



Red Hat OpenShift Data Science 1

Red Hat OpenShift Data Science 入门

了解如何在 OpenShift Data Science 环境中工作

了解如何在 OpenShift Data Science 环境中工作

法律通告

Copyright © 2023 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at

<http://creativecommons.org/licenses/by-sa/3.0/>

. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux[®] is the registered trademark of Linus Torvalds in the United States and other countries.

Java[®] is a registered trademark of Oracle and/or its affiliates.

XFS[®] is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL[®] is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js[®] is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack[®] Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

摘要

登录并启动您的笔记本（notebook）服务器，开始在 Jupyter 中使用笔记本。


目录

第 1 章 登陆到 OPENSIFT DATA SCIENCE	3
第 2 章 OPENSIFT DATA SCIENCE 用户界面	4
2.1. 全局导航	5
2.2. 侧导航	5
第 3 章 OPENSIFT DATA SCIENCE 中的通知	7
第 4 章 创建数据科学项目	8
第 5 章 创建项目工作台	9
5.1. 启动 JUPYTER 并启动笔记本服务器	10
5.2. 笔记本服务器环境的选项	11
第 6 章 数据科学工作者教程	18
6.1. 访问教程	20
第 7 章 启用与 OPENSIFT DATA SCIENCE 连接的服务	22
第 8 章 禁用连接到 OPENSIFT DATA SCIENCE 的应用程序	23
8.1. 从 OPENSIFT DATA SCIENCE 中删除禁用的应用程序	24
第 9 章 支持要求和限制	25
9.1. 支持的浏览器	25
9.2. 支持的服务	25
9.3. 支持的软件包	26
第 10 章 常见问题	27
第 11 章 管理员对 JUPYTER 中的常见问题进行故障排除	28
11.1. 用户收到 404 : PAGE NOT FOUND 错误 (在登录到 JUPYTER	28
11.2. 用户的笔记本服务器没有启动	28
11.3. 用户运行笔记本的 CELLS 时遇到 DATABASE OR DISK IS FULL 错误或 NO SPACE LEFT ON DEVICE 错误。	29
第 12 章 对 JUPYTER 中的常见问题进行故障排除	30
12.1. 我可以看到 403:FORBIDDEN 错误 (当登陆到 JUPYTER 时)	30
12.2. 我的笔记本服务器没有启动	30
12.3. 在运行我的笔记本时遇到一个 DATABASE OR DISK IS FULL 错误或一个 NO SPACE LEFT ON DEVICE 错误	30

第 1 章 登陆到 OPENSIFT DATA SCIENCE

使用浏览器登录到 OpenShift Data Science，以便轻松访问 Jupyter 和您的数据科学项目。

步骤

1. 浏览到 OpenShift Data Science 实例 URL，再点 **Log in with OpenShift**.
 - 如果您是数据科学家用户，您的管理员必须为您提供 OpenShift Data Science 实例 URL，例如 **https://rhods-dashboard-redhat-ods-applications.apps.example.abc1.p1.openshiftapps.com/**
 - 如果您可以访问 OpenShift Dedicated，可以访问 OpenShift Dedicated Web 控制台并点击 **Application Launcher** () → **Red Hat OpenShift Data Science**.
2. 单击身份提供程序的名称，如 **GitHub**。
3. 输入您的凭证并点击 **Log in**（或您的身份提供程序的相关链接）。

验证

- OpenShift Data Science 在 **Enabled applications** 页面中打开。

故障排除

- 如果您在登陆时遇到 **An authentication error occurred** 或 **Could not create user** 问题：
 - 您可能错误地输入您的凭证。确认您的凭证正确。
 - 您可能在多个配置的身份提供程序中都有一个帐户。如果您之前使用其他身份提供程序登录，请尝试使用该身份提供程序。

其他资源

- [启动 Jupyter 并启动笔记本服务器](#)

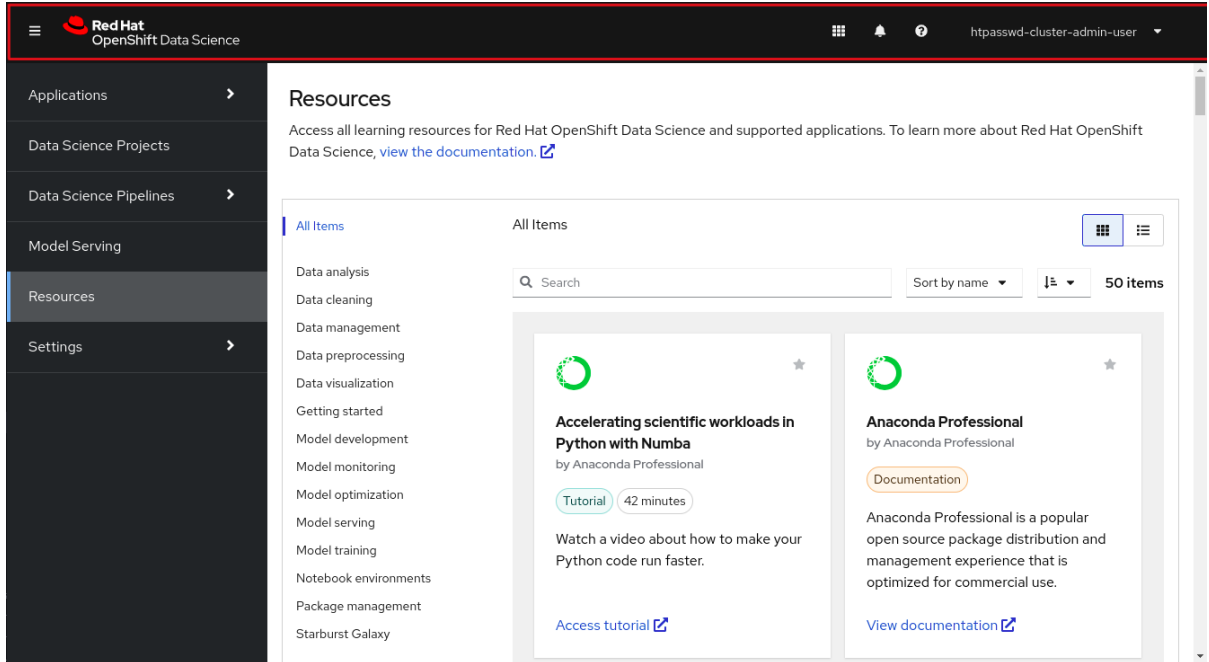
第 2 章 OPENSIFT DATA SCIENCE 用户界面

Red Hat OpenShift Data Science 界面基于 OpenShift Web 控制台界面。

Red Hat OpenShift Data Science 用户界面被分为几个区域：

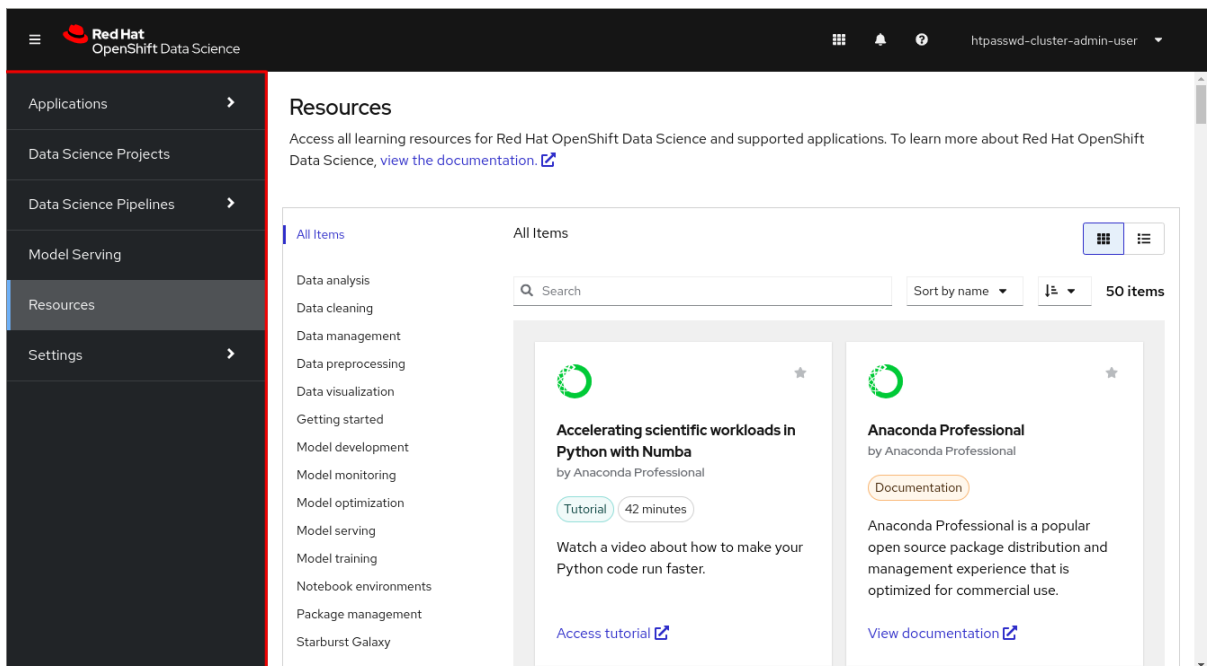
- 全局导航栏，它提供对有用的控制的访问，如帮助和通知。

图 2.1. 全局导航栏



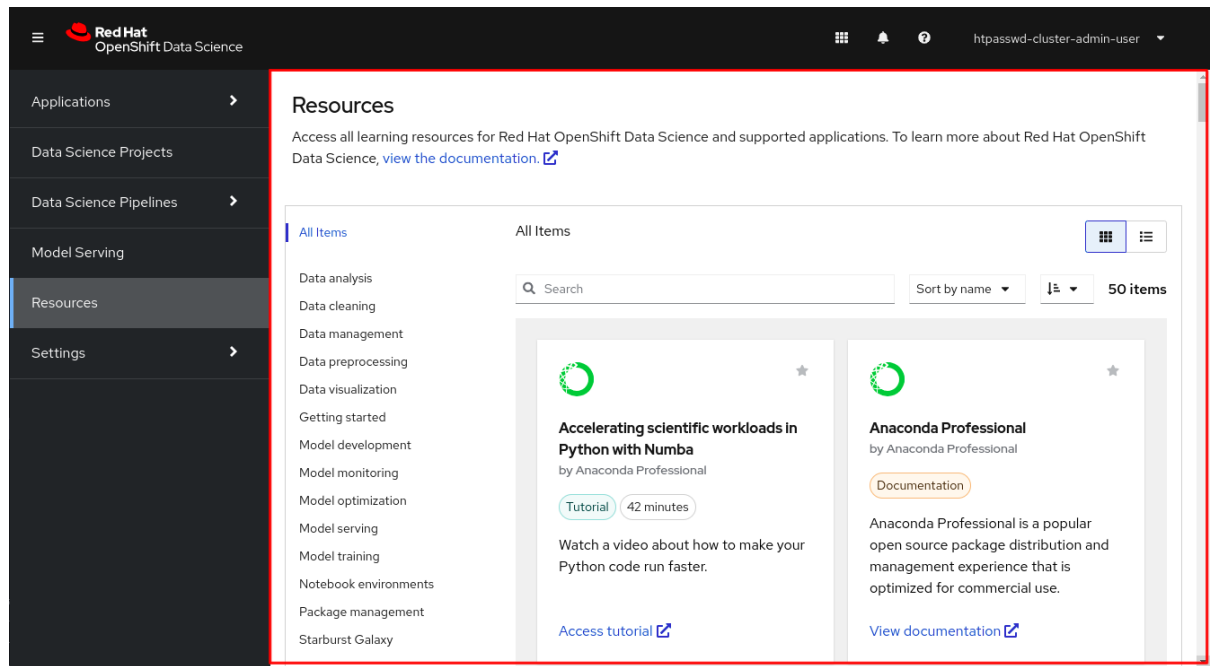
- 侧边导航菜单，其中包含 OpenShift Data Science 中提供的不同页面类别。

图 2.2. 侧边导航菜单






- 主显示区域，其中显示当前页面，并与当前显示信息（如通知或快速启动指南）共享空间。主显示区域还显示 笔记本服务器控制面板，您可以在其中启动 Jupyter 来启动 Jupyter。管理员还可以使用笔记本服务器控制面板管理其他用户的笔记本服务器。

图 2.3. 主显示区域



2.1. 全局导航

顶部导航中有四个项：

- **Toggle 导航菜单按钮** () 可切换是否显示侧边导航。
- **通知 按钮** () 会打开和关闭 **Notifications** 抽屉，您可以更加详细地阅读当前和之前的通知。
- **Help 菜单** () 提供了创建红帽支持问题单的链接，并访问 OpenShift Data Science 文档。
- **用户 菜单** 显示当前登录用户的名称，并提供了 **注销按钮**。

2.2. 侧导航

侧边导航有三个主要部分：

Applications → Enabled

Enabled 页面显示已启用，并可以在 OpenShift Data Science 中使用的应用程序。本页是 OpenShift Data Science 的默认登录页面。

点击应用程序卡上的 **Launch application** 按钮，在新标签页中打开应用程序接口。如果应用程序有一个关联的快速启动导览，点应用程序卡中的下拉菜单，然后选择 **Open quick start** 以访问它。本页还显示您的管理员已禁用的应用程序和组件。禁用的应用程序使用 **Disabled** 在应用程序的卡中表示。点击应用程序卡中的 **Disabled** 可以使用一个链接删除卡本身，以及如果许可证之前已过期，重新验证其许可证。

Applications → Explore

Explore 页面显示可用于 OpenShift Data Science 的应用。点击卡获取应用程序或访问 **启用** 按钮的更多信息。只有在应用程序不需要 OpenShift Operator 安装时，才会看到 **Enable** 按钮。

数据科学项目

Data Science 项目页面允许您将数据科学工作组织到一个项目中。在此页面中，您可以创建和管理数据科学项目。您还可以通过添加工作台、向项目集群添加存储、添加数据连接和添加模型服务器来提高数据科学项目的功能。

Data Science Pipelines → Pipelines

Pipelines 页面允许您导入、管理、跟踪和查看数据科学管道。使用 Red Hat OpenShift Data Science 管道，您可以标准化和自动化机器学习工作流，以便您开发和部署数据科学模型。

Data Science Pipelines → Runs

Runs 页面允许您定义、管理和跟踪数据科学管道的执行。管道运行是数据科学管道的单一执行。您还可以查看之前为数据科学项目执行和调度的运行记录。

Model Serving

Model Serving 页面允许您管理和查看部署模型的状态。您可以使用此页面来部署数据科学模型来提供智能应用程序，或者查看现有的部署模型。您还可以确定已部署模型的 inference 端点。

资源

Resources 页面中显示学习资源，如文档、方法材料以及快速入门导览。您可以使用左侧显示的选项过滤可见的资源，或者在搜索栏中输入术语。

设置 → 笔记本镜像

笔记本镜像 设置页面允许您为项目的特定要求配置自定义笔记本镜像。将自定义笔记本镜像添加到 OpenShift Data Science 部署后，在创建笔记本服务器时可以选择它们。

设置 → 集群设置

集群 设置页面允许您在集群中执行以下管理任务：

- 启用或禁用红帽收集集群上 OpenShift Data Science 使用的数据。
- 通过更改集群持久性卷声明 (PVC) 的默认大小来配置在集群中声明资源的方式。
- 通过停止已闲置的笔记本服务器，减少 OpenShift Data Science 部署中的资源使用。
- 通过添加容限，在污点节点上调度笔记本 pod。

Settings → User management

通过 **User and group settings** 页面，您可以定义 OpenShift Data Science 用户组和 admin 组成员资格。

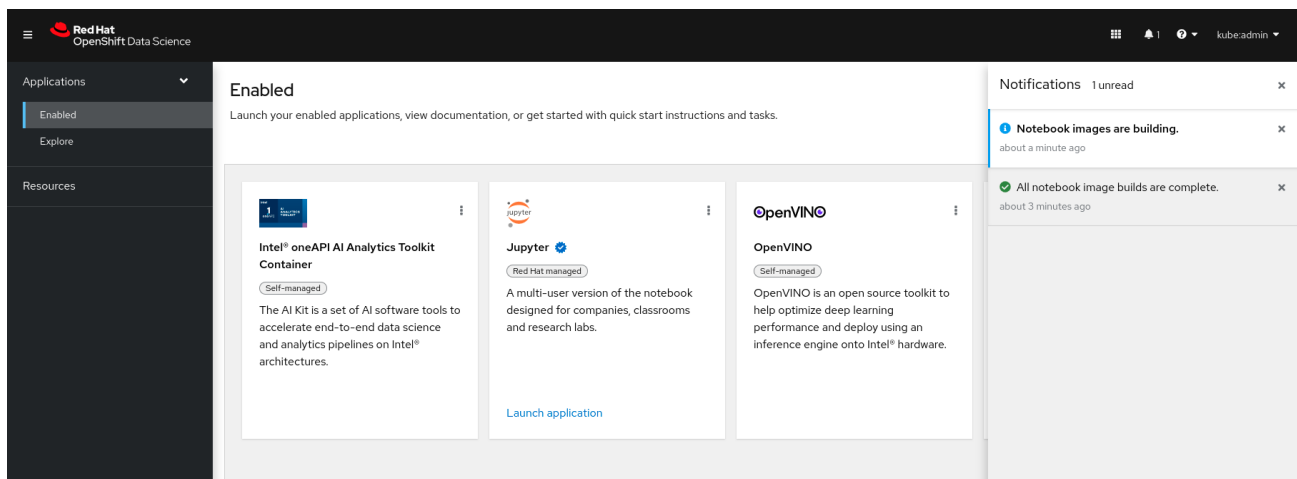
第 3 章 OPENSIFT DATA SCIENCE 中的通知

当集群中的重要事件发生时，Red Hat OpenShift Data Science 会显示通知。

当通知消息被触发时，Red Hat OpenShift Data Science 界面的左下角会显示通知消息。

如果您错过通知信息，请点击 **通知** 按钮()以打开 **Notifications** 抽屉并查看未读取的信息。

图 3.1. 通知抽屉 (Notifications drawer)



第 4 章 创建数据科学项目

要开始数据科学工作，请创建一个数据科学项目。创建项目可帮助您在同一位置组织工作。您还可以通过添加工作台、向项目集群添加存储、添加数据连接和添加模型服务器来提高数据科学项目的功能。

先决条件

- 您已登录到 Red Hat OpenShift Data Science。
- 如果您使用专用的 OpenShift Data Science 组，则作为 OpenShift 中的用户组或 admin 组（如 **rhods-users** 或 **rhods-admin**）的一部分。

步骤

1. 在 OpenShift Data Science 仪表板中，点 **Data Science Projects**。
Data Science 项目页将打开。
2. 点 **Create data Science 项目**。
此时会打开 **Create a data Science project**对话框。
3. 输入您的数据科学项目的**名称**。
4. 可选：编辑 data Science 项目的**资源名称**。资源名称只能包含小写字母数字字符 -，需要以字母数字字符开头和结尾。
5. 输入您的数据科学项目的**描述**。
6. 点 **Create**。
此时会打开项目详情页面。在这里，您可以创建工作台，添加集群存储，以及向项目添加数据连接。

验证

- 您创建的数据科学项目显示在 **Data Science 项目**页中。

第 5 章 创建项目工作台

要在隔离的区域中检查和使用数据模型，您可以创建一个工作台。此工作台可让您从现有笔记本容器镜像创建一个新的 Jupyter 笔记本，以访问其资源和属性。对于需要保留数据的数据科学项目，您可以在您要创建的工作台中添加容器存储。

先决条件

- 您已登陆到 Red Hat OpenShift Data Science。
- 如果您使用专用的 OpenShift Data Science 组，则作为 OpenShift 中的用户组或 admin 组（如 **rhods-users** 或 **rhods-admin**）的一部分。
- 您已创建了一个数据科学项目，您可向其中添加工作台。

步骤

1. 在 OpenShift Data Science 仪表板中，点 **Data Science Projects**。
Data Science 项目页将打开。
2. 点您要在工作台中添加到的项目的名称。
将打开项目的 **Details** 页面。
3. 在 **Workbenches** 部分中，点 **Create workbench**。
Create workbench 页面将打开。
4. 配置您要创建的工作台的属性。
 - a. 输入工作台的**名称**。
 - b. 输入工作台的**描述**。
 - c. 选择用于工作台服务器的**笔记本镜像**。
 - d. 为服务器选择**容器大小**。
 - e. 可选：选择并指定任何新的**环境变量**的值。
 - f. 配置 OpenShift Data Science 集群的存储。
 - i. 选择 **Create new persistent storage** 来创建在从 OpenShift Data Science 后保留的存储。填写相关字段以定义存储。
 - ii. 选择 **Use existing persistent storage** 来重新使用现存的存储，然后从 **Persistent storage** 列表中选择存储。
5. 点 **Create workbench**。

验证

- 您创建的工作台会出现在项目的 **Details** 页面中。
- 您在创建过程中与工作台关联的任何集群存储都会出现在项目的 **Details** 页面中。
- **Status** 列位于 **Details** 页面的 **Workbenches** 部分中，在工作台服务器启动时显示 **Starting** 状态，并在工作台成功启动时显示 **Running**。

5.1. 启动 JUPYTER 并启动笔记本服务器

启动 Jupyter 并启动笔记本服务器以使用您的笔记本。

先决条件

- 您已登录到 Red Hat OpenShift Data Science。
- 您知道笔记本服务器环境中要用于任何环境变量的名称和值，例如 **AWS_SECRET_ACCESS_KEY**。
- 如果您需要使用非常大的数据集，请与管理员合作主动提高笔记本服务器的存储容量。

步骤

1. 在 **Enabled applications** 页面中找到 **Jupyter** 卡。

2. 点 **Launch application**。

如果您看到 **Access 权限所需** 的消息，代表您不在 OpenShift Data Science 的默认用户组或默认管理员组中。请联络您的管理员，将您加入到正确的组中（根据 [Adding users for OpenShift Data Science](#)。）

如果您之前没有授权 **jupyter-nb-<username>** 服务帐户来访问您的帐户，则 **Authorize Access** 页面会显示提示您提供授权。检查默认选择的权限，然后单击 **Allow selected permissions** 按钮。

如果接受凭证，笔记本服务器控制面板将打开开始一个笔记本服务器页面。

3. 启动一个笔记本服务器。

如果您之前打开了 Jupyter，则不需要此项。

- a. 选择用于您的服务器的笔记本镜像。
- b. 如果笔记本镜像包含多个版本，请从 **Versions** 部分中选择笔记本镜像的版本。



注意

当发布新版本的笔记本镜像时，上一版本仍然可用并被支持。这为您提供将工作迁移到最新版本的笔记本镜像。

- c. 为您的服务器选择容器大小。
- d. 可选：为您的服务器选择 **GPU 数**（图像处理单元）。



重要

只有在 PyTorch、TensorFlow 和 CUDA 笔记本服务器镜像中才支持使用 GPU 来加快工作负载性能。另外，只有在集群中启用了 GPU 时，您可以指定笔记本服务器所需的 GPU 数量。要了解如何启用 GPU 支持，请参阅在 [OpenShift Data Science](#) 中启用 GPU 支持。

- e. 可选：选择并指定任何新的环境变量的值。
这个接口会存储这些变量，以便您只需要输入一次。为经常集成的环境和框架（如 Amazon Web Services(AWS)）自动提供常见环境变量的变量名称示例。



重要

确保您为具有敏感值（如密码）的敏感值选择 **Secret** 复选框。

- f. 可选：如果需要，选择 **Start server in current tab**
- g. 单击 **Start server**。
此时会显示 **Starting server** 进程的状态。点 **Expand 事件日志**，以查看服务器创建流程的附加信息。根据您请求的部署大小和资源，启动服务器可能需要几分钟的时间。点 **Cancel** 以取消服务器创建。

服务器启动后，您会看到以下行为之一：

- 如果您之前选择了 **Start server in current tab** 则 JupyterLab 界面会在 Web 浏览器的当前标签页中打开。
- 如果您之前没有选中 **Start server in current tab** 则启动服务器对话框会提示您在新浏览器标签页或当前标签页中打开服务器。
JupyterLab 接口根据您的选择打开。

验证

- 此时会打开 JupyterLab 接口。

其他资源

[笔记本服务器环境的选项。](#)

故障排除

- 如果您遇到 "Unable to load notebook server configuration options" 错误信息，请联系您的管理员来检查与您的 Jupyter pod 相关的日志信息，以进一步分析存在的问题。

5.2. 笔记本服务器环境的选项

当您首次启动 Jupyter 时，或者在停止笔记本服务器后，必须在启动笔记本服务器向导中选择服务器选项，以便您期望的软件和变量在服务器上可用。本节介绍了启动笔记本服务器向导中提供的选项详情。

启动笔记本服务器页面被分为几个部分：

笔记本镜像

指定笔记本服务器所基于的容器镜像。不同的笔记本镜像会默认安装不同的软件包。如果笔记本镜像包含多个版本，您可以从 **Versions** 部分中选择要使用的笔记本镜像版本。



注意

笔记本镜像最少支持一年。预配置笔记本镜像的主要更新大约每 6 个月进行一次。因此，通常会在任何给定时间提供两个支持的笔记本镜像。要使用最新的软件包版本，红帽建议您使用最近添加的笔记本镜像。

启动笔记本镜像后，您可以通过在笔记本单元中运行 **pip** 工具来检查在笔记本服务器上安装了哪些 Python 软件包，以及您已安装的软件包版本。

下表显示了可用笔记本镜像中使用的软件包版本：

表 5.1. 笔记本镜像选项

镜像名称	镜像版本	预安装的软件包
CUDA	2 (推荐)	<ul style="list-style-type: none"> ● Python 3.9 ● CUDA 11.8 ● JupyterLab 3.5 ● Notebook 6.5
	1	<ul style="list-style-type: none"> ● Python 3.8 ● CUDA 11.4 ● JupyterLab 3.2 ● Notebook 6.4
最小 Python (默认)	2 (推荐)	<ul style="list-style-type: none"> ● Python 3.9 ● JupyterLab 3.5 ● Notebook 6.5
	1	<ul style="list-style-type: none"> ● Python 3.8 ● JupyterLab 3.2 ● Notebook 6.4

镜像名称	镜像版本	预安装的软件包
PyTorch	2 (推荐)	<ul style="list-style-type: none">● Python 3.9● JupyterLab 3.5● Notebook 6.5● PyTorch 1.13● CUDA 11.7● TensorBoard 2.11● Boto3 1.26● Kafka-Python 2.0● Matplotlib 3.6● Numpy 1.24● Pandas 1.5● Scikit-learn 1.2● SciPy 1.10
	1	<ul style="list-style-type: none">● Python 3.8● JupyterLab 3.2● Notebook 6.4● PyTorch 1.8● CUDA 10.2● TensorBoard 2.6● Boto3 1.17● Kafka-Python 2.0● Matplotlib 3.4● Numpy 1.19● Pandas 1.2● Scikit-learn 0.24● SciPy 1.6

镜像名称	镜像版本	预安装的软件包
Standard Data Science	2 (推荐)	<ul style="list-style-type: none">● Python 3.9● JupyterLab 3.5● Notebook 6.5● Boto3 1.26● Kafka-Python 2.0● Matplotlib 3.6● Pandas 1.5● Numpy 1.24● Scikit-learn 1.2● SciPy 1.10
	1	<ul style="list-style-type: none">● Python 3.8● JupyterLab 3.2● Notebook 6.4● Boto3 1.17● Kafka-Python 2.0● Matplotlib 3.4● Pandas 1.2● Numpy 1.19● Scikit-learn 0.24● SciPy 1.6

镜像名称	镜像版本	预安装的软件包
TensorFlow	2 (推荐)	<ul style="list-style-type: none"> ● Python 3.9 ● JupyterLab 3.5 ● Notebook 6.5 ● TensorFlow 2.11 ● TensorBoard 2.11 ● CUDA 11.8 ● Boto3 1.26 ● Kafka-Python 2.0 ● Matplotlib 3.6 ● Numpy 1.24 ● Pandas 1.5 ● Scikit-learn 1.2 ● SciPy 1.10
	1	<ul style="list-style-type: none"> ● Python 3.8 ● JupyterLab 3.2 ● Notebook 6.4 ● TensorFlow 2.7 ● TensorBoard 2.6 ● CUDA 11.4 ● Boto3 1.17 ● Kafka-Python 2.0 ● Matplotlib 3.4 ● Numpy 1.19 ● Pandas 1.2 ● Scikit-learn 0.24 ● SciPy 1.6

镜像名称	镜像版本	预安装的软件包
TrustyAI	1	<ul style="list-style-type: none"> ● Python 3.9 ● JupyterLab 3.5 ● Notebook 6.5 ● TrustyAI 0.2 ● Boto3 1.26 ● Kafka-Python 2.0 ● Matplotlib 3.6 ● Numpy 1.24 ● Pandas 1.5 ● Scikit-learn 1.2 ● SciPy 1.10

部署大小

指定笔记本服务器上可用的计算资源。

容器大小 控制 CPU 数量、内存量、容器的最小和最多请求容量。

GPU 数 指定附加到容器的图形处理单元数。



重要

只有在 PyTorch、TensorFlow 和 CUDA 笔记本服务器镜像中才支持使用 GPU 来加快工作负载性能。另外，只有在集群中启用了 GPU 时，您可以指定笔记本服务器所需的 GPU 数量。要了解如何启用 GPU 支持，请参阅在 [OpenShift Data Science 中启用 GPU 支持](#)。

环境变量

指定在笔记本服务器上要设置的变量名称和值。在服务器启动期间设置环境变量意味着，您不需要在笔记本的正文或 Jupyter 命令行界面中定义它们。表中显示了一些推荐的环境变量。

表 5.2. 推荐的环境变量

环境变量选项	推荐的变量名称
AWS	<ul style="list-style-type: none"> ● AWS_ACCESS_KEY_ID 指定 Amazon Web Services 的 Access Key ID。 ● AWS_SECRET_ACCESS_KEY 指定在 AWS_ACCESS_KEY_ID 中指定的帐户的 Secret 访问密钥。

其他资源

- [启动 Jupyter 并启动笔记本服务器](#)

第 6 章 数据科学工作者教程

为了帮助您快速开始工作，您可以访问 Red Hat OpenShift Data Science 及其支持的应用程序的学习资源。这些资源可在 Red Hat OpenShift Data Science 用户界面的 **Resources** 选项卡中找到。

表 6.1. 教程

资源名称	描述
使用 Numba 在 Python 中加速科学工作负载	观看有关如何使您的 Python 代码运行更快的视频。
在 Python 中构建交互式可视化和仪表板	跨越多个笔记本浏览各种数据，了解如何部署完整仪表板和应用程序。
使用 scikit-learn 构建机器学习模型	了解如何利用 scikit-learn（高级学习、非监管学习和分类问题）构建机器学习模型。
构建二进制分类模型	培训模式以预测客户是否有可能订阅银行促销。
为数据可视化选择 Python 工具	使用 PyViz.org 网站以帮助您决定最适合您的开源 Python 数据可视化工具。
探索用于数据科学的 Anaconda	了解 Anaconda，这是 Python 和 R 编程语言的免费开源分发。
Pachyderm 概念入门	通过创建在几个镜像上执行边缘检测的管道，了解 Pachyderm 的主要概念。
带有 Numba 的 Python 中的 GPU 计算	了解如何使用 Numba 创建 GPU 加速功能。
运行 Python 笔记本在 IBM Watson OpenScale 中生成结果	运行 Python 笔记本电脑来创建、培训和部署机器学习模型。
运行 AutoAI 试验来构建模型	观看关于为营销活动构建二进制分类模型的视频。
在 Pachyderm 中培训回归模型	了解如何使用 Pachyderm 集群创建用于试验、分析数据并设置回归集群的示例数据存储库。
使用 Dask 进行并行数据分析	使用 Dask（一种可扩展现有 Python 生态系统的并行计算库）并行分析本地的中型数据集。
在 Watson Studio 中使用 Jupyter 笔记本	观看关于在 Watson Studio 中使用 Jupyter 笔记本的视频。
使用 Pandas 在 Python 中进行数据分析	了解如何使用 pandas（一个 Python 编程语言的数据分析库）。

表 6.2. 快速入门指南

资源名称	描述
创建 Jupyter 笔记本	在 JupyterLab 中创建 Jupyter 笔记本。
使用 NVIDIA GPU 附加组件创建机器学习模型	在 Jupyter 上创建机器学习模型，它使用您已提供的 GPU。
创建启用 Anaconda 的 Jupyterbook	创建一个启用 Anaconda 的 Jupyter 笔记本并访问为安全性和兼容性而策展的 Anaconda 软件包。
使用 Watson Studio 部署模型	在 Watson Studio 中导入笔记本，并使用 AutoAI 构建和部署模型。
使用 Flask 和 OpenShift 部署示例 Python 应用程序	在 Jupyterbook 和 Flask 应用程序中部署您的数据科学模型，以用作开发沙盒。
导入 Pachyderm Beginner 教程说明书	加载 Pachyderm 的初学教程教程，并了解 Pachyderm 的主概念，如数据存储库、管道以及使用单元格中的 pachctl CLI。
安装和验证 NVIDIA GPU 附加组件	了解如何安装并验证 Jupyter 检测到可用的 GPU 可用。
打开金丝雀部署启动和更新 SK 学习模型	打开 SK 学习模型并使用 Canary 部署实践进行更新。
使用 Starburst Galaxy 查询数据	学习如何使用 Starburst Galaxy 从 Jupyter 笔记本查询数据。
使用 Red Hat OpenShift API Management 保护部署模型	使用 Red Hat OpenShift API Management 保护模型服务 API。
使用 Intel® oneAPI AI Analytics Toolkit(AI Kit) 记书本	使用 Intel® oneAPI AI 分析工具包运行数据科学说明书示例。
使用 OpenVINO 工具包	使用 OpenVINO 模型优化器提供 ONNX 计算机愿景模型并使用笔记本中推断的结果。

表 6.3. How to 指南

资源名称	描述
如何在笔记本运行时环境选项之间进行选择	探索用于配置笔记本运行时环境的可用选项。
如何清理、形成和可视化数据	了解如何使用 IBM Watson Studio 数据优化清理和形成表格数据。
如何创建连接以访问数据	了解如何在平台中创建与各种数据源的连接。

资源名称	描述
如何创建部署空间	了解如何为机器学习创建部署空间。
如何在 Watson Studio 中创建笔记本	了解如何在 Watson Studio 中创建基本的 Jupyter 笔记本。
如何在 Watson Studio 中创建项目	了解如何在 Watson Studio 中创建分析项目。
如何创建与 Git 集成的项目	了解如何将 Git 存储库的资产添加到项目中。
如何在笔记本中安装 Python 软件包	了解如何在笔记本服务器中安装其他 Python 软件包。
如何将数据加载到 Jupyter 笔记本	了解如何通过加载数据将数据源整合到 Jupyter 笔记本中。
如何使用 OpenVINO Model Server 提供模型	了解如何使用 OpenVINO 自定义资源使用 OpenVINO 型号服务器部署优化模型。
如何设置 Watson OpenScale	了解如何通过 OpenScale 跟踪和衡量来自模型的成果。
如何更新笔记本服务器设置	了解如何在笔记本服务器上更新设置或笔记本镜像。
如何使用 Amazon S3 存储桶中的数据	了解如何使用环境变量在 S3 存储中连接数据。
如何在笔记本服务器上查看已安装的软件包	了解如何查看在运行的笔记本服务器上安装了哪些软件包。

6.1. 访问教程

您可以访问 Red Hat OpenShift Data Science 及其支持的应用程序的学习资源。

先决条件

- 确保您已登陆到 Red Hat OpenShift Data Science。
- 已登陆到 OpenShift Dedicated Web 控制台。

步骤

1. 在 Red Hat OpenShift Data Science 主页上，单击 **Resources**。
Resources 页面将打开。
2. 点相关卡上的 **Access tutorial**。

验证

- 您可以访问 Red Hat OpenShift Data Science 及其支持的应用程序的学习资源。

其他资源

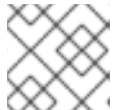
- [常见问题](#).

第 7 章 启用与 OPENSIFT DATA SCIENCE 连接的服务

您必须将基于 SaaS 的服务（如 Anaconda 专业版）与 Red Hat OpenShift Data Science 搭配使用之前，您必须启用这些基于 SaaS 的服务。On-cluster 服务会自动启用。

通常，您可以安装服务，或使用以下方法之一启用连接到 OpenShift Data Science 的服务：

- 从 OpenShift Data Science 仪表板上的 **Explore** 页面启用该服务，如以下步骤所述。
- 从 OperatorHub 为服务安装 Operator。OperatorHub 是一个 web 控制台，供集群管理员用来发现并选择要在其集群上安装的 Operator。它默认部署在 OpenShift Container Platform 中（[使用 Web 控制台从 OperatorHub 安装](#)）。



注意

红帽可能不会完全支持包含 OperatorHub 安装的 Operator 的部署。

- 从 Red Hat Marketplace 为服务安装 Operator ([安装 Operator](#))。
- 将服务作为附加组件安装到 OpenShift Dedicated 集群([将 Operator 添加到集群](#))。

对于一些服务（如 Jupyter），服务端点位于 OpenShift Data Science 的 **Enabled** 页面中的服务的标题中。某些服务无法直接从其标题访问，例如 OpenVINO 和 Anaconda 提供在 Jupyter 中使用的笔记本镜像，且不会从其标题中提供端点链接。此外，将这些端点 URL 存储为环境变量，以便在笔记本环境中轻松引用。

为了帮助您快速启动，您可以访问 **Resources** 页中的学习资源和文档，或者点击 **Enabled** 页面中服务的相关链接。

先决条件

- 您已登陆到 OpenShift Data Science。
- 您的管理员已安装了它，或在 OpenShift 集群上配置了该服务。

步骤

1. 在 OpenShift Data Science 主页，单击 **Explore**。
Explore 页面将打开。
2. 点击您要启用的服务卡。
3. 在服务的抽屉上点 **Enable**。
4. 如有提示，输入服务的密钥并点 **连接**。
5. 点 **Enable** 以确认您要启用该服务。

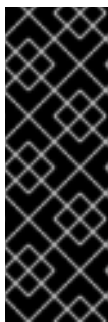
验证

- 您启用的服务出现在 **Enabled** 页面中。
- 服务端点显示在 **Enabled** 页面中的服务的标题中。

第 8 章 禁用连接到 OPENSIFT DATA SCIENCE 的应用程序

您可以禁用应用程序和组件，以便在不再使用它们时（例如，不再需要这些应用程序或应用程序许可证也过期），这些应用程序不再出现在 OpenShift Data Science 仪表板中。

禁用未使用的应用程序后，您的数据科学家可以从其 OpenShift Data Science 仪表板中手动删除这些应用程序卡，以便他们能够专注于最有可能使用的应用程序。如需有关手动删除应用程序卡的更多信息，请参阅[从 OpenShift Data Science 中删除禁用的应用程序](#)。



重要

禁用以下应用程序时不要遵循这个步骤：

- Anaconda 专业版.您不能手动禁用 Anaconda Professional 版本。只有在许可证过期时才会自动禁用。
- Red Hat OpenShift API Management。您只能从 OpenShift Cluster Manager 卸载 Red Hat OpenShift API Management。

- 已登陆到 OpenShift Dedicated Web 控制台。
- 您在 OpenShift Dedicated 中是 **cluster-admins** 用户组的一部分。
- 您已在 OpenShift Dedicated 集群上安装或配置了服务。
- 您要禁用的应用程序或组件被启用并在 **Enabled** 页面中显示。

步骤

1. 在 OpenShift Dedicated Web 控制台中，更改到 **Administrator** 视角。
2. 更改到 **redhat-ods-applications** 项目。
3. 点 **Operators** → **Installed Operators**。
4. 点您要卸载的 Operator。您可以在 **Filter by name** 字段中输入关键字，以帮助您更快地找到 Operator。
5. 使用 Operator 界面中的标签页删除任何 Operator 资源或实例。
在安装过程中，一些 Operator 要求管理员使用 Operator 界面中的标签页创建资源或启动进程实例。这些必须在 Operator 可以正确卸载前删除。
6. 在 **Operator Details** 页面中，点 **Actions** 下拉菜单并选择 **Uninstall Operator**。
此时会显示 **Uninstall Operator?** 对话框。
7. 选择 **Uninstall** 来卸载 Operator、Operator 部署和 pod。完成后，Operator 会停止运行，不再接收更新。



重要

删除 Operator 不会删除 Operator 的任何自定义资源定义或受管资源。自定义资源定义和受管资源仍然存在，必须手动清理。Operator 部署的任何应用程序以及配置的任何非集群资源都会继续运行，必须手动清理。

验证

- Operator 从其目标集群卸载。
- Operator 不再出现在 **Installed Operators** 页面中。
- 您的数据科学家将无法使用禁用的应用程序，并在 OpenShift Data Science 仪表板的 **Enabled** 页面中被标记为 **禁用**。此操作可能需要几分钟时间在删除 Operator 后进行。

8.1. 从 OPENSIFT DATA SCIENCE 中删除禁用的应用程序

在管理员禁用了未使用的应用程序后，您可以从 Red Hat OpenShift Data Science 仪表板中手动删除它们。禁用和删除未使用的应用程序后，您可以专注于最有可能使用的应用程序。

先决条件

- 确保您已登陆到 Red Hat OpenShift Data Science。
- 已登陆到 OpenShift Dedicated Web 控制台。
- 您的管理员之前已禁用了您要删除的应用程序。

步骤

1. 在 OpenShift Data Science 界面中，单击 **Enabled**。
Enabled 页面将打开。禁用的应用程序使用应用程序的卡中的 **Disabled** 表示。
2. 点您要删除的应用程序的卡中的 **Disabled**。
3. 点击链接删除应用程序卡。

验证

- 禁用应用程序的卡不再出现在 **Enabled** 页面中。

第 9 章 支持要求和限制

查阅本节以了解红帽支持的要求以及 Red Hat OpenShift Data Science 的限制。

9.1. 支持的浏览器

Red Hat OpenShift Data Science 支持以下浏览器的最新版本：

- Google Chrome
- Mozilla Firefox
- Safari

9.2. 支持的服务

Red Hat OpenShift Data Science 支持以下服务：

表 9.1. 支持的服务

服务名称	描述
Anaconda 专业版	Anaconda 专业版是一个流行的开源软件包分布和管理体验，针对商业用途进行了优化。
IBM Watson Studio	IBM Watson Studio 是一个可以将 AI 和机器学习集成到您的业务中的平台，并使用您自己的数据创建自定义模型。
Intel® oneAPI AI Analytics Toolkits	AI Kit 是一组 AI 软件工具，可加速 Intel® 架构上的端到端数据科学和分析管道。
Jupyter	<p>Jupyter 是笔记本的多用户版本，专为公司、课堂和研究实验室而设计。</p> <div style="display: flex; align-items: flex-start;"> <div style="width: 30px; height: 100px; background: repeating-linear-gradient(45deg, transparent, transparent 2px, black 2px, black 4px); margin-right: 10px;"></div> <div> <p>重要</p> <p>虽然我们会努力确保 Red Hat OpenShift Data Science 可以应对 OpenShift 节点故障、升级和类似的具有破坏性的操作，但个人用户的笔记本环境在出现这些情况时有可能会中断。如果 OpenShift 节点重启或不可用，则该节点上任何用户笔记本环境都会在不同节点上重启。当发生这种情况时，在用户笔记本环境中执行的任何持续进程都会中断，用户需要在其环境重新可用时重新执行这些进程。</p> <p>由于这个限制，红帽建议在 OpenShift Data Science 上的 Jupyter 笔记本服务器环境中执行那些不允许中断的进程。</p> </div> </div>

服务名称	描述
Pachyderm	<p>利用 Pachyderm 的数据版本、管道和线板功能自动化机器学习生命周期并优化机器学习操作。</p> <div style="display: flex; align-items: flex-start;">  <div> <p>注意</p> <p>Pachyderm 创建的 pachd pod 不会从 OpenShift Web 控制台提供终端访问。如果您尝试访问 pachd pod 的终端视图，您会看到错误。这是预期的行为。</p> </div> </div>
Red Hat OpenShift API Management	OpenShift API 管理是一种服务，可加快时间价值，并降低了提供基于 API 优先的微服务应用的成本。
OpenVINO	OpenVINO 是一个开源工具包，可帮助优化深度学习性能，并使用对 Intel 硬件进行推测引擎进行部署。
Starburst Galaxy	Starburst Galaxy 是一个完全被管理的服务，可使用 SQL 在不同数据源中运行高性能查询。

9.3. 支持的软件包

Red Hat OpenShift Data Science 中最新支持的笔记本服务器镜像默认安装有 Python。有关这些镜像中包含的软件包和版本的完整列表，请查看[笔记本服务器环境的选项](#)中的表。

您可以在任何包含该软件包所需二进制文件的笔记本服务器上安装与支持的 Python 版本兼容的软件包。如果您想要使用的笔记本服务器镜像中不包含所需的二进制文件，请联络红帽支持。

您可以使用 **pip install** 命令临时安装软件包。您还可以使用 **requirements.txt** 文件向 **pip install** 命令提供软件包列表。如需更多信息，请参阅[在笔记本服务器中安装 Python 软件包](#)。

每次启动笔记本服务器时，都必须重新安装这些软件包。

您可以使用 **pip uninstall** 命令删除软件包。

其他资源

- [在笔记本服务器中安装 Python 软件包](#)
- [笔记本服务器环境的选项](#)

第 10 章 常见问题

除了文档外，红帽还为 OpenShift Data Science 和支持的应用程序提供一组丰富的学习资源。

在 OpenShift Data Science 仪表板的 **Resources** 页面中，您可以使用 `category` 链接来过滤数据科学工作流的不同阶段的资源。例如，点 **Model serving** 类别来显示各种部署模型方法的资源。点 **All items** 显示所有类别的资源。

对于所选类别，您可以应用附加选项来过滤可用资源。例如，您可以根据类型进行过滤，如 `How-to article, quick start, tutorials`；这些资源提供了常见问题的回答。

第 11 章 管理员对 JUPYTER 中的常见问题进行故障排除

如果您的用户在与 Jupyter、其笔记本或他们的笔记本服务器相关的 Red Hat OpenShift 数据存储中遇到错误，请阅读本节以了解导致问题的原因，以及如何解决这个问题。

如果在此处或发行注记中无法找到相关的信息，请联系红帽支持团队。

11.1. 用户收到 404 : PAGE NOT FOUND 错误（在登录到 JUPYTER

问题

如果您配置了 OpenShift Data Science 用户组，则可能无法将用户名添加到 OpenShift Data Science 的默认用户组。

诊断

检查用户是否是默认用户组的一部分。

1. 查找允许访问 Jupyter 的组名称。
 - a. 登录到 OpenShift Dedicated Web 控制台。
 - a. 点 **User Management → Groups**。
 - b. 点用户组的名称，如 **rhods-users**。
此时会出现该组的组详细信息页面。
 1. 点组的 **Details** 选项卡，并确认相关组的 **Users** 部分包含有权访问 Jupyter 的用户。

解决方案

- 如果用户没有被添加到允许访问 Jupyter 的组中，请遵循 [OpenShift Data Science 添加用户](#) 来添加它们。
- 如果用户已添加到允许访问 Jupyter 的组中，请联系红帽支持。

11.2. 用户的笔记本服务器没有启动

托管用户笔记本服务器的 OpenShift Dedicated 集群可能无法访问充足的资源，或者 Jupyter pod 可能出现问题。

1. 登录到 OpenShift Dedicated Web 控制台。
1. 删除并重启此用户的笔记本服务器 Pod。
 - a. 点 **Workloads → Pods**，将项目设置为 **rhods-notebooks**。
 - b. 搜索属于此用户的笔记本服务器 pod，例如 **jupyter-nb-<username>-***。
如果笔记本服务器 pod 存在，则笔记本服务器 pod 中可能会出现间歇性失败。

如果用户的笔记本服务器 pod 不存在，请继续诊断。
2. 根据所选笔记本服务器镜像所需的资源，检查 OpenShift 集群中当前可用的资源。
如果有足够 CPU 和 RAM 的 worker 节点可用于在集群中调度，请继续诊断过程。
3. 检查 Jupyter pod 的状态。

解决方案

- 如果笔记本服务器 pod 出现间歇性失败：
 - a. 删除属于用户的笔记本服务器 pod。
 - b. 询问用户再次启动其笔记本服务器。
- 如果笔记本服务器没有足够的资源来运行所选笔记本电脑服务器镜像，请在 OpenShift 集群中添加更多资源，或者选择较小的镜像大小。
- 如果 Jupyter pod 处于 FAILED 状态：
 - a. 检索 `jupyter-nb-*` pod 的日志，并将其发送到红帽支持以进行进一步评估。
 - b. 删除 `jupyter-nb-*` pod。
- 如果没有以前的解决方案，请联系红帽支持。

11.3. 用户运行笔记本的 CELLS 时遇到 DATABASE OR DISK IS FULL 错误或 NO SPACE LEFT ON DEVICE 错误。

问题

用户可能已在其笔记本服务器上耗尽存储空间。

诊断

1. 登录到 Jupyter，并启动属于用户问题的笔记本服务器。如果笔记本服务器没有启动，请按照以下步骤检查用户是否已耗尽存储空间：
 - a. 登录到 OpenShift Dedicated Web 控制台。
 - a. 点 **Workloads** → **Pods**，将项目设置为 `rhods-notebooks`。
 - b. 点属于该用户的笔记本服务器 pod，例如 `jupyter-nb-<idp>-<username>-*`。
 - c. 点 **Logs**。如果您看到类似如下的行，用户已超过其可用容量：

```
Unexpected error while saving file: XXXX database or disk is full
```

解决方案

- 通过扩展其持久性卷来增加用户可用的存储：[扩展持久性卷](#)
- 与用户合作找出可以从 `/opt/app-root/src` 目录中删除的文件，以释放其现有存储空间。

第 12 章 对 JUPYTER 中的常见问题进行故障排除

如果您在与 Jupyter、笔记本电脑或笔记本服务器相关的 Red Hat OpenShift Data Science 中看到错误，请阅读本节以了解可能导致此问题的内容。

如果您无法在此处或发行注记中看到您的问题，请联系红帽支持。

12.1. 我可以看到 403:FORBIDDEN 错误（当登录到 JUPYTER 时）

问题

如果您的管理员配置了专用的 OpenShift Data Science 用户组，您的用户名可能不会添加到默认用户组或 OpenShift Data Science 的默认管理员组中。

解决方案

Contact your administrator so that they can add you to the correct group/s.

12.2. 我的笔记本服务器没有启动

托管您的笔记本服务器的 OpenShift Dedicated 集群可能无法访问充足的资源，或者 Jupyter pod 可能出现问题。

解决方案

检查 OpenShift 中的 **Events** 部分中的日志，以了解与问题相关的错误消息。例如：

```
Server requested
2021-10-28T13:31:29.830991Z [Warning] 0/7 nodes are available: 2 Insufficient memory,
2 node(s) had taint {node-role.kubernetes.io/infra: }, that the pod didn't tolerate, 3 node(s) had taint
{node-role.kubernetes.io/master: },
that the pod didn't tolerate.
```

请联系您的管理员，获取任何相关错误消息的详细信息，以便他们可以执行进一步检查。

12.3. 在运行我的笔记本时遇到一个 DATABASE OR DISK IS FULL 错误或一个 NO SPACE LEFT ON DEVICE 错误

问题

您可能在笔记本服务器上耗尽存储空间。

解决方案

请联系您的管理员，以便他们可以进一步的检查。