Enterprise Steam Installation and Setup

Release 1.6.1.2

H2O.ai

Jan 14, 2020
## CONTENTS

1 Installation 3
   1.1 Before You Begin the Installation ........................................... 3
   1.2 Ubuntu Installation ............................................................... 6
   1.3 RHEL 6 / CentOS 6 Installation ................................................ 6
   1.4 RHEL 7 / CentOS 7 Installation ................................................ 7

2 Logging-in 9

3 Configurations 11
   3.1 Server configuration ............................................................. 12
   3.2 General Settings ................................................................. 13
   3.3 Backends .............................................................................. 29
   3.4 Products .............................................................................. 32

4 Upgrade 41
   4.1 Backup ............................................................................... 41
   4.2 Upgrade Enterprise Steam service ............................................. 41
   4.3 Upgrade Python/R clients ......................................................... 41

5 Restart 43

6 Setup Hadoop Impersonation 45
   6.1 In Cloudera Manager .............................................................. 45
Enterprise Steam is a service for securely using H2O.ai products such as H2O, Sparkling Water and Driverless AI in an enterprise 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 adoption of H2O.ai products 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, Sparkling Water or Driverless AI instances where they can practice data science.

For administrators, Enterprise Steam provides control over the use, configuration and resource usage of H2O products.

This document applies to administrators and describes how to install and start Enterprise Steam to make it accessible to a set of users.
Enterprise Steam supports RHEL/CentOS 6 or newer and Ubuntu 12 or newer. First, review the *Before You Begin the Installation* section, then follow the instructions for your platform.

### 1.1 Before You Begin the Installation

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

- Complete the preinstall worksheet to identify what is necessary for the installation.
- Obtain the Enterprise Steam license key.
- Sudo or root access is required to install Enterprise Steam.
- For H2O and Sparkling Water, a Hadoop admin is needed in order to setup impersonation.
- The installation creates a new system user called “steam” if it does not exist. It is highly recommended to use this service ID to run Enterprise Steam.
- Enterprise Steam requires a single port (:9555) to be open before installation. 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.

### 1.1.1 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.

**General Information**

1. What are the IP address and hostname of the server where Enterprise Steam will run?
   
   _____________________________________________________

2. Who is your IT firewall administrator?
   
   _____________________________________________________

3. What TCP port will Enterprise Steam use (recommended 9555)?
   
   _____________________________________________________
4. Does the Enterprise Steam port need to be opened through a firewall?

5. 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]?

6. Do you have root login access to the server (required for installation of the package using yum or dpkg)?

7. What is the service ID that will be used to run Enterprise Steam? (recommended to use the provided “steam” service ID)

8. Will you provide a TLS certificate and private key for Enterprise Steam to use?

9. Will you use LDAP/AD or SAML to authenticate users?

For Hadoop Users

1. Who is your Hadoop administrator?

2. What is your Hadoop distribution (for example, HDP2.4 or CDH5.5 or MAPR 5.1)?

3. Does your Hadoop environment use Kerberos authentication?

For Spark Users

1. What is your Spark version?

2. What is your SPARK_HOME directory?

3. What is your JAVA_HOME directory?

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)?

___________________________________________________________

For Kerberos Users

1. Who is your Kerberos administrator?

___________________________________________________________

2. Did you create a keytab file for the Enterprise Steam Principal? Where is it stored?

___________________________________________________________

3. What is the name of the Enterprise Steam Principal?

___________________________________________________________

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

___________________________________________________________

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

___________________________________________________________
1.1.2 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. See the license section for more information on how to install the license.

1.2 Ubuntu Installation

This section describes how to install Enterprise Steam on Ubuntu. Same steps should apply for similar Debian based Linux distributions.

1.2.1 Requirements

- Ubuntu 12.04 or greater.
- Enterprise Steam .deb file. This is available on the Enterprise Steam download page.

1.2.2 Install Enterprise Steam

1. Download Enterprise Steam from the download page.
2. Install the Enterprise Steam DEB package.
   
   ```
   sudo dpkg -i steam-X.Y.Z.x86_64.deb
   ```
3. Set the Enterprise Steam administrator username and password.
   
   ```
   sudo service steam set-admin
   ```
4. Start Enterprise Steam.
   
   ```
   sudo service steam start
   ```
5. Validate that the Enterprise Steam service is running.
   
   ```
   sudo service steam status
   ```

If the service is running, you can login to Enterprise Steam. See the Logging-in section for more information.

If the service is not running, please review the logs(`cat /var/log/steam/steam.log`) to identify the issue.

1.3 RHEL 6 / CentOS 6 Installation

This section describes how to install Enterprise Steam on Red Hat Enterprise Linux 6 and CentOS 6.

1.3.1 Requirements

- RHEL 6 or CentOS 6.
- Enterprise Steam .rpm file. This is available on the Enterprise Steam download page.
1.3.2 Install Enterprise Steam

1. Download Enterprise Steam from the download page.
2. Install the Enterprise Steam RPM package.
   ```bash
   yum localinstall steam-X.Y.Z.x86_64.el6.rpm
   ```
3. Set the Enterprise Steam administrator username and password.
   ```bash
   sudo /etc/init.d/steam set-admin
   ```
4. Start Enterprise Steam.
   ```bash
   sudo /etc/init.d/steam start
   ```
5. Validate that the Enterprise Steam service is running.
   ```bash
   sudo /etc/init.d/steam status
   ```

If the service is running, you can login to Enterprise Steam. See the *Logging-in* section for more information.

If the service is not running, please review the logs (`cat /var/log/steam/steam.log`) to identify the issue.

1.4 RHEL 7 / CentOS 7 Installation

This section describes how to install Enterprise Steam on Red Hat Enterprise Linux 7 and CentOS 7.

1.4.1 Requirements

- RHEL 7 or CentOS 7.
- Enterprise Steam .rpm file. This is available on the Enterprise Steam download page.

1.4.2 Install Enterprise Steam

1. Download Enterprise Steam from the download page.
2. Install the Enterprise Steam RPM package.
   ```bash
   yum localinstall steam-X.Y.Z.x86_64.el7.rpm
   ```
3. Set the Enterprise Steam administrator username and password.
   ```bash
   sudo su -s /bin/bash -c "/opt/h2oai/steam/steam set admin" steam
   ```
4. Start Enterprise Steam.
   ```bash
   sudo systemctl enable steam
   sudo systemctl start steam
   ```
5. Validate that the Enterprise Steam service is running.
   ```bash
   sudo systemctl status steam
   ```
If the service is running, you can login to Enterprise Steam. See the Logging-in section for more information.

If the service is not running, please review the logs (`cat /var/log/steam/steam.log`) to identify the issue.
After completing the Enterprise Steam installation and starting the Enterprise Steam service, the next step is to log-in as an administrator to finish the setup.

**Note:** Only Chrome version 50 or newer is supported.

1. Open a Chrome browser and navigate to `https://<hostname>:9555`. If the page does not load, ensure that the Enterprise Steam service is running, the Enterprise Steam port is opened and that the link is correctly using HTTPS. Ignore any warnings about self-signed certificates for now. You will have an option to add TLS certificate once logged in.

2. Log in using the administrator username and password that you created during the installation process.

3. Enter the *Configurations* section to finish the setup.
The Configurations page allows Enterprise Steam Admins to perform a complete setup of Enterprise Steam. To change the server startup settings that cannot be changed in the UI, see the Server configuration section.

Note: Only Admins have access to the Configurations page.

Some changes require Enterprise Steam to be restarted. In that case, the configurations page will show a banner. See the Restart section on how to restart Enterprise Steam.

The Configurations page consists of the following sections.

- **GENERAL**
  - **ACCESS CONTROL**
  - Set the system-wide Authentication method
  - Create a personal token for use in scripts and on the command line
  - Add/edit users
  - Add/edit roles
• Add/edit profiles
• STEAM CONFIGURATION
  • Upload your license and review your license information
  • Configure security
  • Configure logging
  • Import or export current configuration for authentication, YARN, Sparkling Water, Driverless AI, security, and logging
• BACKENDS
  • HADOOP
    • Configure the Hadoop backend to enable H2O and Sparkling Water clusters
  • H2O AGENT
    • Configure H2O Agent(s) to enable on-premise deployment backend for Driverless AI
• PRODUCTS
  – Configure Enterprise Steam to work with H2O, upload a new H2O-3 engine, and specify H2O launch parameters
  – Configure Enterprise Steam to work with Sparkling Water, upload a new Sparkling Water engine, and specify the desired Python environment(s)
  – Configure Enterprise Steam to work with Driverless AI and add new Driverless AI servers

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

3.1 Server configuration

Enterprise Steam has a small configuration file for changes that cannot be made in the Web UI. You can find a template of the configuration file at `/etc/steam/steam.yaml.template`.

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

# This port is the one to be exposed externally.
# Steam's bundled haproxy binds to this port.
# Steam starts and manages it's own haproxy process with it's own
# config file (which is autogenerated).
STEAM_HTTPS_PORT: 9555

# The internal port for the Steam process.
# This is the internal port of the Steam itself. It should never be
# accessed directly. See STEAM_HTTPS_PORT for the public port to use
# when accessing Steam. It is rare that this port should be set except
# in cases of port conflict.
STEAM_INTERNAL_PORT: 29000

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

To make any changes, save the modified configuration file as `/etc/steam/steam.yaml` and Restart Enterprise Steam.
3.2 General Settings

General Settings allow you to control access to Enterprise Steam and to configure Enterprise Steam.

3.2.1 Access Control

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.

CONFIGURATIONS

System-wide configuration for Enterprise Steam. Please refer to the documentation for all setting validation before saving. Some settings require Enterprise Steam to be restarted to take effect.

<table>
<thead>
<tr>
<th>ACCESS CONTROL</th>
<th>STEAM CONFIGURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentication</td>
<td>Licensing</td>
</tr>
<tr>
<td>Token</td>
<td>Security</td>
</tr>
<tr>
<td>Users</td>
<td>Logging</td>
</tr>
<tr>
<td>Roles</td>
<td>Import/Export</td>
</tr>
<tr>
<td>Profiles</td>
<td></td>
</tr>
</tbody>
</table>

Authentication

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 Authentication page.

2. Select LDAP in the Enabled authentication 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>Hostname</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.</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>LDAP 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>LDAP 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>CA 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 Invalidate LDAP cache button invalidates the records in the LDAP cache and forces the cache to retrieve updated records for users.

Configure SAML Connection Settings

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

1. Navigate the Authentication page.

2. Select SAML in the Enabled authentication 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 Metadata 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>SAML 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>SAML Advanced Settings</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.

**Token**

The Token page 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 **Token** page, 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.
**PERSONAL ACCESS TOKEN**

You can 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.

**Users**

By default, the Users page shows all current Enterprise Steam users. (Note that you can also specify to show deactivated Enterprise Steam users as well.) This section describes how to add, edit, and deactivate users.

**Adding Users**

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

1. Click the **Create User** button on the Users Page.
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. **YARN Queues**: Optionally specify a list of YARN queues associated with this user.
6. **Cluster Profile**: Specify the cluster profile(s) that this user will be part of. Note that Enterprise Steam ships with a number of default cluster profiles.
7. Click **Create User** when you are done.
Upon successful completion, the new user will appear in the list of Enterprise Steam users.

**Editing Users**

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

On the Users page, 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, and you can specify YARN queues for the user. Click **Confirm** when you are done.
Note: A message will display in the UI if you remove all roles from a user.

**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.
Deactivating Users

On the Users page, click the Deactivate Steam User link for the user whose Enterprise Steam access you want to revoke.

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.

Creating Roles

1. Click the Create Role button on the Roles page.
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.

Changing Permissions

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

![Role permissions](image)

### Deleting Roles

On the Roles page, 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.

**Note** The Admin role cannot be deleted.

![Support Logon](image)

### Profiles

The Profiles page 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 four 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 Sparkling Water section for more information.)
- default-sparkling-external: This is disabled by default. Configure Spark settings to enable this profile. (See the Sparkling Water section for more information.)
- default-dai: This is disabled by default. Configure Driverless AI settings to enable this profile. (See the Driverless AI section for more information.)
Note: The minimum Sparkling Water versions are 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.

Adding Profiles

1. On the Configurations page, click the Profiles option to open the Profiles page. 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. **Python Environments**: Select the environment(s) that will be associated with this profile. This list of available environments comes from the Python Environments page.

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

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. **Python Environments**: Select the environment(s) that will be associated with this profile. This list of available environments comes from the *Python Environments* page.

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

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.

**Driverless AI**

1. **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.

2. **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.

3. **Instance Limit Per User**: Specify the maximum number of Driverless AI instances that a user with this profile can access.

4. **DAI Servers**: Specify the Driverless AI servers that a user with this profile can access.

5. **Config Toml Override**: Specify and Driverless AI configuration overrides that will be associated with this profile. A list of available configuration options is available in the Driverless AI config.toml file.

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.

**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** page. 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.
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 page. 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.

Deleting Profiles

1. On the Configurations page, click the Profiles page. 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.

3.2.2 Steam Configuration

The Steam Configuration options allow you to configure settings that were previously available in the steam.yaml file.

**CONFIGURATIONS**

System-wide configuration for Enterprise Steam. Please refer to the documentation. All settings are validated before saving. Some settings require Enterprise Steam to be restarted to take effect.

<table>
<thead>
<tr>
<th>ACCESS CONTROL</th>
<th>STEAM CONFIGURATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentication</td>
<td>Licensing</td>
</tr>
<tr>
<td>Token</td>
<td>Security</td>
</tr>
<tr>
<td>Users</td>
<td>Logging</td>
</tr>
<tr>
<td>Roles</td>
<td>Import/Export</td>
</tr>
<tr>
<td>Profiles</td>
<td></td>
</tr>
</tbody>
</table>

**Licensing**

The Licensing page shows you how long you have left on your current license. If your license expires, you will be prompted to enter a new Enterprise Steam license.
By default, a self-signed TLS certificate will be autogenerated by Steam. It is advised to change this to a valid secure certificate.

Security Configuration

- **TLS Certificate Path**: 
- **TLS Key Path**: 
- **Server Strict Transport**: `max-age=631138519`
- **Server X-XSS Protection**: 0
- **Server Content Security Policy**: `style-src 'self' 'unsafe-inline'`
- **Web UI Timeout (Min)**: 480
- **Disable JupyterHub**: 

Save Configuration
1. Specify the certificate file used by both the Steam process and Steam’s haproxy process.

2. Specify the private key PEM file used by both the Steam process and Steam’s haproxy process.

3. Specify the server strict transport value. 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. This value is in seconds, and the default value is equivalent to 20 years (max-age=631135819). Leave this empty to disable this setting.

4. Specify the server X-XSS protection value. 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 this value is set to 1 and a cross-site scripting attack is detected, the browser will sanitize the page (remove the unsafe parts). This value defaults to 0.

5. Specify the Server Content Security Policy (CSP). 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. This defaults to:

```
style-src 'self' 'unsafe-inline' https://fonts.googleapis.com; font-src 'self' 'unsafe-inline';
font-src 'self' https://fonts.gstatic.com data:;
```

6. Specify the web timeout value in minutes.

7. Specify whether to disable JupyterHub.

8. Click Save Configuration when you are finished.

**Logging**

This page allows you to configure or edit Enterprise Steam logging information and download existing logs.

**LOGGING CONFIGURATION**

- **LOG DIRECTORY**
  
  `/var/log/steam`  

- **LOG LEVEL**
  
  `Debug`  

- **LOG FILE PERMISSIONS**
  
  `644`  

1. Specify the Log Directory. Steam will save application logs into this directory.

2. Specify the Log Level. The Steam log level can be set to Panic, Fatal, Error, Warning, Info, or Debug.

3. Specify the Log File Permissions. This represents the Unix permission of the log files and defaults to 644.
4. Click **Save Configuration** when you are done.

**Import/Export**

This page allows you to import or export current configuration for authentication, YARN, Sparkling Water, Driverless AI, security, and logging.

**IMPORT/EXPORT CONFIGURATION**

Authentication (Local, LDAP or SAML), YARN, Sparkling Water, Driverless AI, Security and Logging settings are part of the exported configuration.

**Import**

- **Upload Configuration File**
- **Browse**

**Export**

- **Download Configuration**

You can import a new configuration file by clicking the **Browse** button and navigating to the configuration file. Or you can download an existing configuration file to your local machine.

### 3.3 Backends

The Backends section allows you to configure the Hadoop backend to enable H2O and Sparkling Water clusters and to configure H2O Agent backend to enable Driverless AI servers.

<table>
<thead>
<tr>
<th>BACKENDS</th>
<th>enabled</th>
<th>BACKENDS</th>
<th>enabled</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HADOOP</strong></td>
<td></td>
<td><strong>H2O AGENT</strong></td>
<td></td>
</tr>
<tr>
<td>Configuration</td>
<td></td>
<td>Configuration</td>
<td></td>
</tr>
</tbody>
</table>

### 3.3.1 Hadoop Configuration

In order to enable H2O and Sparkling Water clusters in Enterprise Steam, you must enable the Hadoop backend. If you Hadoop cluster is in secure mode (protected by Kerberos), then you will have to provide a keytab that Enterprise Steam will use to submit jobs on Hadoop.

**Note:** Your Hadoop admin will have to setup impersonation for Steam user. Follow to the *Setup Hadoop Impersonation* section. You will not be able enable Hadoop backend without it.

1. Click **Configuration** under the Hadoop Backend section. This opens the Hadoop Configuration page, which shows your current Hadoop version, distribution, and jar file location.
2. Click **Enabled** to enable the Hadoop backend.

3. Enter the absolute path to the Hadoop configuration directory.

4. Enter the absolute path where Hadoop tmp files are stored.

5. Specify whether to make all usernames to be lowercase when submitting jobs to Hadoop. This is disabled by default.

6. Specify whether to enable Kerberos. This is disabled by default. If enabled, then specify the following additional settings:
   
   (a) Kerberos principal: Enter the full name of the Kerberos principal associated with the keytab.
   
   (b) Kerberos keytab path: Enter the path to the Kerberos keytab on the filesystem.
   
   (c) Custom KRB5_CONFIG PATH: Optionally enter the path to a custom KRB5_CONFIG file.

7. Click **Save Configuration** when you are done.

---

### HADOOP CONFIGURATION

To enable H2O and Sparkling Water clusters you will have to enable the Hadoop backend. If your Hadoop cluster is in secure mode (protected by Kerberos) you will have to obtain a keytab that Enterprise Steam will use to submit jobs on Hadoop. Hadoop admin will also have to set up impersonation as described in the [documentation](#).

**Hadoop info**

- **Version**: 2.6.0
- **Distribution**: hdp6
- **Location**: /usr/hdp/2.2.6.3-1/hadoop/hadoop-common-2.6.0.2.2.6.3-1.jar

---

**Settings**

- **Enabled**: On
- **Hadoop Conf Dir**: /etc/hadoop/conf
- **Hadoop Temp Dir**: /tmp
- **Lowercase Usernames**: Off
- **Kerberos Enabled**: Off
- **Kerberos Principal**: 
- **Kerberos Keytab Path**: 
- **Custom KRB5_CONFIG PATH**: 

**Save Configuration**
3.3.2 H2O Agent Configuration

To enable the on-premise deployment backend, download and install the H2O Agent on your servers. The H2O Agent is available here: https://h2o-agent.s3.amazonaws.com/release/0.1.0/index.html. After installing the Agent on your servers, you will be able to add those agents on the H2O Agents page.

Adding an H2O Agent

1. Click Configuration under the H2O Backend section. This opens the H2O Agent page, which shows currently available H2O agents.
2. Click Add Agent.
3. Enter a unique name for the agent.
4. Enter the server hostname for the agent.
5. Enter the H2O agent port. This defaults to 57344.
6. Enter the H2O agent’s API token. This is located at /etc/h2o-agent/secret_token.
7. Click Create when you are done.

Viewing H2O Agent Details

Click the Information icon beside the agent that you want to review. This opens a page with details about the agent.
Deleting an H2O Agent

Click the Delete icon ( ) beside the agent that you want to delete. You will be prompted to confirm this action. Click Remove to complete the deletion process, or click Cancel to return to the H2O Agent screen without deleting.

3.4 Products

3.4.1 H2O

The H2O section allows you to enable Enterprise Steam to work with H2O. From this section, you can also add new H2O engines and specify launch parameters to be used when starting H2O.

Configuration

H2O Configuration allows you to enable Enterprise Steam to work with H2O. When enabled, the default-h2o profiles will be available. (See the Profiles section for more information.)

1. Click the Configuration option under the H2O section.
2. Enable H2O.
3. Optionally turn on Internal Secure Connections. This enables SSL internal security.
4. Optionally specify any additional Hadoop CLASSPATH options. These will be added to HADOOP_CLASSPATH when launching H2O clusters.
5. Click Save Configuration when you are done.
Engines

Click the Engines option in the H2O section to view a list of available H2O engines. To install a new H2O engine, download the engine that matches your desired H2O version and Hadoop distribution, then click Browse to upload the file. You can also delete engines that are no longer needed by clicking on the Remove engine icon next to the engine you want to delete.

Launch Parameters

The Launch Parameters option in the H2O section allows you to specify launch parameters to be used when starting H2O. A list of parameters that can be specified is available in the H2O-3 User Guide.

1. Click Add Parameter to add a new launch parameter.
2. Enter the name of the parameter.
3. Specify a value for the parameter.
4. Specify if this parameter is the default or an override.
5. Specify if this is a Hadoop or H2O Driver parameter.
6. Click Create when you are done.
3.4.2 Sparkling Water

The Sparkling Water section allows you to enable Enterprise Steam integration with Sparkling Water. From this section, you can also upload new Sparkling Water engines and new Python environments.

**Configuration**

Sparkling Water Configuration allows you to enable Sparkling Water and/or RSparkling. When enabled, the `default-sparkling-internal` and `default-sparkling-external` profiles will be available. (See the Profiles section for more information.)

1. Click the **Configuration** option under Sparkling Water.
2. Enable Sparkling Water and/or R Sparkling.
3. Specify the path to your Spark home directory (`SPARK_HOME`).
4. Specify the path to your Java 8 home directory (`JAVA_HOME`).
5. Enable Internal Secure Connections.
6. Select a Sparkling Water backend to enforce. You can change this to Internal or External backend.
7. Click **Save Configuration** when you are done.
Engines

Click the Engines option to view a list of available Sparkling Water engines. To install a new Sparkling Water engine, download the engine that matches your version of Spark, then click Browse to upload the ZIP file. You can also delete engines that are no longer needed by clicking on the Remove engine icon next to the engine you want to delete.

Note that the minimum supported versions of Sparkling Water in Enterprise Steam are:

- 2.1.41
- 2.2.27
- 2.3.16
- 2.4.1

Python Environments

Click the Python Environments option to view a list of available Python environments. Enterprise Steam ships with two default environments - Python 2.7 and Python 3.7. These environments allow Enterprise Steam to utilize PySparkling.
**Add New Python Environment**

With Enterprise Steam, you can add a new Python environment from a Python path or from a Conda package.

**Add New Conda Pack**

Perform the following steps to add a new Python environment from a Conda package.

**Note:** The OS type where the environment was built must match the OS type of the target machine.

1. On the Python Environments page, click **Add Environment**.
2. Select **Conda pack** from the Environment Type dropdown menu.
3. Enter a unique name for this new environment.
5. Click **Add Environment** when you are done.

**Add New Python Path**

Perform the following steps to add a new Python environment via a Python path.

1. On the Python Environments page, click **Add Environment**.
2. Select **Python path** from the Environment Type dropdown menu.
3. Enter a unique name for this new environment.
4. Specify the PySpark Python path.
5. Click **Add Environment** when you are done.

**Deleting Python Environments**

**Note:** The default environments cannot be deleted.

Click the **Actions > Remove** option beside the Python environment that you want to delete. A confirmation message will display. Click **Confirm** to complete the removal.
3.4.3 Driverless AI

The Driverless AI section allows you to enable and configure Enterprise Steam to manage Driverless AI instances and add new Driverless AI servers.

Configuration

1. Click the **Configuration** option in the Driverless AI section.
2. Click **Enabled** to enable Driverless AI.
3. Specify the path to the storage directory that is mounted and shared across all Driverless AI servers and the Steam server. When a valid path is specified, then the list of available engines will populate.
4. Paste your Driverless AI license file in the Driverless AI License File text field.
5. Click **Save Configuration** when you are done.
Servers

Click the **Servers** option in the Driverless AI section to view a list of Driverless AI servers.

## Adding a Driverless AI Server

1. Click the **Add Server** button.
2. Select an existing H2O Agent.
3. Click the checkbox beside the profile or profiles that should be assigned to this server.
4. Click **Add Server** to complete the configuration.

## Viewing Server and Instance Details

Click on **Actions > Server** or **Actions > Instance** to view details about this Driverless AI server or the Driverless AI instance(s) associated with this server.
Disabling/Removing Servers

You can disable or remove servers that have no running Driverless AI instances.

1. Click on Actions > Stop instance to stop any currently running instances. (Note that instances can also be stopped on the DAI Instances page.)

2. Click on Actions > Disable to disable the server or Actions > Remove to delete the server from your Enterprise Steam environment.
To upgrade, download the official release from the Enterprise Steam download page.

### 4.1 Backup

Backing up your Enterprise Steam installation before an upgrade is optional but highly recommended. Please backup the `/opt/h2oai/steam/var/master` folder. To revert to a previous version, reinstall that version and restore the backup.

### 4.2 Upgrade Enterprise Steam service

1. Install the Enterprise Steam package.
   
   - **RPM:** `yum localinstall steam-X.Y.Z.x86_64.el6.rpm`
   
   - **DEB:** `sudo dpkg -i steam-X.Y.Z.x86_64.deb`

   There is no need to stop the service before upgrading or restart the service after upgrading. Enterprise Steam will restart automatically.

2. Validate that the Enterprise Steam service is running.
   
   - **RHEL 6/CentOS 6:** `sudo /etc/init.d/steam status`
   
   - **RHEL 7/centOS 7:** `sudo systemctl status steam`
   
   - **Ubuntu/Ubuntu:** `sudo service steam status`

   For troubleshooting, please review the logs (`cat /var/log/steam/steam.log`).

### 4.3 Upgrade Python/R clients

After upgrading Enterprise Steam, users of Python and R need to upgrade their Steam API clients. See the Enterprise Steam download page.

- **Python:** `pip install h2osteam-X.Y.Z-py2.py3-none-any.whl`
- **Conda Python 2.7:** `conda install h2osteam-X.Y.Z-py2.py27_0.tar.bz2`
- **Conda Python 3.6:** `conda install h2osteam-X.Y.Z-py2.py36_0.tar.bz2`
- **Conda Python 3.7:** `conda install h2osteam-X.Y.Z-py2.py37_0.tar.bz2`
- **R:** `R CMD INSTALL h2osteam_X.Y.Z.tar.gz`
When modifying some configurations, you will be prompted to restart Enterprise Steam to apply changes.

1. Restart Enterprise Steam.
   - RHEL 6 / CentOS 6: `sudo /etc/init.d/steam restart`
   - RHEL 7 / CentOS 7: `sudo systemctl restart steam`
   - Ubuntu / Ubuntu: `sudo service steam restart`

2. Validate that the Enterprise Steam service is running.
   - RHEL 6 / CentOS 6: `sudo /etc/init.d/steam status`
   - RHEL 7 / CentOS 7: `sudo systemctl status steam`
   - Ubuntu / Ubuntu: `sudo service steam status`

For troubleshooting, please review the logs (`cat /var/log/steam/steam.log`).
SETUP HADOOP IMPERSONATION

For Enterprise Steam to act on behalf of logged-in users when launching clusters on Hadoop/YARN a Hadoop administrator has to allow Enterprise Steam to do so. This requires changes to Hadoop core-site.xml. Do not change core-site.xml manually; instead use Cloudera Manager, Ambari or similar tool that manages Hadoop configuration.

Hadoop administrator needs to add the following properties to core-site.xml:

```xml
<property>
  <name>hadoop.proxyuser.SERVICEID.hosts</name>
  <value>HOST</value>
</property>
<property>
  <name>hadoop.proxyuser.SERVICEID.groups</name>
  <value>*</value>
</property>
```

where:

- SERVICEID is the user ID of Kerberos principal that is associated with the Enterprise Steam Kerberos keytab or Enterprise Steam service ID (usually steam)
- HOST is the hostname of the Enterprise Steam server. Wildcard (*) is accepted.

Following is an example of valid core-site.xml changes to enable Enterprise Steam on steam.mycompany.loc to impersonate any user:

```xml
<property>
  <name>hadoop.proxyuser.SERVICEID.hosts</name>
  <value>steam.mycompany.loc</value>
</property>
<property>
  <name>hadoop.proxyuser.SERVICEID.groups</name>
  <value>*</value>
</property>
```

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

### 6.1 In Cloudera Manager

1. Login to Cloudera Manager as Hadoop administrator capable of changing Hadoop configuration.
2. Go to HDFS service.
3. Go to Configuration.
4. **Search** for Cluster-wide Advanced Configuration Snippet (Safety Valve) for core-site.xml configuration.

5. **Add entry with name** hadoop.proxyuser.SERVICEID.hosts and value HOST as described in the previous section.

6. **Add entry with name** hadoop.proxyuser.SERVICEID.groups and value * as described in the previous section.

7. Save Changes

8. Deploy client configuration and restart the cluster.