Red Hat OpenShift Data Science 1

Release notes

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

Red Hat OpenShift Data Science 1 Release notes

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

Legal Notice

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 AWS.

Table of Contents

PREFACE

This documentation is provided for the Field Trial release of Red Hat OpenShift Data Science.

See the following documents for service and life cycle information related to this Field Trial 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 (AWS).

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, including support for 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.

For administrators, OpenShift Data Science enables data science workloads in an existing Red Hat OpenShift Dedicated or Red Hat OpenShift on AWS 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 on OpenShift Dedicated for testing.

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

Sign up for a Field Trial and then install the OpenShift Data Science as an Add-on to your OpenShift Dedicated cluster using Red Hat Cluster Manager.

Manage users with your existing identity provider

OpenShift Data Science supports the same identity providers as OpenShift Dedicated. 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 knowledge to configure and manage machine pools for your OpenShift Data Science users.

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.

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.

Ten minute wait after notebook server launch failed

If the JupyterHub leader pod failed while the notebook server was being launched, the user could not access their notebook server until the pod restarted, which took approximately ten minutes. This process has been improved so that the user is redirected to their server when a new leader pod is elected. If this process times out, users see a 504 Gateway Timeout error, and can refresh to access their server.

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.

IBM Watson Studio not available in OpenShift Data Science 1.3

IBM Watson Studio is not available in OpenShift Data Science 1.3 because it is not yet compatible with OpenShift Dedicated 4.9, which is used by OpenShift Data Science. There is currently no workaround for this issue.

Incorrect package versions displayed during notebook selection

The **Start a notebook server** page displays incorrect versions of Python for the Tensorflow and PyTorch notebook images. Both images display Python 3.8.6 but actually use Python 3.8.8.

Uninstall does not work when OpenShift API Management is also installed

When OpenShift Data Science and OpenShift API Management are installed together on the same cluster, they use the same Virtual Private Cloud (VPC). The uninstall process for these Add-ons attempts to delete the VPC. When both Add-ons are installed, the uninstall process for one service is blocked because the other service still has resources in the VPC.

Workaround: Before uninstalling OpenShift Data Science, run the following command to edit the operator definition, and remove any lines related to finalizers:

```
$ oc edit postgres.integreatly.org -n redhat-rhods-operator
```

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 502 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.

```
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.

Changing alert notification emails requires pod restart

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

Workaround: To apply the changed configuration:

    a. Change into the **redhat-ods-operator** project and restart the **rhods-operator** pod.

    b. Wait for the **rhods-operator** pod to restart and return to the **Running** state.

    c. Change into the **redhat-ods-monitoring** project and restart the **prometheus-*** pod.

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.