Release notes

Features, Technology Previews, and known issues associated with this release
Features, Technology Previews, and known issues associated with this release
Abstract

These release notes provide an overview of new features, enhancements, major technical changes, and any known bugs in the version of Red Hat OpenShift Data Science currently available in Red Hat OpenShift Dedicated and Red Hat OpenShift Service on Amazon Web Services (ROSA).
Table of Contents

PREFACE .................................................................................................................. 3

CHAPTER 1. OVERVIEW OF OPENSFIFT DATA SCIENCE ........................................ 4

CHAPTER 2. PRODUCT FEATURES .......................................................................... 5
  2.1. FEATURES FOR DATA SCIENTISTS .............................................................. 5
  2.2. FEATURES FOR IT OPERATIONS ADMINISTRATORS ..................................... 5
  2.3. ENHANCEMENTS ............................................................................................ 5

CHAPTER 3. BUG FIXES ......................................................................................... 7

CHAPTER 4. KNOWN ISSUES .................................................................................. 9
See the following documents for service and life cycle information related to this release:

- OpenShift Data Science Service Definition
- OpenShift Data Science Life Cycle
CHAPTER 1. OVERVIEW OF OPENSHIFT DATA SCIENCE

Using Red Hat OpenShift Data Science, users can integrate data, artificial intelligence and machine learning software to execute end-to-end machine learning workflows. OpenShift Data Science is available as an Add-on to Red Hat managed environments such as Red Hat OpenShift Dedicated and Red Hat OpenShift Service on Amazon Web Services (ROSA).

For data scientists, OpenShift Data Science includes JupyterHub and a collection of default notebook images optimized with the tools and libraries required for model development, and the TensorFlow and PyTorch frameworks. Deploy and host your models, integrate models into external applications, and export models to host them in any hybrid cloud environment. You can also accelerate your data science experiments through the use of graphics processing units (GPUs).

For administrators, OpenShift Data Science enables data science workloads in an existing Red Hat OpenShift Dedicated or ROSA environment. Manage users with your existing OpenShift identity provider, and manage the resources available to notebook servers to ensure data scientists have what they require to create, train, and host models.
CHAPTER 2. PRODUCT FEATURES

Red Hat OpenShift Data Science provides a number of features for data scientists and IT operations administrators.

2.1. FEATURES FOR DATA SCIENTISTS

One-page JupyterHub notebook server configuration
Choose from a default set of notebook images pre-configured with the tools and libraries you need for model development.

Collaborate on notebooks using Git
Use JupyterLab’s Git interface to work collaboratively with application developers or add other models to your notebooks.

Integrate with Red Hat OpenShift Streams for Apache Kafka
Integrate fault-tolerant real-time data streams into your notebooks and machine learning models by connecting OpenShift Data Science to Red Hat OpenShift Streams for Apache Kafka.

Deploy using application templates
Red Hat provides application templates designed for data scientists so that you can easily deploy your models and applications for testing purposes on either Red Hat OpenShift Dedicated or Red Hat OpenShift Service on Amazon Web Services (ROSA).

Try it out in the Red Hat Developer sandbox environment
You can try out OpenShift Data Science and access tutorials and activities in the Red Hat Developer sandbox environment.

2.2. FEATURES FOR IT OPERATIONS ADMINISTRATORS

Install as an Add-on
Install the OpenShift Data Science as an Add-on to your Red Hat OpenShift Dedicated or Red Hat OpenShift Service on Amazon Web Services (ROSA) cluster using Red Hat OpenShift Cluster Manager (http://console.redhat.com/openshift/).

Manage users with your existing identity provider
OpenShift Data Science supports the same identity providers as OpenShift Dedicated and ROSA. You can configure existing groups in your identity provider as administrators or users of OpenShift Data Science.

Manage resources with OpenShift Dedicated
Use your existing OpenShift Dedicated or ROSA knowledge to configure and manage machine pools for your OpenShift Data Science users.

Customize PVC size to suit your workloads
Allocate the right amount of persistent storage for your data scientists by default to optimize resource costs and productivity.

2.3. ENHANCEMENTS

This section describes enhancements to existing features in Red Hat OpenShift Data Science.

Default persistent volume claim (PVC) size increased
The default size of a PVC provisioned for a data science user in an OpenShift Data Science cluster has been increased from 2 GB to 20 GB.
**Improved resilience to OpenShift Dedicated node failure**

OpenShift Data Science services now try to avoid being scheduled on the same node so that OpenShift Data Science components are more failure resistant.
CHAPTER 3. BUG FIXES

This section describes the fixes for notable user-facing issues in Red Hat OpenShift Data Science.

Incorrect number of available GPUs were displayed in JupyterHub

When a user attempted to create a notebook instance in JupyterHub, the maximum number of GPUs available for scheduling was not updated as GPUs were assigned. As a result, there was a delay if the user requested a GPU that was already assigned.

Changing alert notification emails required pod restart

Changes to the list of notification email addresses in the Red Hat OpenShift Data Science Add-On were not applied until after the rhods-operator pod and the prometheus-* pod were restarted.

Red Hat OpenShift API Management 1.15.2 add-on installation did not successfully complete

For OpenShift Data Science installations that are integrated with the Red Hat OpenShift API Management 1.15.2 add-on, the Red Hat OpenShift API Management installation process did not successfully obtain the SMTP credentials secret. Subsequently, the installation did not complete.

Pachyderm now compatible with OpenShift Dedicated 4.10 clusters

Pachyderm was not initially compatible with OpenShift Dedicated 4.10, and so was not available in OpenShift Data Science running on an OpenShift Dedicated 4.10 cluster. Pachyderm is now available on and compatible with OpenShift Dedicated 4.10.

Uninstall process failed to complete when both OpenShift Data Science and OpenShift API Management were installed

When OpenShift Data Science and OpenShift API Management are installed together on the same cluster, they use the same Virtual Private Cluster (VPC). The uninstall process for these Add-ons attempts to delete the VPC. Previously, when both Add-ons are installed, the uninstall process for one service was blocked because the other service still had resources in the VPC. The cleanup process has been updated so that this conflict does not occur.

Images were incorrectly updated after upgrading OpenShift Data Science

After the process to upgrade OpenShift Data Science completed, JupyterHub failed to update its notebook images. This was due to an issue with the image caching mechanism. Images are now correctly updating after an upgrade.

Incorrect TensorFlow and TensorBoard versions displayed during notebook selection

The Start a notebook server page displayed incorrect version numbers (2.4.0) for TensorFlow and TensorBoard in the TensorFlow notebook image. These versions have been corrected to TensorFlow 2.7.0 and TensorBoard 2.6.0.

Quick start links did not display for enabled applications

For some applications, the Open quick start link failed to display on the application’s card on the Enabled page. As a result, users did not have direct access to the quick start tour for the relevant application.

Incorrect Python versions displayed during notebook selection

The Start a notebook server page displayed incorrect versions of Python for the TensorFlow and PyTorch notebook images. Both images incorrectly displayed Python 3.8.6. The actual version used was Python 3.8.8. Additionally, the third integer of package version numbers is now no longer displayed.

Missing step in Getting Started with OpenShift Streams for Apache Kafka
The guided tour for OpenShift Streams for Apache Kafka missed a step on assigning read permissions to the service account. This step is now included, allowing users to complete the guided tour without issues.
CHAPTER 4. KNOWN ISSUES

This section describes known issues in Red Hat OpenShift Data Science and any known methods of working around the issues described.

Starburst Galaxy quick start does not provide download link in the instruction steps
The Starburst Galaxy quick start, located on the Resources page on the dashboard, requires the user to open the explore-data.ipynb notebook, but fails to provide a link within the instruction steps. Instead, the link is provided in the quick start’s introduction.

Unavailable images set as default selection in JupyterHub during a rebuild
During a rebuild of the Pytorch and TensorFlow images, if either of these images is specified in the user’s configmap as the last selected image, the image remains selected by default in JupyterHub.

GPU tutorial does not appear on dashboard
The "GPU computing" tutorial, located at Gtc2018-numba, does not appear on the Resources page on the dashboard.

Error can occur when creating a notebook instance
When creating a notebook instance in JupyterHub, a Directory not found error appears intermittently. This error message can be ignored by clicking Dismiss.

Ten minute wait after notebook server launch failed
If the JupyterHub leader pod fails while the notebook server is being launched, the user cannot access their notebook server until the pod restarts. This takes approximately ten minutes.
Workaround: Click Try Again in the notebook server launch interface.

GPU selection persists when GPU nodes are unavailable
If a user provisions a notebook server with GPU support, and the utilized GPU nodes are subsequently removed from the cluster, the user cannot create a notebook server. This occurs because the most recently used setting for the number of attached GPUs is used by default. +
Workaround: Manually edit the jupyterhub-singleuser-profile-<username> ConfigMap in the rhods-notebooks namespace and set the gpuCount to 0.

Seldon unavailable on OpenShift Dedicated 4.9 or higher
Seldon is not available in OpenShift Data Science because it is not yet compatible with OpenShift Dedicated 4.9 or higher. OpenShift Dedicated is used by OpenShift Data Science. There is currently no workaround for this issue.

Terminal access unavailable in Pachyderm pod
The pachd pod created by Pachyderm intentionally does not provide Terminal access to OpenShift Dedicated administrators. The Terminal view in the OpenShift Dedicated web console is provided regardless of whether pods provide access. If you attempt to access the Terminal view for the pachd pod, the following expected error appears:

```
ERRO[0000] exec failed: container_linux.go:367: starting container process caused: exec: "sh": executable file not found in $PATH
command terminated with non-zero exit code: exit status 1
The terminal connection has closed.
```

Actions on dashboard not clearly visible
The dashboard actions to re-validate a disabled application’s license, and to remove a disabled
application's card are not clearly visible to the user. These actions only appear when the user clicks on the application card's Disabled label. As a result, the intended workflows may not be clear to the user.

License re-validation action appears unnecessarily

The dashboard action to re-validate a disabled application's license appears unnecessarily for applications that do not have a license validation or activation system. In addition, when a user attempts to re-validate a license that cannot be re-validated, feedback is not displayed to state why the action cannot be completed.

Error can occur during Pachyderm deployment

When creating an instance of the Pachyderm operator, a webhook error appears intermittently, preventing the creation process from starting successfully. The webhook error is indicative that, either the Pachyderm operator failed a health check, causing it to restart, or that the operator process exceeded its container's allocated memory limit, triggering an Out of Memory (OOM) kill. **Workaround:** Repeat the Pachyderm instance creation process until the error no longer appears.

Error occurs while fetching the generated images in the sample Pachyderm notebook

An error occurs when a user attempts to fetch an image using the sample Pachyderm notebook in JupyterHub. The error states that the image could not be found. **Workaround:** If the error stating that the image could not be found occurs, run the following commands:

```bash
$ oc adm policy add-scc-to-user anyuid -z pachyderm-worker
$ oc scale rc/pipeline-edges-v1 --replicas=0; oc scale rc/pipeline-edges-v1 --replicas=1
```

This command must be performed by an administrator with **cluster-admins** role privileges.

PyTorch and TensorFlow images are unavailable in JupyterHub during a rebuild

Users are unable to launch PyTorch and TensorFlow notebooks in JupyterHub during a rebuild.

PyTorch and TensorFlow images are unavailable when upgrading

When upgrading from OpenShift Data Science 1.3 to a later version, PyTorch and TensorFlow images are unavailable to users for approximately 30 minutes. As a result, users are unable to launch PyTorch and TensorFlow notebooks in JupyterHub during the upgrade process. **Workaround:** Users that launch their TensorFlow or PyTorch images prior to the upgrade do not experience any disruption. Therefore, it is recommended that users launch their PyTorch or TensorFlow notebooks prior to commencing the upgrade.

IBM Watson Studio not available in OpenShift Data Science

IBM Watson Studio is not available when OpenShift Data Science is installed on OpenShift Dedicated 4.9 or higher, because it is not compatible with these versions of OpenShift Dedicated. Contact **Marketplace support** for assistance manually configuring Watson Studio on OpenShift Dedicated 4.9 and higher.

Gateway errors during notebook server creation

If the leader JupyterHub pod fails during notebook server creation and a new leader pod is not selected before a user is redirected to their notebook server, users may see either a 504 Gateway Timeout error page or a 502 Bad Gateway error page. A new leader pod is selected after a few seconds. To recover from this error, wait a few seconds and then refresh the page.
Unnecessary warnings about missing Graphical Processing Units (GPUs)

The TensorFlow notebook image checks for graphical processing units (GPUs) whenever a notebook is run, and issues warnings about missing GPUs when none are present. These messages can safely be ignored, but you can disable them by running the following in a notebook when you launch a notebook server that uses the TensorFlow notebook image.

```python
import os
os.environ['TF_CPP_MIN_LOG_LEVEL'] = '3'
```

Cannot delete Git repositories in JupyterLab file browser

When a user attempts to delete a directory using the JupyterLab file browser, deletion fails if the directory is not empty. Hidden files such as the .git directory in a Git repository are not shown in the JupyterLab file browser, so Git repositories cannot be deleted from the JupyterLab file browser.

**Workaround:** To delete a Git repository from JupyterLab:

a. Use the JupyterLab launcher to open a Terminal.

b. Run the remove command, `rm -rf <path>`, replacing `<path>` with the path to the Git repository directory, for example, `repos/my-project-repo`.

Cannot set container size during notebook server creation

The Container size dropdown menu is intermittently not displayed on the Create a notebook server page. Users cannot select a container size other than the default if this menu does not display. You may be able to trigger the correct behavior by refreshing the page.

Previously authenticated sessions persist after user configuration change

When an administrator logs in to JupyterHub and later configures a custom user group to replace a default user group, the JupyterHub session that was initially authenticated using the default group persists for up to five minutes in the same browser window. This mainly affects administrators attempting to test permissions after adding or removing a custom user group for their identity provider.

**Workaround:** After changing user group configuration, manually log out of all sessions before testing updated user permissions.

OpenShift Data Science hyperlink still visible after uninstall

When the OpenShift Data Science Add-on is uninstalled from an OpenShift Dedicated cluster, the link to the OpenShift Data Science interface remains visible in the application launcher menu. Clicking this link results in a "Page Not Found" error because OpenShift Data Science is no longer available.

User sessions persist in some components

Although users of OpenShift Data Science and its components are authenticated through OpenShift, session management is separate from authentication. This means that logging out of OpenShift Dedicated or OpenShift Data Science does not affect a logged in JupyterHub session running on those platforms. When a user’s permissions change, that user must log out of all current sessions so that changes take effect.

Deleted users stay logged in to JupyterHub for up to 5 minutes

When a user’s permissions for JupyterHub are revoked, it takes up to five minutes for JupyterHub to log the user out. After a user has been removed from a valid user group, the user is able to spawn a new notebook server for about 30 seconds, and is able to continue working in JupyterLab for up to
five minutes before they are logged out.

**Removed users are shown in the JupyterHub administrative interface**

When a user’s permission to access JupyterHub is revoked, they are prevented from creating or starting notebook servers, but their user name still appears in the list of users in the JupyterHub administrative interface. This happens because the cleanup step to remove that user from JupyterHub’s user list is missing. There is currently no customer workaround for this issue.

**Notebook servers shut down after 24 hours**

A JupyterHub user can be logged in for a maximum of 24 hours. After 24 hours, user credentials expire, the user is logged out of JupyterHub, and their notebook server pod is stopped and deleted regardless of any work running in the notebook server. There is currently no workaround for this issue. However, you can configure OAuth tokens to expire after a set period of inactivity. See [Configuring the internal OAuth server](#) for more information.