Clear browser cookies and cache
2. Visit the Enterprise Steam homepage and login as admin
3. (optional) To set up LDAP cache, go to Configuration > User authentication and set the Cache Max Age parameter. Test and save the config.

Regular Clients - Web

1. Clear browser cookies and cache
2. Visit the Enterprise Steam homepage and login.

Regular Clients - Python/R

Install new Python/R Enterprise Steam packages

- Python: pip install h2osteam-1.1.0-py2.py3-none-any.whl
- R: R CMD INSTALL h2osteam_1.1.0.tar.gz

1.6.2 Upgrade from 0.9.2.16 or Newer

The steps for upgrading from 0.9.2.16 or newer are the same as the steps above, but you may omit Server Step 3.
1. The commands to start Enterprise Steam vary depending on your OS.

**RHEL 7**: Start Enterprise Steam by running the following commands on your YARN edge node.

```
sudo systemctl enable steam
sudo systemctl start steam
```

**RHEL 6 and Ubuntu**: Start Enterprise Steam by running the following command on your YARN edge node.

```
sudo /etc/init.d/steam start
```

2. (Optional) Check the log file to verify that Enterprise Steam starts correctly:

```
sudo cat /var/log/steam/steam.log
```

3. Open a Chrome browser and navigate to your Hadoop edge node (where Enterprise Steam is currently running). For example, `https://<hadoop-edge-node>:9000`. Log in using the Administrator username and password that you created during the installation process. (Note that you can also change your password from this page.)

![Welcome to Enterprise Steam](image)

Upon successful completion, the Welcome page will display.
2.1 Uploading a License File

A license is required in order to run Enterprise Steam. The first time you log in to Enterprise Steam, a message will display in the upper-left corner prompting you to enter your license key. This license is obtained from H2O Sales.

1. Click the “Please updated your license” link in the upper-left corner to jump to the Configurations > Licensing page.

2. Click the Browse button, and navigate to your Enterprise Steam license file.
2.1. Uploading a License File
The Configurations page allows Enterprise Steam Admins to add, edit, and deactivate users and roles. Users can be added either individually using the Enterprise Steam SQLite database or through an existing LDAP directory.

Note: Only Admins have access to the Configurations page.

The Configurations page consists of the following tabs.

- The **Users** tab shows the current list of users and their assigned role(s).
- The **Roles** tab provides a table of the permissions assigned to each role.
- The **Profiles** tab allow you to define the size of clusters, the minimum and maximum memory, and the number of cores, threads, and queues for the profile.
- The **Authentication** tab allows you to connect Enterprise Steam to your current user database.
- The **Licensing** tab provides information on your Enterprise Steam License.
• The **YARN** tab allows you to enter YARN configuration settings that will apply when users launch new clusters.

• The **Engines** tab allows you to add a manager file and H2O engines.

• The **Spark** tab allows you to enable Spark and provide paths to your Spark home directory, Hadoop configuration directory, and Java 8 home directory.

• The **Token** tab allows you to create your own personal access tokens for use in scripts and on the command line.

Each of these tabs is described in greater detail in the sections that follow.

### 3.1 Authentication

Enterprise Steam supports Local, LDAP, and SAML authentication. No additional configuration is required for Local authentication. Refer to the sections that follow for information on how to configure LDAP and SAML authentication.

#### 3.1.1 Configure LDAP Connection Settings

Enterprise Steam ships with a built-in SQLite database. By default, Enterprise Steam uses this database to store user and cluster management metadata. You can use this database, or you can configure Enterprise Steam to work with your existing LDAP directory.

1. Navigate to the **Configurations** page and select the **Authentication** tab.

2. Select LDAP in the **User DB Type** drop down menu, then configure the LDAP connection settings. (Refer to the table below and the image that follows.)
<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LDAP Connection Settings</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Host</td>
<td>The LDAP host server address</td>
<td>ldap.0xdata.loc</td>
</tr>
<tr>
<td>Port</td>
<td>The LDAP server port</td>
<td>389</td>
</tr>
<tr>
<td>SSL-Enabled</td>
<td>Enable this if your LDAP supports SSL.</td>
<td></td>
</tr>
<tr>
<td>Bind DN</td>
<td>The Distinguished Name used by the LDAP server if extended access is required. This can be left blank if anonymous bind is sufficient.</td>
<td>cn=admin,dc=0xdata,dc=loc</td>
</tr>
<tr>
<td>Bind DN Password/Confirm</td>
<td>The password for the Bind DN user</td>
<td>h2o</td>
</tr>
<tr>
<td>User Base DN</td>
<td>The location of the LDAP users, specified by the DN of your user subtree</td>
<td>ou=users,dc=0xdata,dc=loc</td>
</tr>
<tr>
<td>User Base Filter</td>
<td>The LDAP search filter used to filter users</td>
<td>department=IT</td>
</tr>
<tr>
<td>User Name Attribute</td>
<td>The User Attribute that contains the username</td>
<td>uid</td>
</tr>
<tr>
<td><strong>Group Settings</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group Names</td>
<td>The Distinguished Name used for group synch</td>
<td>cn=jettygroup,ou=groups,dc=0xdata,dc=loc</td>
</tr>
<tr>
<td>Group Base DN</td>
<td>The location of your LDAP groups, specified by the DN of your user subtree</td>
<td>ou=groups,dc=0xdata,dc=loc</td>
</tr>
<tr>
<td>Group Name Attribute</td>
<td>The Group Attribute that contains the username</td>
<td>cn</td>
</tr>
<tr>
<td>Static Member Attribute</td>
<td>The attribute for static group entries</td>
<td>memberUid</td>
</tr>
<tr>
<td><strong>Advanced Settings</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search Request Size Limit</td>
<td>Limit the size of search results. 0 indicates unlimited.</td>
<td></td>
</tr>
<tr>
<td>Search Request Time Limit</td>
<td>Limit the time allotted for completing search results. 0 indicates unlimited.</td>
<td>0</td>
</tr>
<tr>
<td>Cache Max Age (in mins)</td>
<td>The maximum age in minutes of of LDAP record in cache before forcing a refresh. Use 0 for no cache (not recommended).</td>
<td>5</td>
</tr>
<tr>
<td>Certificate Path</td>
<td>Specify CAs to use for contacting LDAP servers. Leave empty to use system root CAs.</td>
<td></td>
</tr>
</tbody>
</table>
3. Click **Test Config** when you are done. A valid response message indicates that the configuration was successful.

4. Click **Save Config**.

After LDAP is configured, users can log in to Enterprise Steam using their LDAP username and password.

**Notes:**

- The **Reset** button clears all user-specified information in this form and resets any default values.
- The **Invalidate LDAP cache** button invalidates the records in the LDAP cache and forces the cache to retrieve updated records for users.

### 3.1.2 Configure SAML Connection Settings

Perform the following steps to configure Enterprise Steam to use SAML authentication.

1. Navigate to the **Configurations** page and select the **Authentication** tab.

2. Select SAML in the **User DB Type** drop down menu, then configure the following SAML settings:
<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SAML Settings</strong></td>
<td></td>
</tr>
<tr>
<td>IDP Meta-data Path</td>
<td>The path to the SAML Identity Provider (IdP) metadata file on the local file system.</td>
</tr>
<tr>
<td>Keystore Path</td>
<td>The path to the keystore file on the local file system.</td>
</tr>
<tr>
<td>Keystore Password</td>
<td>The keystore password.</td>
</tr>
<tr>
<td>Base URL</td>
<td>The base URL for Enterprise Steam. For example, <a href="http://steam.loc:8888">http://steam.loc:8888</a>.</td>
</tr>
<tr>
<td><strong>Group Settings</strong></td>
<td></td>
</tr>
<tr>
<td>User Name Attribute</td>
<td>The attribute of authorization token that contains usernames.</td>
</tr>
<tr>
<td>Group Name Attribute</td>
<td>The attribute of authorization token that contains group names.</td>
</tr>
<tr>
<td>Admin Group Name</td>
<td>The name of the admin group that will get privileges in Enterprise Steam.</td>
</tr>
<tr>
<td><strong>Advanced</strong></td>
<td></td>
</tr>
<tr>
<td>SAML Entity ID</td>
<td>The PartnerSpID value that will be passed to the IDP. This is optional.</td>
</tr>
<tr>
<td>Logout URL</td>
<td>Specify the URL where the user will be redirected to after logging out. This is optional. By default, users will see the “Logged Out” screen.</td>
</tr>
</tbody>
</table>

3. Click **Save and Enable** when you are done.

**Notes:**

- The **Remove LDAP** button is available on this form if LDAP authentication is already configured. Click this button to remove the LDAP configuration.
• The **Disable SAML** button disables an SAML authentication that is enabled, but it does not delete the configured settings. Note that this will require a server restart.

### 3.2 Profiles

The Profiles tab allow you to define individual cluster sizes and configurations. Admins can then give different users access to the different clusters by specifying a specific profile when launching a new cluster.

Enterprise Steam comes with three profiles:

- **default-h2o**: This is enabled by default.
- **default-sparkling-internal**: This is disabled by default. Configure Spark settings to enable this profile. (See the *Spark* section for more information.)
- **default-sparkling-external**: This is disabled by default. Configure Spark settings to enable this profile. (See the *Spark* section for more information.)

**Note**: The minimum 2.1.41/2.2.27/2.3.16/2.4.*

From this page, you can edit any of the default profiles, add additional profiles, copy profiles, and delete profiles.

#### 3.2.1 Adding Profiles

1. On the Configurations page, click the **Profiles** tab. This page shows a list of available profiles.

2. In the **Create New Profile** section of this page, enter a name for the new profile and select an available type (H2O, Sparkling Water - Internal Backend, Sparkling Water - External Backend). Click **Create** when you are ready. This opens the **Creating Profiles** form. Note that this form varies depending on the Type.
**H2O Type**

1. **YARN Queues**: Optionally specify a comma-separated list of YARN queues available for user of this profile. Leave empty if you want to let the user to specify this parameter when launching the cluster.
2. **LDAP Groups**: Optionally specify a comma-separated list of LDAP groups that will have access to this cluster. Enter * to allow any LDAP user to access this profile. Leave empty if you want to manually assign each cluster profile to each user.
3. **SAML Groups**: Optionally specify a comma-separated list of SAML groups that will have access to this cluster. Enter * to allow any SAML user to access this profile. Leave empty if you want to manually assign each cluster profile to each user.
4. **Cluster Limit Per User**: Specify the maximum number of clusters that a user with this profile can launch.
5. **H2O Nodes**: Specify the minimum, maximum, and default number of allowed H2O nodes (cluster size) for this profile.
6. **H2O Node Memory (GB)**: Specify the minimum, maximum, and default amount of memory to allocate to H2O for each node (in GB).
7. **H2O Node Threads**: Specify the minimum, maximum, and default number of H2O threads (CPUs) to use for each node. 0 defaults to using all CPUs on the host.
8. **H2O Node Extra Memory (%)**: Specify the minimum, maximum, and default extra memory for internal JVM use outside of the Java heap. (This corresponds to the extramempercent Hadoop launch parameter.)
9. **Maximum Idle Time (hrs)**: Specify the minimum, maximum, and default idle time in hours.
10. **Maximum Uptime (hrs)**: Specify the minimum, maximum, and default uptime in hours.
11. **YARN Virtual Cores**: Specify the minimum, maximum, and default number of YARN virtual cores.

**Sparkling Water - Internal Backend Type**

1. **YARN Queues**: Optionally specify a comma-separated list of YARN queues available for user of this profile. Leave empty if you want to let the user to specify this parameter when launching the cluster.
2. **LDAP Groups**: Optionally specify a comma-separated list of LDAP groups that will have access to this cluster. Enter * to allow any LDAP user to access this profile. Leave empty if you want to manually assign each cluster profile to each user.
3. **SAML Groups**: Optionally specify a comma-separated list of SAML groups that will have access to this cluster. Enter * to allow any SAML user to access this profile. Leave empty if you want to manually assign each cluster profile to each user.
4. **Cluster Limit Per User**: Specify the maximum number of clusters that a user with this profile can launch.
5. **Spark Properties**: Optionally enter additional Spark properties for this cluster. Specify one property per line using 'key=value' format.
6. **Python PySpark Paths**: Enter a path to the Python distribution binary. Use one line per entry with the following format: name=/path/to/python.
7. **Driver Cores**: Specify the minimum, maximum, and default number of driver cores.
8. **Driver Memory (GB)**: Specify the minimum, maximum, and default driver memory (in GB).
9. **Number of Executors**: Specify the minimum, maximum, and default number of executors.
10. **Executor Cores**: Specify the minimum, maximum, and default number of cores per executor.
11. **Executor Memory**: Specify the minimum, maximum, and default amount of executor memory per node (in GB).

12. **H2O Node Threads**: Specify the minimum, maximum, and default number of H2O threads (CPUs) to use for each node. 0 defaults to using all CPUs on the host.

13. **Startup Timeout (seconds)**: Specify the minimum, maximum, and default startup timeout in seconds. The cluster will terminate if it cannot start within this time.

### Sparkling Water - External Backend Type

1. **YARN Queues**: Optionally specify a comma-separated list of YARN queues available for user of this profile. Leave empty if you want to let the user to specify this parameter when launching the cluster.

2. **LDAP Groups**: Optionally specify a comma-separated list of LDAP groups that will have access to this cluster. Enter * to allow any LDAP user to access this profile. Leave empty if you want to manually assign each cluster profile to each user.

3. **SAML Groups**: Optionally specify a comma-separated list of SAML groups that will have access to this cluster. Enter * to allow any SAML user to access this profile. Leave empty if you want to manually assign each cluster profile to each user.

4. **Cluster Limit Per User**: Specify the maximum number of clusters that a user with this profile can launch.

5. **Spark Properties**: Optionally enter additional Spark properties for this cluster. Specify one property per line using ‘key=value’ format.

6. **Python PySpark Paths**: Enter a path to the Python distribution binary. Use one line per entry with the following format: name=/path/to/python.

7. **Driver Cores**: Specify the minimum, maximum, and default number of driver cores.

8. **Driver Memory (GB)**: Specify the minimum, maximum, and default driver memory (in GB).

9. **Number of Executors**: Specify the minimum, maximum, and default number of executors.

10. **Executor Cores**: Specify the minimum, maximum, and default number of cores per executor.

11. **Executor Memory**: Specify the minimum, maximum, and default amount of executor memory per node (in GB).

12. **H2O Nodes**: Specify the minimum, maximum, and default number of allowed H2O nodes (cluster size) for this profile.

13. **H2O Node Memory (GB)**: Specify the minimum, maximum, and default amount of memory to allocate to H2O for each node (in GB).

14. **H2O Node Threads**: Specify the minimum, maximum, and default number of H2O threads (CPUs) to use for each node. 0 defaults to using all CPUs on the host.

15. **Startup Timeout (seconds)**: Specify the minimum, maximum, and default startup timeout in seconds. The cluster will terminate if it cannot start within this time.

3. Click **Save** when you are done.

Upon completion, the new profile will appear in the Existing Profile section. If necessary, you can update or delete existing profiles directly from this section.
3.2.2 Editing Profiles

Enterprise Steam comes with a default profile. You can edit this profile or other existing profiles by following the steps below.

1. On the Configurations page, click the Profiles tab. This page shows a list of available profiles.
2. Click the Edit button beside the profile that you want to edit.
3. Edit any properties that you want to change, then click Save at the bottom of the form.

3.2.3 Copying Profiles

Copying profiles is an easy way to create a new profile based on an existing one.

1. On the Configurations page, click the Profiles tab. This page shows a list of available profiles.
2. Click the Copy button beside the profile that you want to copy.
3. Change the name of the profile and change any options that you want to be different from the existing profile.
4. Click Save when you are done.

Upon completion, the new profile will appear in the Existing Profile section.

3.2.4 Deleting Profiles

1. On the Configurations page, click the Profiles tab. This page shows a list of available profiles.
2. Click the Delete button beside the profile that you want to delete.
3. A confirmation page displays. Click Confirm to complete the deletion.

3.3 Users

The Users tab shows all current Enterprise Steam users. This section describes how to add, edit, and deactivate users.

3.3.1 Adding Users

Admins can add users into the Enterprise Steam SQLite database from within the UI.

1. At the top of the Configurations page, click the Create User button.

2. Username: Enter the name of the user. Note that the name must match with a username in your YARN system.
3. **Password/Confirm Password**: Specify and confirm a password for the user.

4. **Role**: Specify the role(s) for this user. Note that Enterprise Steam ships with two default roles: admin and standard user.

5. **Cluster Profile**: Specify the cluster profile that this user will be part of. Note that Enterprise Steam ships with a default cluster profile.

6. Click **Create User** when you are done.

Upon successful completion, the new user will appear in the list of Enterprise Steam users.

### 3.3.2 Editing Users

This section describes how to edit a user’s role.

On the Users tab, click the **Edit** link beside the user you want to edit. This opens the Edit User Details form. Change the user’s roles or cluster profile. You can also specify an authentication type of LDAP, Local, or SAML. Click **Confirm** when you are done.
3.3.3 Resetting a User’s Password

If a user is added with Local Authentication, then admins can reset the user’s password by clicking the **Reset Local Password** link for the desired user. A new password will display at the top of the screen for approximately 5 seconds. This new password should then be provided to the user so that he/she can log in to Enterprise Steam. Note that this option is not available for users added with LDAP or SAML authentication.
3.3.4 Deactivating/Reactivating Users

On the Users tab, click the **Deactivate Steam User** link for the user whose Enterprise Steam access you want to revoke. Click **Reactivate Steam User** to once again grant access for that user.
3.4 Roles

Roles determine the activities/permissions that an Enterprise Steam user can perform within your environment. Enterprise Steam ships with two default roles: admin and standard user. These default roles are sufficient for most Enterprise Steam deployments and, in general, should not be changed. You can create additional roles, however, if you require more granularity in the way that your users access and utilize Enterprise Steam.

3.4.1 Creating Roles

1. At the top of the Configurations page, click the **Create Role** button.

2. Specify a name and description for the role.

3. Select the permissions that will be granted to this role.

4. Click **Create Role** at the bottom of the form when you are done.
3.4.2 Changing Permissions

Admins can add or remove permissions for each role directly on the Roles tab.

1. Select the checkbox for the corresponding permission and role that you want to change.
2. Click Review Changes at the bottom of the page. A popup displays, providing you with a summary of the changes.
3. Click the Confirm button beside each change that you want to make, then click Save Changes to complete the update.
3.4.3 Deleting Roles

On the Roles tab, scroll down to the bottom of the page, and click the trashcan icon under the Role column that you want to delete. A confirmation page will display, prompting you to confirm the deletion. Click Confirm to remove the role.
3.5 YARN

A YARN tab is available for Enterprise Steam customers that want to control the way that users launch new clusters, whether it is specifying a YARN queue or limiting the number of nodes that a user can create when launching a new cluster.

3.5.1 Adding a New YARN Config Entry

Perform the following steps to add a new config entry.

**Note:** Use caution when adding a new config entry. These values are currently not validated, and incorrect entries will result in a failure on cluster launch.

1. **Name**: Specify a valid config name. This name must be a valid Hadoop launch parameter (such as network, driverif, driverport, nodes, etc.)
2. **Value**: Specify a value for the config. This value must correspond to the name. For example, you cannot enter an IP address if the new entry is for nodes.
3. **Priority**: Specify whether this configuration will act as a default or whether this will override any values that the user sets.
4. **Type**: Specify whether this entry is a Hadoop type (passed directly to the YARN command line) or an H2O Driver type (specifying how Enterprise Steam will run).
5. Click **Add Entry** when you are done. The new entry will display on this page. Repeat these steps to add additional YARN config entries.
3.5.2 Updating a YARN Config Entry

On the YARN tab, enter new values in the entry that you want to update, then click the **Update Entry** button.

![YARN Config Entry Form]

3.5.3 Deleting a YARN Config Entry

On the YARN tab, click the **Remove Entry** button below the entry that you want to remove.

![YARN Config Entry Form]
3.6 Engines

The Engines tab allows you to upload h2o-3 drivers and Sparkling Water drivers and optionally tie the Enterprise Steam environment to a Steam Hadoop helper file. Once added, these drivers (engines) will be available for Enterprise Steam users.

1. On your local machine, download the h2odriver from the H2O Download page. Be sure to select the version that corresponds with your version of Hadoop. For example:

```
→zip
```

2. You can also optionally download a new Steam Helper file from the H2O Download page Steam Manager (Hadoop Helper) tab. For example:

```
wget https://s3.amazonaws.com/steam-release/enterprise-steam/steam-hadoop-
→helper/1.1.0/steam-hadoop-helper-hadoop2.4-1.1.0.jar
```

3. Navigate to the Configurations > Engines tab.

4. If you downloaded a new Steam Helper in step 2, then you can upload the new manager file here. The manager file is the Steam Hadoop helper, which is used to prevent clusters from using impersonation. This can be deleted after being added.

5. Browse to and select the H2O driver that you want to add.

6. Optionally browse and select the Sparkling Water driver that you want to add. Note that the Sparkling Water driver can only be used if Spark is enabled. (See the Spark section for more information.)
Upon completion, the new driver(s) will appear on this page and will be available when launching a new cluster.
You can delete drivers from this page by selecting the trashcan icon beside the driver that you want to remove.

### 3.7 Spark

The Spark tab allows you to enable Spark and provide paths to your Spark home directory and Hadoop configuration directory. When Spark is enabled, then the `default-sparkling-internal` and `default-sparkling-external` profiles will be available. (See the *Profiles* section for more information.)

1. Navigate to the **Configurations > Spark** tab.

2. Click **Enabled** to enable Spark.
3. **Spark Home**: Specify the path to your Spark home directory (SPARK_HOME).
4. **Hadoop Conf Dir**: Specify the path to your Hadoop configuration directory (HADOOP_CONF_DIR).
5. **Java8 Home**: Specify the path to your Java 8 home directory (JAVA_HOME).
6. **Default Sparkling Water Backend**: Select a Sparkling Water backend to enforce (internal or external); otherwise leave it as a choice for the user.
7. Click **Save Config** when you are done.

### 3.8 Token

The Token tab allows you to generate a personal access tokens for use in scripts and on the command line. **Note**: Be careful, these tokens are like passwords so you should guard them carefully. The advantage to using a token over putting your password into a script is that a token can be revoked.

On the **Configuration > Token** tab, click **Generate New Token** to generate and retrieve your token. **Note**: For security reasons the token will be shown only once after generating. If you lose your token, you must generate a new one. You can only have one token at a time.
CONFIGURATIONS (ENTERPRISE STEAM GLOBAL)

All settings on this page affect the Enterprise Steam installation globally for administrators. All users on Enterprise Steam can potentially be affected. Please be careful and consult support if you are unsure about the implications of updating these settings.

Personal Access Token

You can now create your own personal access tokens for use in scripts and on the command line. Be careful, these tokens are like passwords so you should guard them carefully. The advantage to using a token over putting your password into a script is that a token can be revoked. For security reasons the token will be shown only once after generating. If you lose your token you must generate a new one. You can only have one token at the same time.

[Generate New Token]
The **Clusters** page shows all H2O clusters that Enterprise Steam is connected to along with the state of the cluster, the cluster type, the ID of the user who started the cluster, and the cluster creation date.

### 4.1 Launching a New Cluster

The options for launching a new cluster vary based on the selected cluster type.

#### 4.1.1 Launch a New H2O Cluster

1. In the Enterprise Steam UI, navigate to the **Clusters** page and select **Launch New Cluster**.
2. Select **H2O** from the Cluster Type dropdown menu.
3. Select a Cluster Profile from the dropdown menu to use when setting up the new cluster. Cluster profiles are configured by the Admin on the Configurations page and provide the allowed min and max values for each option in a cluster profile.

4. Specify values for the options below. Once added, other Enterprise Steam users will be able to connect to this cluster.
   - **Cluster Name**: Specify a name for this cluster.
   - **H2O Version**: For new H2O clusters, specify the H2O version to use.
   - **Number of Nodes**: Specify the number of nodes for the cluster.
   - **Java Memory per Node [GB]**: Specify the amount of memory that should be available on each node.
   - **YARN Virtual Cores per Node**: Specify the number of virtual cores.
   - **H2O Threads per Node**: Specify the number of threads (CPUs) to use in the cluster. Leave this blank to use all available threads.
   - **Extra Memory**: Specify the amount of extra memory for internal JVM use outside of the Java heap. This is a percentage of memory per node. The default (and recommended) value is 10%.
   - **Maximum Idle Time [HRS]**: Specify the maximum number of hours that the cluster can be idle before gracefully shutting down. Leave this blank to turn off this setting and allow the cluster to remain idle for an unlimited amount of time.
   - **Maximum Uptime [HRS]**: Specify the maximum number of hours that the cluster can be running. Leave this blank to turn off this setting and allow the cluster to remain up for an unlimited amount of time.
   - **Leader Node ID**: Optionally specify whether to connect to a different leader node.
   - **YARN Queue**: If your cluster contains queues for allocating cluster resources, optionally specify a queue for this cluster. Note that the YARN Queue cannot contain spaces. Leave this empty to use the default YARN queue.
5. Click the **Launch New Cluster** button to start the new cluster.

Upon successful completion, the cluster will appear on the **Clusters** page.

### 4.1.2 Launch a New Sparkling Water - Internal Backend Cluster

1. In the Enterprise Steam UI, navigate to the **Clusters** page and select **Launch New Cluster**.

2. Select **Sparkling Water - Internal Backend** from the Cluster Type dropdown menu.

3. Select a Cluster Profile from the dropdown menu to use when setting up the new cluster. Cluster profiles are configured by the Admin on the Configurations page and provide the allowed min and max values for each option in a cluster profile.

4. Specify values for the options below. Once added, other Enterprise Steam users will be able to connect to this cluster.
   - **Cluster Name**: Specify a name for this cluster.
   - **Sparkling Water Version**: For new Sparkling Water clusters, specify the Sparkling Water version to use.
• **Driver Cores**: Specify the number of driver cores for the cluster.

• **Driver Memory [GB]**: Specify the amount of driver memory that should be available on each core.

• **Number of Executors**: Specify the number of cores per executor.

• **Executor Memory**: Specify the amount of executor memory per node (in GB).

• **H2O Threads per Node**: Specify the number of threads (CPUs) to use in the cluster. 0 indicates to use all available threads.

• **Startup Timeout [SEC]**: Specify the startup timeout in seconds. The cluster will terminate if it cannot start within this time.

• **YARN Queue**: If your cluster contains queues for allocating cluster resources, optionally specify a queue for this cluster. Note that the YARN Queue cannot contain spaces. Leave this empty to use the default YARN queue.

• **PySpark Python Path**: If you have a custom Python environment, you can specify the path here; otherwise, leave empty to use the default Python path.

• **Spark Properties**: This shows a list of additional Spark properties for the cluster. This is maintained on by Admins from the Profiles tab.

5. Click the Launch New Cluster button to start the new cluster.

Upon successful completion, the cluster will appear on the Clusters page.

### 4.1.3 Launch a New Sparkling Water - External Backend Cluster

1. In the Enterprise Steam UI, navigate to the Clusters page and select Launch New Cluster.

2. Select Sparkling Water - External Backend from the Cluster Type dropdown menu.

3. Select a Cluster Profile from the dropdown menu to use when setting up the new cluster. Cluster profiles are configured by the Admin on the Configurations page and provide the allowed min and max values for each option in a cluster profile.

4. Specify values for the options below. Once added, other Enterprise Steam users will be able to connect to this cluster.

• **Cluster Name**: Specify a name for this cluster.

• **Sparkling Water Vesion**: For new Sparkling Water clusters, specify the Sparkling Water version to use.

• **Driver Cores**: Specify the number of driver cores for the cluster.

• **Driver Memory [GB]**: Specify the amount of driver memory that should be available on each core.

• **Number of Executors**: Specify the number of cores per executor.

• **Executor Cores**: Specify the number of cores per executor.

• **Executor Memory**: Specify the amount of executor memory per node (in GB).

• **H2O Nodes**: Specify the number of H2O nodes.

• **H2O Memory Per Node [GB]**: Specify the amount of memory in GB to allocate to H2O for each node.

• **H2O Threads per Node**: Specify the number of threads (CPUs) to use in the cluster. 0 indicates to use all available threads.

• **Startup Timeout [SEC]**: Specify the startup timeout in seconds. The cluster will terminate if it cannot start within this time.
• **YARN Queue**: If your cluster contains queues for allocating cluster resources, optionally specify a queue for this cluster. Note that the YARN Queue cannot contain spaces. Leave this empty to use the default YARN queue.

• **PySpark Python Path**: If you have a custom Python environment, you can specify the path here; otherwise, leave empty to use the default Python path.

• **Spark Properties**: This shows a list of additional Spark properties for the cluster. This is maintained on by Admins from the Profiles tab.

5. Click the **Launch New Cluster** button to start the new cluster.

Upon successful completion, the cluster will appear on the **Clusters** page.

### 4.2 Deleting Clusters

Click the **Delete** icon beside the cluster that you want to delete. A confirmation message will display. Click **Confirm** to continue. This action stops and then removes the cluster.

**Note**: When deleting a “Stopped” cluster, you will be given the opportunity to also delete all “Stopped” clusters. This option is not available when deleting a “Started” cluster.
4.3 Notebooks

Enterprise Steam allows you to upload and run PySparkling and RSparking Jupyter Notebooks from within the cluster. More information about Jupyter Notebooks is available here: https://jupyter.org/.

4.3.1 Requirements for RSparkling Jupyter Notebooks

- The Hadoop cluster must have R installed along with devtools and sparklyr libraries.
- In Steam you must have an H2O engine uploaded as well as Sparkling Water engine that was built for the same H2O version. The supported minimum Sparkling Water versions for this feature include:
  - 2.1.53
  - 2.2.39
  - 2.3.28
  - 2.4.10

4.3.2 Launching Notebooks

In the Enterprise Steam UI, navigate to the Clusters page and click the My Notebooks button in the upper-right corner to view available notebooks.

4.3.3 Creating New Notebooks

On the Jupyter Notebook Files tab, click New dropdown and select the type of notebook or other file that you want to create. Then create and save the new notebook.
4.3.4 Adding Notebooks

1. On the Jupyter Notebook Files tab, click the **Upload** button.
2. Browse to the location on your local machine where your notebooks is stored.
3. Click **Upload** to complete the notebook upload process.