



Red Hat OpenShift AI Cloud Service 1

使用数据科学项目

在项目和工作台中组织您的工作，创建并协作笔记本、培训和部署模型、配置模型服务器和实施管道

Red Hat OpenShift AI Cloud Service 1 使用数据科学项目

在项目和工作中组织您的工作，创建并协作笔记本、培训和部署模型、配置模型服务器和实施管道

法律通告

Copyright © 2024 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

将工作组织到项目和工作台中，创建并协作笔记本、培训和部署模型、配置模型服务器和实施管道。

目录

| | |
|----------------------------------|-----------|
| 第 1 章 使用数据科学项目 | 3 |
| 1.1. 使用数据科学项目 | 3 |
| 1.2. 使用项目工作台 | 5 |
| 1.3. 使用数据连接 | 13 |
| 1.4. 配置集群存储 | 16 |
| 1.5. 配置对数据科学项目的访问 | 18 |
| 第 2 章 在数据科学 IDE 中使用 | 22 |
| 2.1. 访问工作台 IDE | 22 |
| 2.2. 在 JUPYTERLAB 中工作 | 22 |
| 2.3. 在 CODE-SERVER 中工作 | 29 |
| 第 3 章 使用数据科学项目管道 | 31 |
| 3.1. 启用数据科学管道 2.0 | 31 |
| 3.2. 管理数据科学项目 | 33 |
| 3.3. 管理管道运行 | 49 |
| 3.4. 使用管道日志 | 69 |
| 3.5. 在 JUPYTERLAB 中使用管道 | 73 |
| 3.6. 其他资源 | 87 |

第 1 章 使用数据科学项目

作为数据科学家，您可以将数据科学工作组织到一个项目中。OpenShift AI 中的数据科学项目可由以下组件组成：

工作台

通过创建工作台，您可以在项目中添加 Jupyter 笔记本。

集群存储

对于需要保留数据的数据科学项目，您可以在项目中添加集群存储。

数据连接

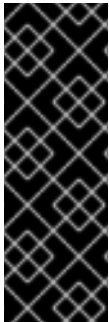
通过向项目添加数据连接，您可以将数据输入连接到工作台。

Pipelines

通过标准化和自动化机器学习工作流，您可以进一步增强和部署数据科学模型。

模型和模型服务器

部署经过培训的数据科学模型，为智能应用程序提供服务。您的模型使用端点部署，允许应用程序向模型发送请求。



重要

如果您在 OpenShift AI 用户界面外创建 OpenShift 项目，则项目不会在 **Data Science Projects** 页面中显示。另外，您无法使用 OpenShift AI 专用的功能，如工作台和模型服务，以及标准 OpenShift 项目。

要将 OpenShift 项目分为数据科学项目，并且为 OpenShift AI 提供专用功能，您必须将标签 `opendatahub.io/dashboard: 'true'` 添加到项目命名空间中。添加此标签后，项目随后会在 **Data Science Projects** 页面中显示。

1.1. 使用数据科学项目

1.1.1. 创建数据科学项目

要实施数据科学工作流，您必须创建一个项目。在 OpenShift 中，项目是具有额外注解的 Kubernetes 命名空间，是您可以管理用户对资源访问的主要方法。个项目将您的数据科学工作组织在一个位置，并允许您与机构中的其他开发人员和数据科学家合作。

在一个项目中，您可以添加以下功能：

- 数据连接，以便您可以访问数据，而无需硬编码信息，如端点或凭证。
- 使用和处理数据的工作台以及开发模型。
- 部署模型，以便您可以对其进行测试，然后将其集成到智能应用程序中。部署模型使其可作为服务使用 API 访问。
- 用于自动化 ML 工作流的管道。

先决条件

- 您已登陆到 Red Hat OpenShift AI。

- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh **oai-users** 或 **rhoai-admins**）。

流程

1. 在 OpenShift AI 仪表板中，选择 **Data Science Projects**。
2. 点 **Create data Science** 项目。
3. 在 **Create a data Science project** 对话框中，输入项目的显示名称。
4. 可选：编辑数据科学项目的资源名称。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。
注：在创建项目后，您可以更改项目显示名称，但您无法更改资源名称。
5. 输入您的数据科学项目的描述。
6. 点 **Create**。

验证

- 此时会打开项目详情页面。在这个页面中，您可以添加数据连接、创建工作台、配置管道和部署模型。

1.1.2. 更新数据科学项目

您可以通过更改项目的名称和描述文本来更新数据科学项目的详细信息。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh **oai-users** 或 **rhoai-admins**）。
- 您已创建了数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 **Data Science Projects**。
Data Science Projects 页面将打开。
2. 点您要更新详情的项目旁的操作菜单 (⋮)，然后点 **Edit project**。
此时会打开 **Edit data Science** 项目对话框。
3. 可选：更新数据科学项目的名称。
4. 可选：更新您的数据科学项目的描述。
5. 点 **Update**。

验证

- 您更新的数据科学项目显示在 **Data Science Projects** 页面中。

1.1.3. 删除数据科学项目

您可以删除数据科学项目，以便在不再使用它们时不会在 OpenShift AI Data Science Projects 页面中显示它们。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，{oai-user-group}）。
- 您已创建了数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要删除的项目旁边的操作菜单(ProductShortName)，然后点 Delete project。Delete project 对话框将打开。
3. 在文本字段中输入项目的名称，以确认您想要删除它。
4. 点 Delete project。

验证

- 您删除的数据科学项目不再显示在 Data Science Projects 页面中。
- 删除数据科学项目会删除任何关联的工作台、数据科学项目、集群存储和数据连接。这个数据会被永久删除，且无法恢复。

1.2. 使用项目工作台

1.2.1. 创建工作台并选择 IDE

工作台是一种隔离区域，您可以检查并使用 ML 模型。您还可以使用数据和运行程序，例如准备和清理数据。例如，如果工作台不需要现有模式，但大多数数据科学工作流程任务（如编写代码处理数据或培训模型）都需要一个工作台。

当您创建工作台时，您可以指定一个镜像(IDE、软件包和其他依赖项)。支持的 IDE 包括 JupyterLab、code-server（技术预览）和 RStudio（技术预览）。

IDE 基于服务器客户端架构。每个 IDE 都提供一个在 OpenShift 集群上的容器中运行的服务器，而用户界面（客户端）则显示在 Web 浏览器中。例如，Jupyter 笔记本服务器在 Red Hat OpenShift 集群的容器中运行。客户端是在您的本地计算机上的 Web 浏览器中打开的 JupyterLab 界面。您在 JupyterLab 中输入的所有命令都由笔记本服务器执行。同样，code-server 或 RStudio 服务器等其他 IDE 提供了一种在 OpenShift 集群上的容器中运行的服务器，而用户界面显示在 Web 浏览器中。此架构允许您在浏览器环境中通过本地计算机交互，同时在集群中进行所有处理。集群提供更大的可用资源和安全性优势，因为正在处理的数据永远不会离开集群。

在工作台中，您还可以配置数据连接（以访问培训模型的外部数据，并保存模型，以便您可以部署它们）和集群存储（用于保留数据）。同一项目中的工作台可以通过带有数据科学管道和模型服务器的对象存储共享模型和数据。

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。

在一个项目中，您可以创建多个工作台。当创建新工作台时，需要考虑事项，例如：

- 工作台配置（如 CPU、RAM 或 IDE）。如果要避免将现有工作台配置编辑为一个新的任务，您可以创建一个新的工作台。
- 分离任务或活动。例如，您可能希望在大型语言模型(LLM)试验活动中使用一个工作台、专用于演示的另一个工作台，另一个工作台用于测试。

1.2.1.1. 关于工作台镜像

工作台镜像（有时称为笔记本镜像）使用模型开发所需的工具和库优化。您可以使用提供的工作台镜像或 OpenShift AI admin 用户创建符合您需求的自定义工作台镜像。

为了为您的模型开发提供一致的稳定平台，许多提供的工作台镜像都包含相同的 Python 版本。OpenShift AI 上的大多数工作台镜像已预先构建，并在安装或升级 OpenShift AI 立即使用。

有关红帽支持工作台镜像和软件包的详情，请参考 [Red Hat OpenShift AI: 支持的配置](#)。

Red Hat OpenShift AI 包含以下默认可用的笔记本镜像。



重要

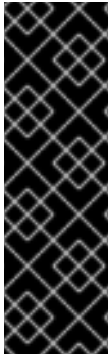
此表中带有（技术预览）的笔记本镜像不被红帽产品服务等级协议(SLA)支持，且可能无法正常工作。红帽不推荐在生产环境中使用技术预览功能。这些技术预览功能可以使用户提早试用新的功能，并有机会在开发阶段提供反馈意见。有关红帽技术预览功能支持范围的更多信息，请参阅[技术预览功能支持范围](#)。

表 1.1. 默认笔记本镜像

| 镜像名称 | 描述 |
|-----------------------|--|
| CUDA | 如果您正在与需要 GPU 支持的计算密集型数据科学模型合作，请使用 Compute Unified Device Architecture(CUDA)笔记本镜像来访问 NVIDIA CUDA 工具包。使用这个工具包，您可以使用 GPU 加速的库和优化工具优化您的工作。 |
| Standard Data Science | 对于不需要 TensorFlow 或 PyTorch 的型号，请使用标准数据科学书镜像。此镜像包含常用的库，可帮助您开发机器学习模型。 |
| TensorFlow | TensorFlow 是供机器学习的开源平台。通过 TensorFlow，您可以构建、培训和部署机器学习模式。TensorFlow 包含高级数据可视化功能，如计算图形可视化。它还允许您轻松监控和跟踪模型的进度。 |
| PyTorch | PyTorch 是一个开源机器学习库，专为深入学习而优化。如果您正在操作计算机愿景或自然语言处理模型，请使用 Pytorch 笔记本电脑镜像。 |
| Minimal Python | 如果您不需要高级机器学习功能，或计算密集型数据科学工作的其他资源，您可以使用最小 Python 镜像开发您的模型。 |
| TrustyAI | 使用 TrustyAI 笔记本镜像通过模型解释性、追踪和责任以及运行时监控利用您的数据科学工作。 |

| 镜像名称 | 描述 |
|---------------------------------|---|
| HabanaAI | HabanaAI 笔记本镜像使用 Habana Gaudi 设备优化了高性能深度学习(DL)。Habana Gaudi 设备加速了 DL 培训工作负载，并最大化培训吞吐量和效率。 |
| code-server (技术预览) | <p>使用 code-server 笔记本镜像，您可以自定义笔记本环境，以满足使用各种扩展来添加新语言、主题、调试器和连接到其他服务的需求。通过语法突出显示、自动和括号匹配，提高数据科学工作的效率，以及用于无缝自动化的自动任务运行程序。如需更多信息，请参阅 GitHub 中的 code-server。</p> <p> 注意</p> <p>基于 Elyra 的管道在 code-server 笔记本镜像中不可用。</p> |
| RStudio 服务器 (技术预览) | <p>使用 RStudio 服务器笔记本镜像访问 RStudio IDE，这是统计计算和图形的集成开发环境。如需更多信息，请参阅 RStudio 服务器网站。</p> <p>要使用 RStudio 服务器 笔记本镜像，您必须首先通过创建机密并触发 BuildConfig 来构建它，然后通过编辑 rstudio-rhel9 镜像流在 OpenShift AI UI 中启用它。如需更多信息，请参阅 构建 RStudio 服务器工作台镜像。</p> <p> 重要</p> <p>免责声明： 红帽支持在 OpenShift AI 中管理工作台。但是，红帽不提供对 RStudio 软件的支持。RStudio 服务器可以通过 https://rstudio.org/ 访问，并遵循 RStudio 许可条款。使用此示例工作台前，请查看许可证条款。</p> |
| CUDA - RStudio 服务器 (技术预览) | <p>使用 CUDA - RStudio 服务器笔记本镜像访问 RStudio IDE 和 NVIDIA CUDA 工具包。RStudio 是 R 的集成开发环境，这是统计计算和图形的编程语言。使用 NVIDIA CUDA 工具包，您可以使用 GPU 加速的库和优化工具优化工作。如需更多信息，请参阅 RStudio 服务器网站。</p> <p>要使用 CUDA - RStudio Server 笔记本镜像，您必须首先通过创建 secret 并触发 BuildConfig 来构建它，然后在 OpenShift AI UI 中通过编辑 cuda-rstudio-rhel9 镜像流启用它。如需更多信息，请参阅 构建 RStudio 服务器工作台镜像。</p> <p> 重要</p> <p>免责声明： 红帽支持在 OpenShift AI 中管理工作台。但是，红帽不提供对 RStudio 软件的支持。RStudio 服务器可以通过 https://rstudio.org/ 访问，并遵循 RStudio 许可条款。使用此示例工作台前，请查看许可证条款。</p> <p>CUDA - RStudio 服务器 笔记本镜像包含 NVIDIA CUDA 技术。CUDA 许可信息位于 https://docs.nvidia.com/cuda/。使用此示例工作台前，请查看许可证条款。</p> |

1.2.1.2. 构建 RStudio 服务器工作台镜像



重要

RStudio Server 和 CUDA - RStudio Serverworkbench 镜像目前在 Red Hat OpenShift AI 中作为技术预览功能提供。

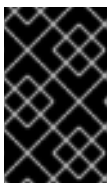
技术预览功能不受红帽产品服务等级协议 (SLA) 支持，且功能可能并不完整。红帽不推荐在生产环境中使用它们。这些技术预览功能可以使用户提早试用新的功能，并有机会在开发阶段提供反馈意见。

有关红帽技术预览功能支持范围的更多信息，请参阅[技术预览功能支持范围](#)。

Red Hat OpenShift AI 包括以下 RStudio Server workbench 镜像：

- RStudio Server workbench 镜像

使用 RStudio Server workbench 镜像，您可以访问 RStudio IDE，这是 R 编程语言的集成开发环境。R 用于统计计算和图形来支持数据分析和预测。

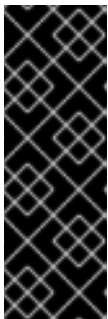


重要

免责声明：红帽支持在 OpenShift AI 中管理工作台。但是，红帽不提供对 RStudio 软件的支持。RStudio 服务器可以通过 [r Studio.org](https://rstudio.org) 提供，并遵循其许可条款。在使用此示例工作台前，您应该查看其许可条款。

- CUDA - RStudio Server workbench 镜像

使用 CUDA - RStudio Serverworkbench 镜像，您可以访问 RStudio IDE 和 NVIDIA CUDA Toolkit。RStudio IDE 是用于统计计算和图形的 R 编程语言的集成开发环境。使用 NVIDIA CUDA 工具包，您可以使用 GPU 加速的库和优化工具来增强您的工作。



重要

免责声明：红帽支持在 OpenShift AI 中管理工作台。但是，红帽不提供对 RStudio 软件的支持。RStudio 服务器可以通过 [r Studio.org](https://rstudio.org) 提供，并遵循其许可条款。在使用此示例工作台前，您应该查看其许可条款。

CUDA - RStudio Serverworkbench 镜像包含 NVIDIA CUDA 技术。CUDA 许可信息包括在 [CUDA Toolkit 文档](#) 中。在使用此示例工作台前，您应该查看其许可条款。

要使用 RStudio Server and CUDA - RStudio Serverworkbench 镜像，您必须首先通过创建 secret 并触发 BuildConfig 来在 OpenShift AI UI 中启用它们，然后通过编辑 `r Studio-rhel9` 和 `cuda-r Studio-rhel9` 镜像流在 OpenShift AI UI 中启用。

先决条件

- 在启动 RStudio 服务器构建过程前，您至少需要 1 个 CPU 和 2Gi 内存 (`r Studio-server-rhel9`)，以及 1.5 个 CPU 和 8Gi 内存用于集群中 `cuda-r Studio-server-rhel9`。
- 已登录到 OpenShift 集群。
- 在 OpenShift 中具有 `cluster-admin` 角色。
- 您有一个有效的 Red Hat Enterprise Linux (RHEL) 订阅。

流程

1. 使用 Subscription Manager 凭证创建 secret。这些通常是您的红帽客户门户网站用户名和密码。
注：secret 必须命名为 `rhel-subscription-secret`，其 `USERNAME` 和 `PASSWORD` 键必须采用大写字母。

```
oc create secret generic rhel-subscription-secret --from-literal=USERNAME=<username> --from-literal=PASSWORD=<password> -n redhat-ods-applications
```

2. 启动构建：

- a. 启动轻量级 RStudio 服务器构建：

```
oc start-build rstudio-server-rhel9 -n redhat-ods-applications --follow
```

- b. 要启动启用了 CUDA 的 RStudio 服务器构建，请触发 `cuda-rhel9 BuildConfig`：

```
oc start-build cuda-rhel9 -n redhat-ods-applications --follow
```

`cuda-rhel9` 构建是 `cuda-r Studio-rhel9` 的先决条件。`cuda-r Studio-rhel9` 构建会自动启动。

3. 使用以下命令确认构建过程成功完成。成功构建显示为 **Complete**。

```
oc get builds -n redhat-ods-applications
```

4. 构建成功完成后，使用以下命令使工作台镜像在 OpenShift AI UI 中提供。

- a. 启用 RStudio 服务器工作台镜像：

```
oc label -n redhat-ods-applications imagestream rstudio-rhel9 opendatahub.io/notebook-image='true'
```

- b. 启用 CUDA - RStudio Server workbench 镜像：

```
oc label -n redhat-ods-applications imagestream cuda-rstudio-rhel9 opendatahub.io/notebook-image='true'
```

验证

- 您可以在 Red Hat OpenShift AI 仪表板的 Applications → Enabled 菜单中看到 RStudio Server Server and CUDA - RStudio Server 镜像。
- 您可以在 Data Science Projects → Workbenches → Create workbench → Notebook image → Image selection 下拉列表中看到 R Studio Server 或 CUDA - RStudio Server。

1.2.1.3. 创建工作台

当您创建工作台时，您可以指定一个镜像(IDE、软件包和其他依赖项)。您还可以配置数据连接、集群存储并添加容器存储。

先决条件

- 您已登录到 Red Hat OpenShift AI。

- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您创建了一个项目。
- 如果您在 Red Hat OpenShift AI 之外创建了 Simple Storage Service (S3) 帐户，并希望创建到现有 S3 存储桶的数据连接，则具有存储存储桶的以下凭证信息：
 - 端点 URL
 - access key
 - Secret 密钥
 - 区域
 - bucket 名称

有关使用存储在 AWS S3 中的数据的详情，请参考从 [Amazon S3 集成数据](#)。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。
2. 点您要将工作台添加到的项目的名称。
此时会打开项目的 Details 页面。
3. 在 Workbenches 部分中，单击 Create a workbench。
4. 在 Create workbench 页面中，配置您要创建的工作台的属性。
 - a. 在 Name 字段中输入工作台的名称。
 - b. 可选：在 Description 字段中输入用于定义工作台的描述。
 - c. 在 Notebook image 部分，完成字段以指定要与工作台搭配使用的工作台镜像。
从 Image selection 列表中，选择一个适合您用例的工作台镜像。工作台镜像包含 IDE 和 Python 软件包（可重复利用的代码）。（可选）点 View package information 选项查看您选择的镜像中包含的软件包列表。

如果工作台镜像有多个可用的版本，请从 Versions 部分中选择要使用的工作台镜像版本。要使用最新的软件包版本，红帽建议您使用最近添加的镜像。



注意

在创建工作台后，您可以更改工作台镜像。

- d. 在 Deployment size 部分中，从 Container size 列表中选择服务器的容器大小。容器大小控制 CPU 数量、内存量以及容器的最小和最大请求容量。
- e. 可选：选择并指定任何环境变量的值。
在工作台配置期间设置环境变量可帮助您稍后节省时间，因为您不需要在笔记本的正文或使用 IDE 命令行界面定义它们。

如果使用 S3 兼容存储，请添加这些推荐的环境变量：

- **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。

如果在添加变量时选择 Secret，OpenShift AI 将凭证作为 Kubernetes secret 存储在受保护的命名空间中。

f. 为您的工作台配置存储。选择以下选项之一：

- 创建新的持久性存储 来创建在关闭工作台后保留的存储。完成相关字段以定义存储。
- 使用现有的持久性存储 来重复使用现有存储并从 Persistent storage 列表中选择存储。

g. 另外，您还可以在工作台中添加数据连接。数据连接是一个资源，其中包含连接数据源或对象存储桶所需的配置参数。目前，只支持 S3-Compatible 数据连接。您可以使用存储存储桶来存储数据、型号和管道工件。您还可以使用数据连接来指定您要部署的模型的位置。在 Data connections 部分中，选择 Use a data connection 复选框。

- 创建新数据连接，如下所示：
 - 选择 Create new data connection
 - 在 Name 字段中输入数据连接的唯一名称。
 - 在 Access key 字段中，输入 S3 兼容对象存储供应商的访问密钥 ID。
 - 在 Secret key 字段中，为您指定的 S3 兼容对象存储帐户输入 secret 访问密钥。
 - 在 Endpoint 字段中，输入 S3 兼容对象存储桶的端点。
 - 在 Region 字段中，输入 S3 兼容对象存储帐户的默认区域。
 - 在 Bucket 字段中，输入 S3 兼容对象存储桶的名称。
- 使用现有数据连接，如下所示：
 - 选择 Use existing data connection
 - 在 Data connection 列表中选择您之前定义的数据连接。

5. 点 Create workbench。

验证

- 您创建的工作台会出现在项目的 Workbenches 选项卡中。
- 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 Cluster storage 选项卡中。
- Workbenches 选项卡上的 Status 列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为 Running。
- （可选）点击 Open 链接在新窗口中打开 IDE。

1.2.2. 启动工作台

您可以从项目详情页面的 Workbenches 选项卡手动启动数据科学项目的工作台。默认情况下，工作台会在创建后马上启动。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了一个包含工作台的数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要启动其工作台的项目的名称。此时会打开项目详情页面。
3. 单击 Workbenches 选项卡。
4. 点相关工作台的 Status 列中的切换，以启动未运行的工作台。您开始的工作台的状态从 Stopped 更改为 Running。工作台启动后，点 Open 以打开工作台的笔记本。

验证

- 您启动的工作台会出现在项目的 Workbenches 选项卡中，状态为 Running。

1.2.3. 更新项目工作台

如果您的数据科学工作需要您更改工作台的笔记本镜像、容器大小或识别信息，您可以更新项目工作台的属性。如果您需要额外的电源以用于大型数据集，您可以将加速器分配给工作台来优化性能。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了一个具有工作台的数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要更新其工作台的项目的名称。此时会打开项目详情页面。
3. 单击 Workbenches 选项卡。
4. 点击您要更新的工作台旁的操作菜单(HBAC)，然后点 Edit workbench。

这会打开 Edit workbench 页面。

5. 更新任何工作台属性，然后点 Update workbench。

验证

- 您更新的工作台会出现在项目的 Workbenches 选项卡中。

1.2.4. 从数据科学项目中删除工作台

您可以从数据科学项目中删除工作台，以帮助删除不再与您的工作相关的 Jupyter 笔记本。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了一个具有工作台的数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要从中删除工作台的项目的名称。此时会打开项目详情页面。
3. 单击 Workbenches 选项卡。
4. 点您要删除的工作台旁的操作菜单（需要），然后点 Delete workbench。Delete workbench 对话框将打开。
5. 在文本字段中输入工作台名称，以确认您想要删除它。
6. 点删除工作台。

验证

- 您删除的工作台不再显示在项目的 Workbenches 选项卡中。
- 与工作台 of Jupyter 笔记本关联的自定义资源 (CR) 已删除。

1.3. 使用数据连接

1.3.1. 在您的数据科学项目中添加数据连接

您可以通过添加与数据源的连接来提高数据科学项目。当您想要使用非常大的数据集时，您可以将数据存储在兼容 S3 的对象存储桶中，以便不会填满本地存储。您还可以选择将数据连接与尚未连接的现有工作台关联。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了数据科学项目，您可向其添加数据连接。
- 您可以访问 S3 兼容对象存储。
- 如果要向现有的工作台添加数据连接，您已在工作台中保存任何数据，以避免丢失工作。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要添加数据连接的项目名称。此时会打开项目详情页面。
3. 点 Data connections 选项卡。
4. 点 Add data connection。此时会打开 Add data connection 对话框。
5. 输入数据连接的名称。
6. 在 Access key 字段中，输入 S3 兼容对象存储供应商的访问密钥 ID。
7. 在 Secret key 字段中，为您指定的 S3 兼容对象存储帐户输入 secret 访问密钥。
8. 在 Endpoint 字段中，输入 S3 兼容对象存储桶的端点。
9. 在 Region 字段中，输入 S3 兼容对象存储帐户的默认区域。
10. 在 Bucket 字段中，输入 S3 兼容对象存储桶的名称。
11. 可选：在 Connected workbench 列表中，选择一个要连接的工作台。
12. 点 Add data connection。

验证

- 您添加的数据连接会出现在项目的 Data connections 选项卡中。
- 如果您选择了一个工作台，则工作台会在项目的 Data connections 选项卡中的 Connected workbenches 列中可见。

1.3.2. 删除数据连接

您可以从数据科学项目中删除数据连接，以帮助删除不再与工作相关的连接。

先决条件

- 您已登陆到 Red Hat OpenShift AI。

- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了带有数据连接的数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要从中删除数据连接的项目名称。此时会打开项目详情页面。
3. 点 Data connections 选项卡。
4. 点您要删除的数据连接旁的操作菜单(HBAC)，然后点 Delete data connection。Delete data connection 对话框将打开。
5. 在文本字段中输入数据连接的名称，以确认您想要删除它。
6. 点 Delete data connection。

验证

- 您删除的数据连接不再显示在项目的 Data connections 选项卡中。

1.3.3. 更新连接的数据源

要使用具有不同工作台的现有数据源，您可以更改连接到项目的 workbench 的数据源。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了数据科学项目，创建了一个工作台，并定义了数据连接。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要更改其数据源的项目的名称。此时会打开项目详情页面。
3. 点 Data connections 选项卡。
4. 点击您要更改的数据源旁的操作菜单（需要修改），然后点 Edit data connection。这会打开 Edit data connection 对话框。
5. 在 Connected workbench 部分中，从列表中选择现有的工作台。
6. 点 Update data connection。

验证

- 更新的数据连接显示在项目的 Data connections 选项卡中。
- 您可以使用连接的 workbench 中的环境变量访问 S3 数据源。

1.4. 配置集群存储

1.4.1. 在数据科学项目中添加集群存储

对于需要保留数据的数据科学项目，您可以在项目中添加集群存储。另外，您还可以将集群存储连接到特定项目的工作台。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了可添加集群存储的数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要添加集群存储的项目的名称。此时会打开项目详情页面。
3. 点 Cluster storage 选项卡。
4. 点 Add cluster storage。此时会打开 Add storage 对话框。
5. 为集群存储输入一个名称。
6. 输入集群存储的描述。
7. 在持久性存储大小 中，以 GB 为单位输入新大小。最小值为 1 GiB，最大大小为 16384 GiB。
8. 可选：从列表选择一个 workbench，将集群存储连接到现有的工作台。
9. 如果您选择了一个工作台将存储连接到，请在 Mount folder 字段中输入存储目录。
10. 点 Add storage。

验证

- 您添加的集群存储会出现在项目的 Cluster storage 选项卡中。
- 使用您定义的存储大小创建新的持久性卷声明 (PVC)。
- 持久卷声明(PVC)在项目的 Workbenches 选项卡中作为附加存储可见。

1.4.2. 更新集群存储

如果您的数据科学工作需要您更改项目集群存储的识别信息，或者存储所连接的工作台，您可以更新项目的群集存储以更改这些属性。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了包含集群存储的数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要更新存储的项目名称。此时会打开项目详情页面。
3. 点 Cluster storage 选项卡。
4. 点击您要更新的存储旁的操作菜单(HBAC)，然后点 Edit storage。这会打开 Edit storage 页面。
5. 更新存储的属性。
 - a. 如果需要，更新存储的名称。
 - b. 如果需要，更新存储的描述。
 - c. 如果需要，为存储增加持久性存储大小。请注意，您只能增加存储大小。更新存储大小会重启工作台，并使其在一段时间内与大小更改成比例不可用。
 - d. 更新存储所连接的工作台（如果适用）。
 - e. 如果您选择了一个新的工作台将存储连接到，请在 Mount folder 字段中输入存储目录。
6. 点 Update storage。

如果您增加存储大小，则工作台会在一段时间内重启并不可用，这通常与大小更改成比例。

验证

- 您更新的存储会出现在项目的 Cluster storage 选项卡中。

1.4.3. 从数据科学项目中删除集群存储

您可以从数据科学项目中删除集群存储，以帮助您释放资源并删除不需要的存储空间。

先决条件

- 您已登录到 Red Hat OpenShift AI。

- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了带有集群存储的数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要从中删除存储的项目名称。此时会打开项目详情页面。
3. 点 Cluster storage 选项卡。
4. 点您要删除的存储旁的操作菜单(HBAC)，然后点 Delete storage。此时会打开 Delete storage 对话框。
5. 在文本字段中输入存储名称，以确认您想要删除它。
6. 点 Delete storage。

验证

- 您删除的存储不再显示在项目的 Cluster storage 选项卡中。
- 与集群存储关联的持久性卷 (PV) 和持久性卷声明 (PVC) 都会被永久删除。这些数据不可恢复。

1.5. 配置对数据科学项目的访问

1.5.1. 配置对数据科学项目的访问

为了让您与其他用户合作的数据科学项目，您可以共享对项目的访问权限。在创建项目后，您可以从 OpenShift AI 用户界面设置适当的访问权限。

您可以为数据科学项目分配以下访问权限级别：

- admin - 用户可以修改项目的所有区域，包括其详细信息（项目名称和描述）、组件和访问权限。
- 编辑 - 用户可以修改项目的组件，如其工作台，但不能编辑项目的访问权限或其详情（项目名称和描述）。


1.5.2. 共享对数据科学项目的访问

为了让您的机构能够协作工作，您可以将对数据科学项目的访问权限与其他用户和组共享。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了数据科学项目。


流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 从数据科学项目列表中，点您要共享访问权限的数据科学项目的名称。此时会打开项目详情页面。
3. 单击权限选项卡。此时会打开项目的 Permissions 页面。
4. 为一个或多个用户提供项目访问权限。
 - a. 在 Users 部分中，点 Add user。
 - b. 在 Name 字段中输入您要提供项目访问权限的用户的用户名。
 - c. 在 Permissions 列表中，选择以下访问权限级别之一：
 - Admin：具有此访问权限级别的用户可以编辑项目详细信息并管理项目的访问权限。
 - Edit：具有此访问权限级别的用户可以查看和编辑项目组件，如其工作台、数据连接和存储。
 - d. 要确认您的条目，请点击 Confirm ()。
 - e. 可选：要添加其他用户，请点 Add user 并重复此过程。
5. 为一个或多个 OpenShift 组提供项目访问权限。
 - a. 在 Groups 部分，点 Add group。
 - b. 从 Name 列表中选择一个组来提供项目访问权限。



注意

如果您没有 **cluster-admin** 或 **dedicated-admin** 权限，则 Name 列表不可见。相反，会显示一个输入字段，供您配置组权限。

- c. 在 Permissions 列表中，选择以下访问权限级别之一：
 - Admin：具有此访问权限级别的组可以编辑项目详细信息并管理项目的访问权限。
 - Edit：具有此访问权限级别的组可以查看和编辑项目组件，如其工作台、数据连接和存储。
- d. 要确认您的条目，请点击 Confirm ()。
- e. 可选：要添加额外的组，点 Add group 并重复此过程。

验证

- 您向谁提供项目访问权限的用户只能执行其访问权限级别所允许的操作。
- Permissions 选项卡上的 Users 和 Groups 部分显示您提供对项目访问权限的相应用户和组。


1.5.3. 更新对数据科学项目的访问

要更改数据科学项目上的协作级别，您可以更新有权访问您的项目的用户和组的访问权限。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了数据科学项目。
- 您之前已与其他用户或组共享对项目的访问权限。
- 有管理员权限或您是项目所有者。

流程


1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。
Data Science Projects 页面将打开。
2. 点您要更改访问权限的项目名称。
此时会打开项目详情页面。
3. 单击权限选项卡。
此时会打开项目的 Permissions 页面。
4. 更新项目的用户访问权限。
 - a. 在 Name 字段中，更新您要提供项目访问权限的用户的用户名。
 - b. 在 Permissions 列表中，选择以下之一来更新用户访问权限：
 - Admin：具有此访问权限级别的用户可以编辑项目详细信息并管理项目的访问权限。
 - Edit：具有此访问权限级别的用户可以查看和编辑项目组件，如其工作台、数据连接和存储。
 - c. 要确认对该条目的更新，请点击 Confirm ()。
5. 更新 OpenShift 组对项目的访问权限。
 - a. 从 Name 列表中，通过从列表中选择另一个组来更新有权访问项目的组。



注意

如果您没有 **cluster-admin** 或 **dedicated-admin** 权限，则 Name 列表不可见。相反，输入字段会显示可让您配置组权限。

- b. 在 Permissions 列表中，选择以下之一来更新组访问权限：
 - Admin：具有此访问权限级别的组可以编辑项目详细信息并管理项目的访问权限。
 - Edit：具有此访问权限级别的组可以查看和编辑项目组件，如其工作台、数据连接和存储。

- c. 要确认对该条目的更新，请点击 Confirm ()。

验证

- Permissions 选项卡上的 Users 和 Groups 部分显示您更改的项目访问权限的对应用户和组。

1.5.4. 删除对数据科学项目的访问

如果您不再想在数据科学项目上协作工作，您可以通过删除之前提供项目访问权限的用户和组来限制对项目的访问。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了数据科学项目。
- 您之前已与其他用户或组共享对项目的访问权限。
- 有管理员权限或您是项目所有者。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。Data Science Projects 页面将打开。
2. 点您要更改访问权限的项目名称。此时会打开项目详情页面。
3. 单击权限选项卡。此时会打开项目的 Permissions 页面。
4. 点您要撤销的用户或组旁的操作菜单 (⋮)，然后点 Delete。

验证

- 您撤销了访问权限的用户无法再执行其访问权限级别所允许的操作。

第 2 章 在数据科学 IDE 中使用

在 Red Hat OpenShift AI 中，在创建工作台时，您可以选择一个工作台镜像，其中包含用于开发机器学习模型的集成开发环境(IDE)。

OpenShift AI 支持以下数据科学 IDE 以开发 ML 模型：

- JupyterLab
- code-server（技术预览功能）
- RStudio 服务器（技术预览功能）

2.1. 访问工作台 IDE

要访问工作台 IDE，请使用 OpenShift AI 接口中提供的链接。

前提条件

- 您已创建了数据科学项目和一个工作台。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。
2. 点包含工作台的项目的名称。
3. 如果工作台的状态为 Running，请跳至第 4 步。
否则，单击相关工作台的 Status 列中的切换，并等待状态从 Stopped 更改为 Running。
4. 点 Open。

验证

- 为工作台 IDE 打开一个新浏览器窗口。

2.2. 在 JUPYTERLAB 中工作

JupyterLab 是笔记本、代码和数据的最新基于 web 的互动开发环境。您可以在数据科学和机器学习中配置和安排 workflow。JupyterLab 是一个开源 Web 应用程序，支持超过 40 种编程语言，包括 Python 和 R。

2.2.1. 创建并导入笔记本

您可以创建一个空白笔记本，或者从几个不同的源导入笔记本。

2.2.1.1. 创建新笔记本

您可以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。笔记本服务器控制面板包含可用容器镜像的列表，您可以作为单用户笔记本服务器运行。

先决条件

- 确保您已登录到 Red Hat OpenShift AI。

- 确保您已启动了笔记本服务器并登录到 JupyterLab。
- 笔记本镜像存在于 registry、镜像流中，并可访问。

流程

1. 点 File → New → Notebook。
2. 如果出现提示，请从列表中为您的笔记本选择一个内核。
如果要使用内核，点 Select。如果您不想使用内核，点 No Kernel。

验证

- 检查笔记本文件是否在 JupyterLab 接口中看到。


2.2.1.2. 从本地存储上传现有的笔记本文件

您可以从本地存储将现有笔记本加载到 JupyterLab 中，以继续工作，或针对新的用例调整项目。

先决条件

- 用于登录到 JupyterLab 的凭证。
- 启动并运行 Jupyter 笔记本服务器。
- 本地存储中存在笔记本文件。

流程

1. 在 JupyterLab 界面左侧边栏的文件浏览器中，点 Upload Files ()。
2. 找到并选择笔记本文件，再单击 打开。
文件显示在文件浏览器中。

验证

- 笔记本文件显示在 JupyterLab 界面左侧边栏的 File Browser 中。
- 您可以在 JupyterLab 中打开笔记本文件。

2.2.2. 使用 Git 协作笔记本

如果您的笔记本或其他文件存储在 Git 版本控制中，您可以将其从 Git 存储库导入到笔记本服务器，以便在 JupyterLab 中使用它们。准备就绪后，您可以将您的更改推送回 Git 存储库，以便其他人可以查看或使用您的模型。

2.2.3. 使用 JupyterLab 从 Git 存储库上传现有的笔记本文件

您可以使用 JupyterLab 用户界面将 Git 存储库克隆到工作区，以继续您的工作或从外部项目集成文件。

先决条件


- 启动并运行 Jupyter 笔记本服务器。

- 要克隆的 Git 存储库的读取访问权限。

流程

1. 复制 Git 存储库的 HTTPS URL。

- 在 GitHub 中，点击 zfc Code → HTTPS，然后点击 Clipboard 按钮。
- 在 GitLab 上，单击 Clone，再单击 Clone with HTTPS 下的 Clipboard 按钮。

2. 在 JupyterLab 界面中，点 Git Clone 按钮()。

您还可以在菜单中点击 Git → Clone a repository，或者点击 Git 图标()并点 Clone a repository 按钮。

此时会显示克隆存储库对话框。

3. 输入包含笔记本的存储库的 HTTPS URL。
4. 点 CLONE。
5. 若有提示，请输入您的 Git 存储库的用户名和密码。

验证

- 检查存储库的内容是否在 JupyterLab 中的文件浏览器中看到，或者在终端中运行 ls 命令，以验证存储库是否显示为一个目录。

2.2.3.1. 使用命令行界面从 Git 存储库上传现有的笔记本文件

您可以使用命令行界面将 Git 存储库克隆到工作区，以继续工作或从外部项目集成文件。

先决条件

- 启动并运行 Jupyter 笔记本服务器。

流程

1. 复制 Git 存储库的 HTTPS URL。

- 在 GitHub 中，点击 zfc Code → HTTPS，然后点击 Clipboard 按钮。
- 在 GitLab 上，单击 Clone，再单击 Clone with HTTPS 下的 Clipboard 按钮。

2. 在 JupyterLab 中，点 File → New → Terminal 打开终端窗口。

3. 进入 git clone 命令。

```
git clone <git-clone-URL>
```

将 ``<git-clone-URL>`` 替换为 HTTPS URL，例如：

```
[1234567890@jupyter-nb-jdoe ~]$ git clone https://github.com/example/myrepo.git
Cloning into myrepo...
remote: Enumerating objects: 11, done.
```

```
remote: Counting objects: 100% (11/11), done.
remote: Compressing objects: 100% (10/10), done.
remote: Total 2821 (delta 1), reused 5 (delta 1), pack-reused 2810
Receiving objects: 100% (2821/2821), 39.17 MiB | 23.89 MiB/s, done.
Resolving deltas: 100% (1416/1416), done.
```

验证

- 检查存储库的内容是否在 JupyterLab 中的文件浏览器中看到，或者在终端中运行 `ls` 命令，以验证存储库是否显示为一个目录。



2.2.3.2. 使用远程 Git 存储库中的更改更新项目

您可以从远程 Git 存储库将其他用户所做的更改拉取到您的 `data science` 项目中。

先决条件

- 您已配置了远程 Git 存储库。
- 您已将 Git 存储库导入到 JupyterLab，存储库的内容在 JupyterLab 中的文件浏览器中可见。
- 有将文件从远程 Git 存储库拉取到本地存储库的权限。
- 您有登录 Jupyter 的凭证。
- 您有一个启动并运行 Jupyter 服务器。

流程

1. 在 JupyterLab 界面中，点 Git 按钮()。
2. 点 Pull latest changes 按钮()。

验证

- 您可以在 Git 窗格的 History 选项卡中查看从远程存储库拉取的更改。


2.2.3.3. 将项目更改推送到 Git 存储库

要在生产环境中构建和部署您的应用，请将您的工作上传到远程 Git 存储库。

先决条件

- 您已在 JupyterLab 界面中打开了笔记本。
- 您已将相关的 Git 存储库添加到笔记本服务器中。
- 有权限将更改推送到相关的 Git 存储库。
- 已安装 Git 版本控制扩展。

流程

1. 点 File → Save All 保存任何未保存的更改。
2. 点击 Git 图标()在 JupyterLab 界面中打开 Git 窗格。
3. 确认更改的文件显示在 Changed 下。
如果您的更改的文件出现在 Untracked 中，请点击 Git → Simple Staging 以启用简化的 Git 进程。
4. 提交您的更改。
 - a. 确保 Changed 下的所有文件都有一个蓝色复选标记。
 - b. 在 Summary 字段中输入您所做的更改的简短描述。
 - c. 点 Commit。
5. 点 Git → Push to Remote 将您的更改推送到远程存储库。
6. 出现提示时，输入您的 Git 凭据，再单击 OK。

验证

- 您最近推送的更改在远程 Git 存储库中可见。

2.2.4. 管理 Python 软件包

在 JupyterLab 中，您可以查看在 notebook 镜像上安装的 Python 软件包并安装额外的软件包。

2.2.4.1. 查看在笔记本服务器中安装的 Python 软件包

您可以通过在笔记本电脑单元中运行 pip 工具来检查在笔记本电脑服务器上安装了哪些 Python 软件包以及您已有的软件包版本。

先决条件

- 登录 JupyterLab 并打开一个笔记本。

流程

1. 在笔记本中输入以下内容：

```
!pip list
```

2. 运行单元。

验证

- 输出显示了所有已安装的 Python 软件包及其版本的字母顺序列表。例如，如果您在使用 Minimal 镜像创建笔记本服务器后立即使用这个命令，则显示的第一个软件包类似如下：

```
Package              Version
-----
aiohttp              3.7.3
alembic              1.5.2
```

| | |
|-----------------|--------|
| appdirs | 1.4.4 |
| argo-workflows | 3.6.1 |
| argon2-cffi | 20.1.0 |
| async-generator | 1.10 |
| async-timeout | 3.0.1 |
| attrdict | 2.0.1 |
| attrs | 20.3.0 |
| backcall | 0.2.0 |

其他资源

- [在笔记本服务器中安装 Python 软件包](#)

2.2.4.2. 在笔记本服务器中安装 Python 软件包

您可以通过将软件包和版本添加到 `requirements.txt` 文件中，安装不是默认笔记本服务器一部分的 Python 软件包，然后在笔记本单元中运行 `pip install` 命令。



注意

您还可以直接安装软件包，但红帽建议您使用 `requirements.txt` 文件，以便在不同的笔记本中轻松重复使用该文件中声明的软件包。此外，在使用 S2I 构建来部署模型时，使用 `requirements.txt` 文件也很有用。

先决条件

- 登录 JupyterLab 并打开一个笔记本。

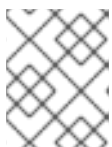
流程

1. 使用以下方法之一创建新文本文件：
 - 单击 `+` 以打开新启动程序，然后单击文本文件。
 - 点 `File` → `New` → `Text File`。
2. 将文本文件重命名为 `requirements.txt`。
 - a. 在文件的名称上单击鼠标右键，然后单击重命名文本。此时会打开 `Rename File` 对话框。
 - b. 在 `New Name` 字段中输入 `requirements.txt`，然后单击 `Rename`。
3. 将要安装的软件包添加到 `requirements.txt` 文件中。

```
altair
```

您可以使用 `==` (等于) 操作符指定要安装的确切版本，例如：

```
altair==4.1.0
```



注意

红帽建议指定准确的软件包版本，以便随着时间的推移增强笔记本服务器的稳定性。在环境行为中，新的软件包版本可能会带来不必要的更改或意外更改。

要同时安装多个软件包，请将每个软件包放在单独的行中。

4. 使用笔记本电脑单元将 requirements.txt 中的软件包安装到您的服务器中。

- a. 在笔记本中创建新单元并输入以下命令：

```
!pip install -r requirements.txt
```

- b. 按 Shift 和 Enter 运行单元。



重要

此命令会在笔记本电脑服务器上安装软件包，但您仍必须在代码单元中运行 `import` 指令，以便在您的代码中使用软件包。

```
import altair
```

验证

- 确认 requirements.txt 中的软件包显示在笔记本电脑服务器上安装的软件包列表中。详情请参阅 [笔记本服务器上安装的 Python 软件包](#)。

2.2.5. 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 AI 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，请联系红帽支持团队。

2.2.5.1. 当登录到 Jupyter 时，看到 403: Forbidden 错误

问题

如果您的管理员为 OpenShift AI 配置了特殊的用户组，您的用户名可能不会添加到默认用户组或 OpenShift AI 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.

2.2.5.2. 我的笔记本服务器没有启动

问题

托管您的笔记本服务器的 OpenShift 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。

2.2.5.3. 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检查。

2.3. 在 CODE-SERVER 中工作

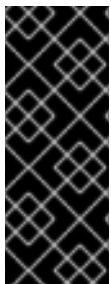
Red Hat OpenShift AI 包含 code-server workbench 镜像。

如需有关 code-server 的更多信息，请参阅 [GitHub 中的 code-server](#)。



注意

基于 Elyra 的管道不适用于 code-server workbench 镜像。



重要

code-server 笔记本镜像目前在 Red Hat OpenShift AI 中作为技术预览功能提供。技术预览功能不受红帽产品服务等级协议 (SLA) 支持，且功能可能并不完整。红帽不推荐在生产环境中使用它们。这些技术预览功能可以使用户提早试用新的功能，并有机会在开发阶段提供反馈意见。

有关红帽技术预览功能支持范围的更多信息，请参阅[技术预览功能支持范围](#)。

2.3.1. 使用 code-server 安装扩展

使用 code-server workbench 镜像，您可以使用扩展来自定义 code-server 环境，来添加新的语言、主题和调试器，并连接到其他服务。您还可以通过扩展来增强数据科学工作的效率，以实现语法高亮显示、自动缩进和括号匹配。

有关您可以使用 code-server 安装的第三方扩展的详情，请查看 [Open VSX 注册表](#)。

先决条件

- 已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了具有 code-server 工作台的数据科学项目。

流程

1. 在 OpenShift AI 仪表板中，单击 Data Science Projects。

Data Science Projects 页面将打开。

2. 点包含您要启动的 code-server workbench 的项目名称。
此时会打开项目详情页面。
3. 单击 Workbenches 选项卡。
4. 点相关工作台的 Status 列中的切换，以启动未运行的工作台。
您开始的工作台的状态从 Stopped 更改为 Running。
5. 工作台启动后，点 Open 以打开工作台笔记本。



6. 在活动栏中，点 Extensions 图标()。
7. 搜索要安装的扩展名称。
8. 点 Install 将扩展添加到您的 code-server 环境中。

验证

- 在 Browser - Installed list on the Extensions 面板中，您会看到您安装的扩展。

第 3 章 使用数据科学项目管道

作为数据科学家，您可以使用 Docker 容器构建可移植机器学习(ML)工作流来增强 OpenShift AI 上的数据科学项目。这可以让您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。

例如，机器学习工作流中的步骤可能包括数据提取、数据处理、功能提取、模型培训、模型验证和模型服务等项目。自动化这些活动可让您的组织开发基于新接收的数据重新培训和更新模型的持续流程。这有助于解决与构建集成机器学习部署以及持续在生产中操作相关的挑战。

您还可以使用 Elyra JupyterLab 扩展在 JupyterLab 中创建并运行数据科学项目。如需更多信息，请参阅在 [JupyterLab 中使用管道](#)。

OpenShift AI 中的数据科学管道现在基于 [KubeFlow Pipelines \(KFP\)版本 2.0](#)。如需更多信息，请参阅[启用数据科学管道 2.0](#)。

OpenShift AI 中的数据科学项目管道由以下组件组成：

- Pipeline 服务器：附加到数据科学项目的服务器，并托管您的数据科学项目。
- Pipeline：管道定义机器学习工作流的配置以及工作流中各个组件之间的关系。
 - Pipeline 代码：YAML 文件中管道的定义。
 - Pipeline 图：图形显示管道运行中执行的步骤及其之间的关系。
- Pipeline 运行：管道的执行。
 - Active run：一个管道运行处于执行阶段，或已停止。
 - 已调度运行：调度的至少执行一次的管道运行。
 - 归档运行：位于运行存档中的管道运行，不再需要。

此功能基于 Kubeflow Pipelines 2.0。使用最新的 Kubeflow Pipelines 2.0 SDK 在 Python 代码中构建您的数据科学项目管道。构建管道后，使用 SDK 将其编译到 Intermediate Representation (IR) YAML 文件中。OpenShift AI 用户界面允许您跟踪和管理管道和管道运行。您可以使用版本控制管理 OpenShift AI 中管道的增量更改。这可以让您迭代开发和部署管道，保留更改的记录。

您可以将管道工件存储在 S3 兼容对象存储存储桶中，以便不消耗本地存储。要做到这一点，您必须首先在存储帐户中配置 S3 存储桶的写入访问权限。

3.1. 启用数据科学管道 2.0

OpenShift AI 中的数据科学管道现在基于 [KubeFlow Pipelines \(KFP\)版本 2.0](#)。在 OpenShift AI 中默认启用并部署了 DSP 2.0。



注意

PipelineConf 类已弃用，且没有 KFP 2.0 等效。

 **重要**

Data Science Pipelines (DSP) 2.0 包含 Argo 工作流的安装。OpenShift AI 不支持直接客户使用此 Argo 工作流安装。

要使用 DSP 2.0 安装或升级到 OpenShift AI，请确保您的集群没有安装 OpenShift AI 的 Argo 工作流的现有安装。

OpenShift AI 创建的 Argo Workflows 资源在 `argoproj.io` 组中 Administration > CustomResourceDefinitions 下有以下标签：

```
labels:
  app.kubernetes.io/part-of: data-science-pipelines-operator
  app.opendatahub.io/data-science-pipelines-operator: 'true'
```

3.1.1. 使用 DSP 2.0 安装 OpenShift AI

要使用 DSP 2.0 安装 OpenShift AI，请确保没有安装 Argo 工作流，没有在集群中安装 DSP，并[按照安装和卸载 OpenShift AI Cloud Service 中描述的安装](#)步骤进行操作。

如果集群中还没有由 DSP 安装的 Argo 工作流安装，则 DSP 会在安装 OpenShift AI 后被禁用。

要启用数据科学项目管道，从集群中删除 Argo 工作流的独立安装。Data Science Pipelines 将自动启用。

3.1.2. 升级到 DSP 2.0

 **重要**

使用 DSP 2.0 升级到 OpenShift AI 后，使用 DSP 1.0 创建的管道将继续运行，但无法从 OpenShift AI 仪表板访问。如果您是当前的 DSP 用户，请不要升级到带有 DSP 2.0 的 OpenShift AI，直到您准备好迁移到新的管道解决方案。

要升级到 DSP 2.0，请按照以下步骤执行：

1. 确保集群没有 OpenShift AI 安装的 Argo 工作流的现有安装，然后按照[升级 OpenShift AI AI Cloud Service](#)中所述的升级步骤进行操作。
如果您升级到启用了 DSP 2.0 的 OpenShift AI，且集群中没有安装 DSP 的 Argo 工作流，则 OpenShift AI 组件不会被升级。要完成组件升级，请禁用 DSP 或者从集群中删除 Argo 工作流的独立安装。然后，组件升级将自动完成。
2. 更新工作台，以使用 notebook 镜像版本 2024.1 或更高版本。如需更多信息，请参阅[更新项目工作台](#)。
3. 手动将管道从 DSP 1.0 迁移到 2.0。如需更多信息，请参阅[管道从 DSP 1.0 迁移到 2.0](#)。

3.1.3. 将管道从 DSP 1.0 迁移到 2.0

OpenShift AI 不会自动将现有 DSP 1.0 实例迁移到 2.0。要使用 DSP 2.0 的现有管道，您必须手动迁移它们。

1. 在带有 DSP 2.0 的 OpenShift AI 上，创建一个新的数据科学项目。
2. 配置新的管道服务器。

- 更新并重新编译您的 DSP 1.0 管道，如从 [KFP SDK v1 迁移到 v2 迁移说明和破坏更改](#) 中所述。

DSP 2.0 does not use the `kfp-tekton` library. In most cases, you can replace usage of `kfp-tekton` with the `kfp` library.

- 将更新的管道导入到基于 DSP 2.0 的数据科学项目。
- (可选) 删除 DSP 1.0 管道服务器。



重要

Data Science Pipelines 1.0 使用 `kfp-tekton` Python 库。Data Science Pipelines 2.0 不使用 `kfp-tekton`。当集群中没有使用剩余的 DSP 1.0 管道服务器时，您可以卸载 `kfp-tekton`。

对于 Data Science Pipelines 2.0，请使用 KFP SDK 的最新版本。如需更多信息，请参阅 [Kubeflow Pipelines SDK API 参考](#)。

3.1.4. 访问 DSP 1.0 管道和历史记录

您可以在 OpenShift Console 的 Pipelines > Project > PipelineRuns 下查看历史 DSP 1.0 管道运行信息。

您仍然可以使用 `kfp-tekton` SDK 进行编程访问管道和管道运行历史记录来连接到 KFP API 服务器。如需更多信息，请参阅 [Tekton 的 Kubeflow Pipelines SDK](#)。

3.1.5. 卸载 OpenShift Pipelines Operator

当迁移到 DSP 2.0 完成时，如果您没有为 Data Science Pipelines 1.0 以外的任何目的使用 OpenShift Pipelines，您可以删除 OpenShift Pipelines Operator。



重要

在删除 OpenShift Pipelines Operator 前，请确保将 DSP 1.0 管道迁移到 2.0 已完成，且集群中没有剩余的 DSP 1.0 管道服务器。

其他资源

- [PyPI: kfp](#)
- [Kubeflow Pipelines SDK API 参考](#)。
- [创建数据科学项目](#)
- [配置管道服务器](#)
- [导入数据科学项目](#)
- [删除管道服务器](#)

3.2. 管理数据科学项目

3.2.1. 配置管道服务器

在 OpenShift AI 中成功创建管道前，您必须配置管道服务器。此任务包括配置管道工件和数据存储位置。



注意

在为管道服务器配置数据连接时，您不需要指定任何存储目录。导入管道时，会在存储桶的根目录中创建 `/pipelines` 文件夹，其中包含管道的 YAML 文件。如果您上传同一管道的新版本，则具有不同 ID 的新 YAML 文件将添加到 `/pipelines` 文件夹中。

运行管道时，工件会存储在存储桶的根目录下的 `/pipeline-name` 文件夹中。



重要

如果您使用外部 MySQL 数据库并使用 DSP 2.0 升级到 OpenShift AI，数据库将迁移到 DSP 2.0 格式，使它与早期版本的 OpenShift AI 不兼容。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，`rh oai-users` 或 `rhoai-admins`）。
- 您已创建了可添加管道服务器的数据科学项目。
- 您有一个兼容 S3 的对象存储桶，您已在存储帐户中配置了 S3 存储桶的写入访问权限。
- 如果您要配置带有外部 MySQL 数据库的管道服务器，您的数据库必须至少使用 MySQL 版本 5.x。但是，红帽建议您使用 MySQL 版本 8.x。
- 如果您要使用 MariaDB 数据库配置管道服务器，您的数据库必须使用 MariaDB 版本 10.3 或更高版本。但是，红帽建议您至少使用 MariaDB 版本 10.5。

流程

1. 在 OpenShift AI 仪表板中，单击 **Data Science Projects**。

Data Science Projects 页面将打开。

2. 点您要为其配置管道服务器的项目名称。

此时会打开项目详情页面。

3. 点 **Pipelines** 选项卡。

4. 点 **Configure pipeline server**。

此时会出现 **Configure pipeline server** 对话框。

5. 在 **Object storage connection** 部分中，为必填字段提供值：

- a. 在 **Access key** 字段中，输入 S3 兼容对象存储供应商的访问密钥 ID。
- b. 在 **Secret key** 字段中，为您指定的 S3 兼容对象存储帐户输入 **secret** 访问密钥。
- c. 在 **Endpoint** 字段中，输入 S3 兼容对象存储桶的端点。
- d. 在 **Region** 字段中，输入 S3 兼容对象存储帐户的默认区域。
- e. 在 **Bucket** 字段中，输入 S3 兼容对象存储桶的名称。



重要

如果您指定了不正确的数据连接设置，则无法在同一管道服务器上更新这些设置。因此，您必须删除管道服务器并配置另一个服务器。

6. 在 **Database** 部分中，点 **Show advanced database options** 来指定用于存储管道数据的数据库，并选择以下一组操作：

- 选择 **Use default database stored on your cluster**，以在项目中部署 MariaDB 数据库。
 - 选择 **Connect to external MySQL database**，以添加与管道服务器可访问的外部数据库的新连接。
 - i. 在 **Host** 字段中，输入数据库的主机名。
 - ii. 在 **Port** 字段中，输入数据库的端口。
 - iii. 在 **Username** 字段中，输入连接到数据库的默认用户名。
 - iv. 在 **Password** 字段中，输入默认用户帐户的密码。
 - v. 在 **Database** 字段中，输入数据库名称。
7. 点 **Configure pipeline server**。

验证

在项目的 **Pipelines** 选项卡中：

- **Import pipeline** 按钮可用。
- 当您点操作菜单（需要），然后点 **View pipeline server configuration** 时，会显示管道服务器详情。

3.2.2. 定义管道

Kubeflow Pipelines SDK 可让您定义端到端机器学习和数据管道。使用最新的 Kubeflow Pipelines 2.0 SDK 在 Python 代码中构建您的数据科学项目管道。构建管道后，使用 SDK 将其编译到

Intermediate Representation (IR) YAML 文件中。定义管道后，您可以将 YAML 文件导入到 OpenShift AI 仪表板，以便您可以配置其执行设置。

您还可以使用 Elyra JupyterLab 扩展在 JupyterLab 中创建并运行数据科学项目。有关在 JupyterLab 中创建管道的更多信息，请参阅在 [JupyterLab 中使用管道](#)。有关 Elyra JupyterLab 扩展的更多信息，请参阅 [Elyra 文档](#)。

其他资源

- [Kubeflow Pipelines 2.0 文档](#)
- [Elyra 文档](#)

3.2.3. 导入数据科学项目

为了帮助您开始使用 OpenShift AI 中的数据科学管道，您可以将包含管道代码的 YAML 文件导入到活跃的管道服务器，或者从 URL 导入 YAML 文件。此文件包含一个使用 Kubeflow 编译器编译的 Kubeflow 管道。将管道导入到管道服务器后，您可以通过创建管道运行来执行管道。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您已使用 Kubeflow 编译器编译了管道，并可以访问生成的 YAML 文件。
- 如果您要从 URL 上传管道，则 URL 可以公开访问。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中，选择要导入管道的项目。
3. 点 **Import pipeline**。
4. 在 **Import pipeline** 对话框中，输入您要导入的管道的详情。
 - a. 在 **Pipeline name** 字段中输入您要导入的管道的名称。
 - b. 在 **Pipeline description** 字段中，输入您要导入的管道的描述。
 - c. 通过执行以下操作之一选择您要从中导入管道的位置：
 - 选择 **Upload a file** 以从本地机器的文件系统上传管道。点 **upload** 或拖放文件来导入管道。
 - 选择 **Import by url** 从 URL 上传管道，然后在文本框中输入 URL。
 - d. 点 **Import pipeline**。

验证

- 您导入的管道会出现在 **Pipelines** 页面中，并在项目详情页面的 **Pipelines** 选项卡中。

3.2.4. 下载数据科学项目

要进一步更改您之前上传到 OpenShift AI 的数据科学项目，您可以从用户界面下载管道的代码。

先决条件

- 已安装 OpenShift Pipelines operator。
- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您已创建了管道并将其导入到可用于下载的活跃管道服务器。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。

Pipelines 页面将打开。

2. 从 **Project** 列表中，选择要下载的管道的项目。
3. 在 **Pipeline name** 列中，点您要下载的管道名称。

Pipeline 详情页面会打开 **Graph** 选项卡。

4. 点 **YAML** 标签。

页面会重新加载以显示管道代码的嵌入式 **YAML** 编辑器。

5. 点 **Download** 按钮(



)将包含管道代码的 **YAML** 文件下载到本地机器。

验证

- 管道代码下载到浏览器的默认目录中，以下载的文件。

3.2.5. 删除数据科学项目管道

如果您不再需要访问仪表板上的数据科学项目管道，您可以将其删除，使其不会出现在 **Data Science Pipelines** 页面中。

先决条件

- 您已登录到 **Red Hat OpenShift AI**。
- 如果您使用专用的 **OpenShift AI** 组，则作为 **OpenShift** 中的用户组或 **admin** 组的一部分（例如，`rh oai-users` 或 `rhoai-admins`）。
- **Pipelines** 页面中有活跃的管道可用。
- 要删除的管道不包含任何管道版本。
- 要删除的管道不包含任何管道版本。如需更多信息，请参阅 [删除管道版本](#)。

流程

1. 在 **OpenShift AI** 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中，选择包含您要从 **Project** 列表中删除的管道的项目。
3. 点您要删除的管道旁的操作菜单 (⋮)，然后点 **Delete pipeline**。
4. 在 **Delete pipeline** 对话框中，在文本字段中输入管道名称，以确认您想要删除它。
5. 点 **Delete pipeline**。

验证

- 您删除的数据科学管道不再出现在 **Pipelines** 页面中。

3.2.6. 删除管道服务器

运行完数据科学管道后，您可以删除管道服务器。删除管道服务器会自动删除其所有相关管道、管道版本并运行。如果您的管道数据存储存储在数据库中，数据库也会及其 **meta-data** 被删除。另外，在删除管道服务器后，您无法创建新的管道或管道运行，直到创建另一个管道服务器为止。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含管道服务器。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中，为您要删除的管道服务器选择项目。
3. 在 **Pipeline 服务器操作** 列表中，选择 **Delete pipeline server**。
4. 在 **Delete pipeline server** 对话框中，在文本字段中输入管道服务器的名称，以确认您想要删除它。
5. 点击 **Delete**。

验证

- 之前分配给已删除的管道服务器的 **Pipelines** 不再出现在相关数据科学项目的 **Pipelines** 页

面中。

- 之前分配给已删除的管道服务器的管道运行不再出现在相关数据科学项目的 **Runs** 页面中。

3.2.7. 查看管道服务器的详情

您可以查看 OpenShift AI 中配置的管道服务器详情，如管道的数据连接详情及其数据存储位置。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 您之前已创建了包含活跃和可用的管道服务器的数据科学项目。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中，选择您要查看的管道服务器 的项目。
3. 在 **Pipeline 服务器操作** 列表中，选择 **View pipeline server configuration**。

验证

- 您可以在 **View pipeline server** 对话框中查看相关的管道服务器详情。


3.2.8. 查看现有的管道

您可以查看您导入到 Red Hat OpenShift AI 的管道详情，如管道最后一次运行、创建的时间、管道执行的运行以及任何关联的管道版本的详情。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 现有的管道可用。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中，为您要查看的管道选择相关项目。
3. 研究列表上的管道。
4. 可选：点相关行中的 **Expand** () 查看与管道关联的管道版本的详情。

验证

- 之前创建的数据科学项目管道列表会出现在 **Pipelines** 页面中。

3.2.9. Pipeline 版本概述

您可以使用版本控制管理 OpenShift AI 中管道的增量更改。这可让您迭代开发和部署管道，保留更改的记录。您可以在 OpenShift AI 仪表板中跟踪和管理您的更改，允许您根据管道的所有可用版本调度和执行运行。

3.2.10. 上传管道版本

您可以将 YAML 文件上传到包含管道最新版本的活跃管道服务器，或者从 URL 上传 YAML 文件。YAML 文件必须包含使用 Kubeflow 编译器编译的 Kubeflow 管道。将管道版本上传到管道服务器后，您可以通过创建管道运行来执行它。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您有一个管道版本，并准备好上传。
- 如果您要从 URL 上传管道版本，则 URL 可以公开访问。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中，选择要将管道版本上传到的项目。
3. 单击 **Import pipeline** 下拉列表，然后选择 **Upload new version**。
4. 在 **Upload new version** 对话框中，输入您要上传的管道版本的详情。
 - a. 在 **Pipeline** 列表中，选择要将管道版本上传到的管道。
 - b. 在 **Pipeline** 版本 **name** 字段中，确认管道版本的名称，并在需要时更改它。

- c. 在 Pipeline 版本 description 字段中输入管道版本的描述。
- d. 通过执行以下操作之一选择要上传管道版本的位置：
 - 选择 Upload a file 以从本地机器的文件系统上传管道版本。点 upload 或拖放文件来导入管道版本。
 - 选择 Import by url 从 URL 上传管道版本，然后在文本框中输入 URL。
- e. 点 Upload。

验证

- 您上传的管道版本显示在 Pipelines 页面中。在包含管道的行中点 Expand (>)查看其版本。
- 包含在 Pipelines 页面中上传的管道版本的行上的 Version 列会递增一个。

3.2.11. 删除管道版本

当不再需要管道时，您可以删除管道的特定版本。删除默认管道版本会自动将默认管道版本改为下一个最新版本。如果没有管道版本，管道会在没有默认版本的情况下保留。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含管道服务器。

- 您已将管道导入到活跃的管道服务器。


流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。

Pipelines 页面将打开。

2. 删除不再需要的管道版本：

- 删除单个管道版本：

- a. 从 **Project** 列表中，选择包含您要删除的管道版本的项目。
- b. 在包含管道的行中，点 **Expand** ()。
- c. 点您要删除的项目版本旁边的操作菜单(HBAC)，然后点 **Delete pipeline version**。

Delete pipeline version 对话框将打开。

- d. 在文本字段中输入管道版本名称，以确认您想要删除它。
- e. 点击 **Delete**。

- 删除多个管道版本：

- a. 在包含您要删除的每个管道版本的行中，选择复选框。
- b. 点 **Import pipeline** 下拉菜单旁边的操作菜单（需要），然后从列表中选择

Delete。

验证

- 您删除的管道版本不再出现在 **Pipelines** 页面或 **data Science** 项目的 **Pipelines** 选项卡中。

3.2.12. 查看管道版本

您可以在 **Pipelines** 页面中查看管道的所有版本。

先决条件

- 已安装 **OpenShift Pipelines operator**。
- 您已登陆到 **Red Hat OpenShift AI**。
- 如果您使用专用的 **OpenShift AI** 组，则作为 **OpenShift** 中的用户组或 **admin** 组的一部分（例如，**rh oai-users** 或 **rhoai-admins**）。
- 您之前已创建了可用的数据科学项目，并包含管道服务器。
- 在活跃和可用的管道服务器上有一个管道。

流程

1. 在 **OpenShift AI** 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
Pipelines 页面将打开。
2. 从 **Project** 列表中，选择包含您要查看的管道版本的项目。
3. 在包含您要查看版本的管道的行中点 **Expand** (
➤

)。

验证

- 您可以在 **Pipelines** 页面中查看管道的版本。

3.2.13. 查看管道版本的详情


您可以查看上传到 Red Hat OpenShift AI（如图形和 YAML 代码）的管道版本的详情。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含管道服务器。
- 在活跃和可用的管道服务器上有一个管道。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。

Pipelines 页面将打开。
2. 从 **Project** 列表中，选择包含您要查看详情的管道版本的项目。
3. 在包含您要查看版本的管道的行中点 **Expand** ()。

4. 点您要查看详情的管道版本。

Pipeline 详情页面 将打开，显示 **Graph** 和 **YAML** 选项卡。

验证

- 在 **Pipeline 详情页面** 中，您可以查看管道图形和 **YAML** 代码。

3.3. 管理管道运行

3.3.1. 管道运行概述

管道运行是数据科学管道的单一执行。作为数据科学家，您可以使用 **OpenShift AI** 定义、管理和跟踪数据科学管道的执行。您可以从 **OpenShift AI** 用户界面的 **Runs** 页面中查看以前执行、调度和归档运行的记录。

您可以优化管道运行的使用以实现可移植性。您可以克隆管道运行来相应地复制和扩展它们，或者在您要保留其执行记录时归档它们，但不再需要它们。您可以删除不再保留的归档运行，也可以将其恢复到以前的状态。

您可以在创建后执行一次运行一次、立即执行，或者基于重复操作。重复运行由带有所有参数值和 **run** 触发器的管道副本组成。运行触发器指示重复运行的时间执行。您可以定义以下运行触发器：

- **Periodic**：调度运行，以间隔方式执行。
- **Cron**：调度运行，以 **cron** 任务执行。

您还可以将同一运行的多个实例配置为同时执行，范围从一到 10 到 10。执行后，您可以从 **OpenShift AI** 用户界面的 **run Details** 页面中跟踪运行的进度。在这里，您可以查看运行的图表和输出工件。管道运行可以处于以下状态之一：

- **已调度运行**：调度的至少执行一次的管道运行。
- **Active run**：一个管道运行处于执行阶段，或已停止。

- 归档运行：位于运行存档中的管道运行，不再需要。

您可以使用 `catch up run` 来确保管道运行在暂停时不会永久落后调度。例如，如果您重新启用一个暂停的重复运行，则每个丢失的运行间隔运行调度程序回填。如果您禁用 `catch up run`，且您有一个调度的运行间隔，则运行调度程序只会调度最新运行间隔的运行执行。默认启用捕获运行。但是，如果您的管道在内部处理回填，红帽建议您禁用捕获运行以避免重复回填。

您可以在活跃的管道运行中查看和分析每个步骤的日志。使用日志查看器，您可以搜索特定的日志消息，查看每个步骤的日志，并将步骤日志下载到本地机器。

3.3.2. 使用数据科学管道存储数据

当您运行数据科学项目时，OpenShift AI 存储管道 YAML 配置文件，并在存储桶的根目录中运行工件。包含管道运行工件的目录可能会因您执行管道运行的位置而有所不同。如需更多信息，请参阅下表：

表 3.1. Pipeline 配置文件和工件存储位置

| Pipeline 运行源 | Pipeline 存储目录 | 运行工件存储目录 |
|---------------------|---|--|
| OpenShift AI 仪表盘 | <code>/pipelines/<pipeline_version_id></code> 示例： <code>/pipelines/1d01c4eb-d2ab-4916-9935-a73a5580f1fb</code> | <code>/<pipeline_name>/<pipeline_run_id></code> 示例： <code>iris-training-pipeline/2g48k8pw-a8ib-4884-9145-h41j7599h3ds</code> |
| JupyterLab Elyra 扩展 | <code>/pipelines/<pipeline_version_id></code> | <code>/<pipeline_name_timestamp></code> 示例： <code>/hello-generic-world-0523161704</code> 使用 JupyterLab Elyra 扩展，您还可以设置 对象存储路径前缀 。 示例： <code>/iris-project/hello-generic-world-0523161704</code> |

3.3.3. 查看活跃的管道运行

您可以查看之前在 OpenShift AI 执行的管道运行列表。在这个列表中，您可以查看与管道运行相关的详情，如运行所属的管道版本，以及运行的状态、持续时间和执行开始时间。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并有一个管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 您之前已执行管道运行。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Runs**。
2. 在 **Runs** 页面中，为您要查看的活动管道运行选择项目。
3. 在 **Run details** 页面上，点 **Active** 选项卡。

运行完成后，运行的状态会显示在表中的 **Status** 列中，指示运行成功还是失败。

验证

- 活跃的运行列表会出现在 **Runs** 页面上的 **Active** 选项卡中。

3.3.4. 执行管道运行

您可以通过创建一个在创建后立即执行的活跃管道运行来实例化一个管道执行。

先决条件

- 您已登录到 Red Hat OpenShift AI。

- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您已将管道导入到活跃的管道服务器。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Runs**。
2. 在 **Runs** 页面上，选择要为其创建运行的项目。
3. 点 **Create run**。
4. 在 **Create run** 页面中，配置运行：
 - a. 在 **Name** 字段中输入运行的名称。
 - b. 在 **Description** 字段中，输入 run 的描述。
 - c. 在 **Pipeline** 列表中选择您要为其创建运行的管道。另外，要创建新管道，请点 **Create new pipeline**，并完成 **Import pipeline** 对话框中的相关字段。
 - d. 在 **Pipeline 版本** 列表中，选择要为其创建运行的管道版本。或者，要上传新版本，请单击 **Upload new version**，并完成 **Upload new version** 对话框中的相关字段。
 - e. 通过从列表中选择参数来配置运行的输入参数。
 - f. 点 **Create run**。

验证

- 您创建的管道运行会出现在 **Runs** 页面上的 **Active** 选项卡中。

3.3.5. 停止活跃的管道运行

如果您不再需要活跃的管道运行继续执行，您可以在其定义的结束日期前停止运行。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 之前创建的数据科学项目包含管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 活跃的管道运行当前正在执行。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Runs**。
2. 在 **Runs** 页面中，选择包含您要停止的活跃运行的管道 的项目。
3. 在 **Active** 选项卡中，点您要删除的活跃运行旁的操作菜单(ProductShortName)，然后点 **Stop**。

运行停止时可能会有短暂的延迟。

验证

- 在活跃运行列表中，运行的状态为 "stopped"。

3.3.6. 复制活跃的管道运行

为了更轻松地使用相同的配置执行管道运行，您可以重复它们。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 在 **Runs** 页面上的 **Active** 选项卡中，可以使用活跃的运行来重复。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Runs**。
2. 在 **Runs** 页面中，选择具有您要复制的管道运行的项目。
3. 点相关活跃运行旁的操作菜单(HBAC)，然后点 **Duplicate**。
4. 在 **Duplicate run** 页面中，配置重复的运行：
 - a. 在 **Name** 字段中输入重复运行的名称。

- b. 在 **Description** 字段中，输入重复运行的描述。
- c. 在 **Pipeline** 列表中选择包含重复运行的管道。
- d. 在 **Pipeline 版本** 列表中，选择包含重复运行的管道版本。
- e. 在 **Parameters** 部分中，通过从列表中选择适当的参数，为您要复制的运行配置输入参数。
- f. 点 **Create run**。

验证

- 重复管道运行会出现在 **Runs** 页面上的 **Active** 选项卡中。

3.3.7. 查看调度的管道运行

您可以查看在 **OpenShift AI** 中计划执行的管道运行列表。在这个列表中，您可以查看与管道运行相关的详情，如运行所属的管道版本。您还可以查看运行状态、执行频率和调度。

先决条件

- 您已登陆到 **Red Hat OpenShift AI**。
- 如果您使用专用的 **OpenShift AI** 组，则作为 **OpenShift** 中的用户组或 **admin** 组的一部分（例如，**rh oai-users** 或 **rhoai-admins**）。
- 您之前已创建了可用的数据科学项目，并包含管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 您已调度的可用于查看的管道运行。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Runs**。

Runs 页面将打开。

2. 从 **Project** 列表中，选择您要查看的调度管道运行的项目。

3. 单击 **Schedules** 选项卡。

4. 查看显示调度的运行的列表。

调度了运行后，运行的状态指示运行是否已准备好执行，或者无法执行。要更改其执行可用性，请点 run 的 **Status** 图标。

验证

- 调度的运行列表会出现在 **Runs** 页面上的 **Schedules** 选项卡中。

3.3.8. 使用 cron 任务调度管道运行

您可以使用 **cron** 作业来调度管道运行在特定时间执行。**Cron Job** 可用于创建周期性和重复的任务，也可以为特定时间调度个别任务，例如，要在低活动期间调度运行。要在 OpenShift AI 中成功运行，您必须使用支持的格式。如需更多信息，请参阅 [Cron Expression Format](#)。

以下示例显示了正确的格式：

| 运行发生 | Cron 格式 |
|--------------|----------------------|
| 每五分钟 | @every 5m |
| 每 10 分钟 | 0 */10 * * * * |
| 每天 16:16 UTC | 0 16 16 * * * |
| 每小时的每日 | 0 0,15,30,45 * * * * |

| 运行发生 | Cron 格式 |
|---------------------------------|---------------------|
| 在 Monday 和 Tuesday 上为 15:40 UTC | 0 40 15 * * MON,TUE |

其他资源

- [Cron Expression 格式](#)

3.3.9. 调度管道运行

要重复运行管道，您可以创建一个调度的管道运行。


先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您已将管道导入到活跃的管道服务器。

流程

1. 在 OpenShift AI 仪表板中，点 Data Science Pipelines → Runs。

Runs 页面将打开。
2. 从 Project 列表中，选择要为其调度运行的项目。
3. 单击 Schedules 选项卡。

4. 单击 **Schedule run**。
5. 在 **Schedule run** 页面中，配置您要调度的运行：
 - a. 在 **Name** 字段中输入运行的名称。
 - b. 在 **Description** 字段中，输入 run 的描述。
 - c. 在 **Trigger type** 列表中，选择以下选项之一：
 - 选择 **Periodic** 来指定执行频率。在 **Run every** 字段中，输入数字值并从列表中选择执行频率。
 - 选择 **Cron** 以 **cron** 格式指定执行计划。这会创建一个 **cron** 作业来执行运行。点 **Copy** 按钮()将 **cron** 作业调度复制到剪贴板。左侧的字段代表秒。有关使用支持的 **cron** 格式调度任务的更多信息，请参阅 [Cron Expression Format](#)。
 - d. 在 **Maximum concurrent run** 字段中，指定可同时执行的运行数量，范围从 1 到 10。
 - e. 对于 **Start date**，指定 运行的开始日期。使用 **Calendar** 和从时间列表中选择开始日期。
 - f. 对于 **End date**，指定 运行的结束日期。使用 **Calendar** 和 **end time** 从时间列表中选择结束日期。
 - g. 对于 **Catch up**，请启用或禁用捕获运行。您可以使用 **catch up run** 来确保管道运行在暂停时不会永久落后调度。例如，如果您重新启用一个暂停的重复运行，则每个丢失的运行间隔运行调度程序回填。
 - h. 在 **Pipeline** 列表中选择您要为其创建运行的管道。另外，要创建新管道，请点 **Create new pipeline**，并完成 **Import pipeline** 对话框中的相关字段。

- i. 在 **Pipeline 版本** 列表中，选择要为其创建运行的管道版本。或者，要上传新版本，请单击 **Upload new version**，并完成 **Upload new version** 对话框中的相关字段。
- j. 通过从列表中选择参数来配置运行的输入参数。
- k. 单击 **Schedule run**。

验证

- 您创建的管道运行会出现在 **Runs** 页面上的 **Schedules** 选项卡中。

3.3.10. 复制调度的管道运行


为了更轻松地调度运行，以作为管道配置的一部分执行，您可以复制现有的调度运行。

先决条件

- 您已登陆到 **Red Hat OpenShift AI**。
- 如果您使用专用的 **OpenShift AI** 组，则作为 **OpenShift** 中的用户组或 **admin** 组的一部分（例如，**rh oai-users** 或 **rhoai-admins**）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 在 **Runs** 页面上的 **Schedules** 选项卡中，可以使用调度的运行重复。

流程

1. 在 **OpenShift AI** 仪表板中，点 **Data Science Pipelines** → **Runs**。

2. 在 **Runs** 页面中，选择具有您要复制的管道运行的项目。
3. 单击 **Schedules** 选项卡。
4. 单击您要重复的运行旁的操作菜单(HBAC)，然后点 **Duplicate**。
5. 在 **Duplicate schedule** 页面中，配置重复运行：
 - a. 在 **Name** 字段中输入重复运行的名称。
 - b. 在 **Description** 字段中，输入重复运行的描述。
 - c. 在 **Trigger type** 列表中，选择以下选项之一：
 - 选择 **Periodic** 来指定执行频率。在 **Run every** 字段中，输入数字值并从列表中选择执行频率。
 - 选择 **Cron** 以 **cron** 格式指定执行计划。这会创建一个 **cron** 作业来执行运行。点 **Copy** 按钮()将 **cron** 作业调度复制到剪贴板。左侧的字段代表秒。有关使用支持的 **cron** 格式调度任务的更多信息，请参阅 [Cron Expression Format](#)。
 - d. 对于 **最大并发运行**，请指定可同时执行的运行数量，范围为 1 到 10。
 - e. 对于 **Start date**，请指定重复运行的开始日期。使用 **Calendar** 和从时间列表中选择开始日期。
 - f. 对于 **End date**，为重复运行指定结束日期。使用 **Calendar** 和 **end time** 从时间列表中选择结束日期。
 - g.

对于 **Catch up**，请启用或禁用捕获运行。您可以使用 **catch up run** 来确保管道运行在暂停时不会永久落后调度。例如，如果您重新启用一个暂停的重复运行，则每个丢失的运行间隔运行调度程序回填。

- h. 在 **Pipeline** 列表中选择您要为其创建重复运行的管道。另外，要创建新管道，请点击 **Create new pipeline**，并完成 **Import pipeline** 对话框中的相关字段。
- i. 在 **Pipeline 版本** 列表中，选择您要为其创建重复运行的管道版本。或者，要上传新版本，请点击 **Upload new version**，并完成 **Upload new version** 对话框中的相关字段。
- j. 通过从列表中选择参数来配置运行的输入参数。
- k. 单击 **Schedule run**。

验证

- 您复制的管道运行会出现在 **Runs** 页面上的 **Schedules** 选项卡中。

3.3.11. 删除调度的管道运行

要丢弃之前调度的管道运行，但不再需要，您可以删除它们，使其不会出现在 **Runs** 页面中。

先决条件

- 您已登陆到 **Red Hat OpenShift AI**。
- 如果您使用专用的 **OpenShift AI** 组，则作为 **OpenShift** 中的用户组或 **admin** 组的一部分（例如，**rh oai-users** 或 **rhoai-admins**）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您已将管道导入到活跃的管道服务器。

- 您之前已调度了一个可以删除的运行。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Runs**。

Runs 页面将打开。

2. 从 **Project** 列表中，选择包含您要删除的已调度运行的管道的项目。

页面刷新，以显示管道的调度在 **Schedules** 选项卡上运行。

3. 点您要删除的已调度运行旁的操作菜单 (⋮)，然后点 **Delete**。

此时会打开 **Delete schedule** 对话框。

4. 在文本字段中输入运行的名称，以确认您想要删除它。

5. 点击 **Delete**。

验证

- 您删除的运行不再出现在 **Schedules** 选项卡中。

3.3.12. 查看管道运行的详情

要更好地了解管道运行，您可以查看之前触发的管道运行的详情，如图形、执行详情和运行输出。


先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分

(例如, `rh oai-users` 或 `rhoai-admins`)。

- 您之前已创建了可用的数据科学项目, 并包含管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 之前您已触发了管道运行。

流程

1. 在 OpenShift AI 仪表板中, 点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中, 选择您要查看运行详情的项目。
3. 对于您要查看运行详情的管道, 点 **Expand** ()。
4. 点管道版本的操作菜单(`ProductShortName`), 然后点 **View run**。
5. 在 **Runs** 页面上, 点您要查看详情的运行的名称。

验证

- 在 **Run details** 页面中, 您可以查看运行的图形、执行详情、输入参数、步骤日志和运行输出。

3.3.13. 查看归档的管道运行

您可以查看已归档的管道运行列表。您可以查看归档的管道运行的详情, 如管道版本、运行状态、持续时间和执行开始日期。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并有一个管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 归档的管道运行存在。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Runs**。
2. 在 **Runs** 页面中，为您要查看的归档管道运行选择项目。
3. 点 **归档** 选项卡。

验证

- 在 **Runs** 页面上的 **Archived** 选项卡中会出现一个归档运行的列表。

3.3.14. 归档管道运行

您可以通过存档来保留管道运行的记录。如果需要，您可以从存档中恢复运行，以重复使用或删除不再需要的运行。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分

(例如, `rh oai-users` 或 `rhoai-admins`)。

- 您之前已创建了可用的数据科学项目, 并有一个管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 您之前已执行了一个可用的管道运行。

流程

1. 在 OpenShift AI 仪表板中, 点 **Data Science Pipelines** → **Runs**。
2. 在 **Runs** 页面中, 从 **Project** 列表中选择您要归档的管道运行的项目。
3. 在 **Run details** 页面中, 点您要归档的运行旁的操作菜单(**ProductShortName**), 然后点 **Archive**。
4. 在 **Archiving run** 对话框中, 在文本字段中输入运行名称以确认您要归档它。
5. 点 **归档**。

验证

- 归档的运行不会出现在 **Active** 选项卡中, 而是显示在 **Runs** 页面上的 **Archived** 选项卡中。

3.3.15. 恢复归档的管道运行

您可以将归档的运行恢复到 **active** 状态。

先决条件

- 您已登陆到 Red Hat OpenShift AI。

- 如果您使用专用的 **OpenShift AI** 组，则作为 **OpenShift** 中的用户组或 **admin** 组的一部分（例如，**rh oai-users** 或 **rhoai-admins**）。
- 您之前已创建了可用的数据科学项目，并有一个管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 您的项目中已存在一个归档的运行。

流程

1. 在 **OpenShift AI** 仪表板中，点 **Data Science Pipelines** → **Runs**。
2. 在 **Runs** 页面中，选择您要恢复的归档管道运行的项目。
3. 在 **Run details** 页面上，点 **Archived** 选项卡。
4. 点击您要恢复的运行旁的操作菜单(HBAC)，然后点 **Restore**。
5. 在 **Restore run** 对话框中，在文本字段中输入运行名称以确认您要恢复它。
6. 单击 **Restore**。

验证

- 恢复的运行会出现在 **Runs** 页面上的 **Active** 选项卡中。

3.3.16. 删除归档的管道运行

您可以从 **OpenShift AI run** 归档中删除管道运行。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并有一个管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 之前您已存档了管道运行。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Runs**。
2. 在 **Runs** 页面中，为您要删除的归档管道运行选择项目。
3. 在 **Run details** 页面中，单击 **Archived**。
4. 点您要删除的运行旁的操作菜单(HBAC)，然后点 **Delete**。
5. 在 **Delete run** 对话框中，在文本字段中输入运行名称以确认您想要删除它。
6. 单击 **Delete**。

验证

- 您删除的归档运行不再出现在 **Runs** 页面上的 **Archived** 选项卡中。

3.3.17. 复制已归档的管道运行

为了更容易地重现与存档中运行的相同配置运行，您可以重复它们。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含配置的管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 在 Runs 页面上的 Archived 选项卡中，可以使用一个归档的运行来重复。

流程

1. 在 OpenShift AI 仪表板中，点 Data Science Pipelines → Runs。
2. 在 Runs 页面中，选择具有您要复制的管道运行的项目。
3. 点 归档 选项卡。
4. 点相关归档运行旁的操作菜单(HBAC)，然后点 Duplicate。
5. 在 Duplicate run 页面中，配置重复的运行：
 - a. 在 Name 字段中输入重复运行的名称。
 - b. 在 Description 字段中，输入重复运行的描述。

- c. 在 Pipeline 列表中选择包含重复运行的管道。
- d. 在 Pipeline 版本 列表中，选择包含重复运行的管道版本。
- e. 在 Parameters 部分中，通过从列表中选择适当的参数，为您要复制的运行配置输入参数。
- f. 点 Create run。

验证

- 重复管道运行会出现在 Runs 页面上的 Active 选项卡中。

3.4. 使用管道日志

3.4.1. 关于管道日志

您可以在触发的管道运行中查看和分析每个步骤的步骤日志。

为了帮助您排除故障并审核管道，您可以使用 OpenShift AI 仪表板中的 log viewer 查看和分析这些步骤日志。在这里，您可以搜索特定的日志消息，查看每个步骤的日志，并将步骤日志下载到本地机器。

如果步骤日志文件超过其容量，日志查看器上会出现一个警告，表示日志窗口显示部分内容。扩展警告会显示更多信息，如日志查看器如何每三秒刷新一次，并且每个步骤日志会显示收到的日志消息的最后 500 行。另外，您可以点 下载所有步骤日志，将所有步骤日志下载到本地机器。

每个步骤都有一组容器日志。您可以通过在日志查看器列表选择一个容器来查看这些容器日志。Step-main 容器日志由步骤的日志输出组成。step-copy-artifact 容器日志包含与发送到 s3 兼容存储的工件数据相关的输出。如果管道中的步骤间传输的数据大于 3 KB，则通常提供五个容器日志。这些日志包含与持久性卷声明(PVC)间传输的数据相关的输出。

3.4.2. 查看管道步骤日志


为了帮助您排除故障并审核管道，您可以使用日志查看器来查看和分析每个管道步骤的日志。在这里，您可以搜索特定的日志消息，并下载管道中每个步骤的日志。如果管道正在运行，您也可以暂停并从

日志查看器恢复日志。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您之前已创建了可用的数据科学项目，并包含管道服务器。
- 您已将管道导入到活跃的管道服务器。
- 之前您已触发了管道运行。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中，选择要查看日志 的项目。
3. 对于您要查看日志的管道，点 **Expand** ()。
4. 点击包含您要查看管道日志的项目版本所在行中的操作菜单(HBAC)，然后点 **View run**。
5. 在 **Runs** 页面上，点击您要查看日志的运行名称。
6. 在 **Run details** 页面上的图表中，点您要查看日志的管道步骤。

7. 点 **Logs** 选项卡。
8. 要查看另一个管道步骤的日志，请在 **Steps** 列表中选择您要查看日志的步骤。
9. 使用日志查看器分析日志。
 - 要搜索特定的日志消息，请在搜索栏中输入至少部分消息。
 - 要在单独的浏览器窗口中查看完整日志，请点操作菜单（需要）并选择 **View raw logs**。另外，要扩展日志查看器的大小，请点操作菜单（需要）并选择 **Expand**。

验证

- 您可以查看管道中每个步骤的日志。

3.4.3. 下载管道步骤日志



您可以使用 **OpenShift AI** 仪表板上的 **log viewer** 查看管道运行的步骤日志，您可以下载它们以进一步分析。您可以选择下载属于管道中所有步骤的日志，或者您只能为日志查看器中显示的步骤日志下载日志。

先决条件

- 您已登陆到 **Red Hat OpenShift AI**。
- 如果您使用专用的 **OpenShift AI** 组，则作为 **OpenShift** 中的用户组或 **admin** 组的一部分（例如，**rh oai-users** 或 **rhoai-admins**）。
- 您之前已创建了可用的数据科学项目，并包含管道服务器。
- 您已将管道导入到活跃的管道服务器。

- 之前您已触发了管道运行。

流程

1. 在 OpenShift AI 仪表板中，点 **Data Science Pipelines** → **Pipelines**。
2. 在 **Pipelines** 页面中，选择要下载日志 的项目。
3. 对于您要下载日志的管道，点 **Expand** ()。
4. 点 **View** 在包含您要下载日志的管道版本的行中运行。
5. 在 **Runs** 页面上，点您要下载日志的运行名称。
6. 在 **Run details** 页面的图表中，点您要下载日志的管道步骤。
7. 点 **Logs** 选项卡。
8. 在日志查看器中，点 **Download** 按钮()。
 - a. 选择 **Download current stop log** 以下载当前管道步骤的日志。
 - b. 选择 **Download all step logs**，为管道运行中的所有步骤下载日志。

验证

- 步骤将下载至浏览器的默认目录以下载文件。

3.5. 在 JUPYTERLAB 中使用管道

3.5.1. JupyterLab 中的管道概述

您可以使用 Elyra 在 JupyterLab 中创建视觉端到端管道工作流。Elyra 是 JupyterLab 的扩展，它为您提供 **Pipeline Editor** 来创建管道工作流，可在 OpenShift AI 中执行。

在创建以下笔记本镜像的最新版本时，您可以访问 JupyterLab 中的 Elyra 扩展：

- **Standard Data Science**
- **PyTorch**
- **TensorFlow**
- **TrustyAI**
- **HabanaAI**

当使用 **Pipeline Editor** 以视觉方式设计管道时，需要最少的编码来创建和运行管道。有关 Elyra 的更多信息，请参阅 [Elyra 文档](#)。有关 **Pipeline Editor** 的更多信息，请参阅 [Visual Pipeline Editor](#)。创建管道后，您可以在本地的 JupyterLab 中运行它，或使用 OpenShift AI 中的数据科学管道远程运行。

管道创建过程由以下任务组成：

- 创建包含工作台的数据科学项目。
- 创建管道服务器。
- 在 JupyterLab 中的 **Pipeline Editor** 中创建新管道。

- 通过添加 Python 笔记本或 Python 脚本并定义其运行时属性来开发管道。
- 定义执行依赖项。
- 运行或导出管道。

在 JupyterLab 中运行管道前，您的管道实例必须包含运行时配置。运行时配置定义了管道实例的连接信息，以及 S3 兼容云存储。

如果您创建工作台作为数据科学项目的一部分，则会自动创建默认的运行配置。但是，如果您在 OpenShift AI 仪表板中从 Jupyter 标题创建笔记本，则必须在 JupyterLab 中运行管道前创建一个运行时配置。有关运行时配置的更多信息，请参阅[运行时配置](#)。作为先决条件，在创建工作台前，请确保已在与工作台相同的数据科学项目中创建并配置管道服务器。

您可以使用 S3 兼容云存储，在笔记本和脚本执行期间使数据提供给笔记本和脚本。您的云存储必须可从运行 JupyterLab 且托管 Data Science Pipelines 的集群中的机器访问。在 JupyterLab 中创建并运行管道前，请确保您具有兼容 s3 的存储凭证。

其他资源

- [Elyra 文档](#)
- [Visual Pipeline Editor](#)
- [运行时配置](#)

3.5.2. 访问管道编辑器

您可以使用 Elyra 在 JupyterLab 中创建视觉端到端管道工作流。Elyra 是 JupyterLab 的扩展，它为您提供 Pipeline Editor 来创建管道工作流，可在 OpenShift AI 中执行。

先决条件

- 您已登录到 Red Hat OpenShift AI。

- 如果您使用专用的 **OpenShift AI** 组，则作为 **OpenShift** 中的用户组或 **admin** 组的一部分（例如，**rh oai-users** 或 **rhoai-admins**）。
- 您已创建了数据科学项目。
- 您已创建了带有 **Standard Data Science** 笔记本镜像的工作台。
- 您已在包含工作台的 **data Science** 项目中创建和配置了管道服务器。
- 您已从包含 **Elyra** 扩展（标准数据科学、**TensorAI**、**PyAITor** 或 **HabanaAI**）的笔记本镜像创建并启动 **Jupyter** 服务器。
- 您可以访问 **S3** 兼容存储。

流程

1. 打开 **JupyterLab** 后，确认 **JupyterLab launcher** 会自动显示。
2. 在 **JupyterLab launcher** 的 **Elyra** 部分中，点 **Pipeline Editor** 标题。

Pipeline Editor 将打开。

验证

- 您可以在 **JupyterLab** 中查看 **Pipeline Editor**。



3.5.3. 创建运行时配置

如果您创建工作台作为数据科学项目的一部分，则会自动创建默认的运行时代配置。但是，如果您在 **OpenShift AI** 仪表板中从 **Jupyter** 标题创建笔记本，则必须在 **JupyterLab** 中运行管道前创建一个运行时代配置。这可让您为管道实例和 **S3** 兼容云存储指定连接信息。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您可以访问兼容 S3 的云存储。
- 您已创建了一个包含工作台的数据科学项目。
- 您已在包含工作台的数据 Science 项目中创建和配置了管道服务器。
- 您已从包含 Elyra 扩展（标准数据科学、TensorAI、PyAITor 或 HabanaAI）的笔记本镜像创建并启动 Jupyter 服务器。

流程

1. 在 JupyterLab 的左侧边栏中，点 Runtimes ()。
2. 点 Create new runtime configuration 按钮()。
Add new Data Science Pipelines runtime configuration 页面会打开。
3. 完成相关字段以定义运行时配置。
 - a. 在 Display Name 字段中输入运行时配置的名称。
 - b. 可选：在 Description 字段中输入用于定义运行时配置的描述。

- c. 可选：在 **Tags** 字段中，点 **Add Tag** 为管道实例定义类别。输入标签的名称，然后按 **Enter** 键。
- d. 定义数据科学项目管道的凭证：
- i. 在 **Data Science Pipelines API Endpoint** 字段中，输入数据科学项目管道的 **API** 端点。不要在此字段中指定 **pipelines** 命名空间。
- ii. 在 **Public Data Science Pipelines API Endpoint** 字段中，输入数据科学项目管道的公共 **API** 端点。

**重要**

您可以从仪表板中的 **Data Science Pipelines** → **Runs** 页面获取 **Data Science Pipelines API** 端点。复制相关的端点，并在 **Public Data Science Pipelines API Endpoint** 字段中输入它。

- iii. 可选：在 **Data Science Pipelines User Namespace** 字段中输入相关用户命名空间来运行管道。
- iv. 在 **Authentication Type** 列表中，选择验证管道所需的身份验证类型。

**重要**

如果您直接从仪表板上的 **Jupyter** 标题创建了笔记本，请从 **Authentication Type** 列表中选择 **EXISTING_BEARER_TOKEN**。

- v. 在 **Data Science Pipelines API Endpoint Username** 字段中，输入身份验证类型所需的用户名。
- vi. 在 **Data Science Pipelines API Endpoint Password Or Token** 中，输入身份验证类型所需的密码或令牌。



重要

要获取 Data Science Pipelines API 端点令牌，在 OpenShift Web 控制台右上角，点您的用户名并选择 **Copy login command**。登录后，点 **Display token**，并复制来自 **Log in with this token** 命令的 `--token=` 的值。

e.

定义 S3 兼容存储的连接信息：

i.

在 **Cloud Object Storage Endpoint** 字段中，输入 S3 兼容存储的端点。有关 Amazon s3 端点的更多信息，请参阅 [Amazon Simple Storage Service 端点和配额](#)。

ii.

可选：在 **Public Cloud Object Storage Endpoint** 字段中输入 S3 兼容存储的 URL。

iii.

在 **Cloud Object Storage Bucket Name** 字段中输入存储管道工件的存储桶名称。如果存储桶名称不存在，则会自动创建。

iv.

从 **Cloud Object Storage Authentication Type** 列表中，选择访问 S3 兼容云存储所需的身份验证类型。如果使用 AWS S3 存储桶，请从列表中选择 **KUBERNETES_SECRET**。

v.

在 **Cloud Object Storage Credentials Secret** 字段中，输入包含存储用户名和密码的 **secret**。此 **secret** 在相关用户命名空间中定义（如果适用）。另外，它必须存储在托管管道运行的集群中。

vi.

在 **Cloud Object Storage Username** 字段中输入要连接到 S3 兼容云存储的用户名（如果适用）。如果使用 AWS S3 存储桶，请输入 **AWS Secret Access Key ID**。

vii.

在 **Cloud Object Storage Password** 字段中，输入要连接到 S3 兼容云存储的密码（如果适用）。如果使用 AWS S3 存储桶，请输入 **AWS Secret Access Key**。

f.

点 **Save & Close**。

验证

- 您创建的运行时配置显示在 JupyterLab 左侧边栏中的 **Runtimes** 选项卡()中。


3.5.4. 更新运行时配置

为确保运行时配置准确和更新，您可以更改现有运行时配置的设置。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您可以访问 S3 兼容存储。
- 您已创建了一个包含工作台的数据科学项目。
- 您已在包含工作台的数据 Science 项目中创建和配置了管道服务器。
- 之前创建的运行时配置在 JupyterLab 界面中可用。
- 您已从包含 Elyra 扩展（标准数据科学、TensorAI、PyAITor 或 HabanaAI）的笔记本镜像创建并启动 Jupyter 服务器。

流程

1. 在 JupyterLab 的左侧边栏中，点 **Runtimes** ()。

- 2.

将光标悬停在您要更新的运行时配置上，然后点 **编辑** 按钮(



)。

Data Science Pipelines 运行时配置页面将打开。

3.

填写相关字段以更新运行时配置。

a.

在 **Display Name** 字段中，更新运行时配置的名称（如果适用）。

b.

可选：在 **Description** 字段中，更新运行时配置的描述（如果适用）。

c.

可选：在 **Tags** 字段中，点 **Add Tag** 为管道实例定义类别。输入标签的名称，然后按 **Enter** 键。

d.

定义数据科学项目管道的凭证：

i.

在 **Data Science Pipelines API Endpoint** 字段中，更新数据科学项目管道的 **API** 端点（如果适用）。不要在此字段中指定 **pipelines** 命名空间。

ii.

在 **Public Data Science Pipelines API Endpoint** 字段中，更新数据科学项目管道的 **API** 端点（如果适用）。

iii.

可选：在 **Data Science Pipelines User Namespace** 字段中，更新相关用户命名空间以运行管道（如果适用）。

iv.

在 **Authentication Type** 列表中，选择验证管道所需的新身份验证类型（如果适用）。



重要

如果您直接从仪表板上的 **Jupyter** 标题创建了笔记本，请从 **Authentication Type** 列表中选择 **EXISTING_BEARER_TOKEN**。

- v. 在 **Data Science Pipelines API Endpoint Username** 字段中，更新身份验证类型所需的用户名（如果适用）。
- vi. 在 **Data Science Pipelines API Endpoint Password Or Token** 中，更新身份验证类型所需的密码或令牌（如果适用）。



重要

要获取 **Data Science Pipelines API** 端点令牌，在 **OpenShift Web** 控制台右上角，点您的用户名并选择 **Copy login command**。登录后，点 **Display token**，并复制来自 **Log in with this token** 命令的 `--token=` 的值。

- e. 定义 **S3** 兼容存储的连接信息：
 - i. 在 **Cloud Object Storage Endpoint** 字段中，更新 **S3** 兼容存储的端点（如果适用）。有关 **Amazon s3** 端点的更多信息，请参阅 [Amazon Simple Storage Service 端点和配额](#)。
 - ii. 可选：在 **Public Cloud Object Storage Endpoint** 字段中，更新 **S3** 兼容存储的 **URL**（如果适用）。
 - iii. 在 **Cloud Object Storage Bucket Name** 字段中，更新存储管道工件的存储桶的名称（如果适用）。如果存储桶名称不存在，则会自动创建。
 - iv. 从 **Cloud Object Storage Authentication Type** 列表中，更新访问 **S3** 兼容云存储所需的身份验证类型（如果适用）。如果使用 **AWS S3** 存储桶，则必须从列表中选择 **USER_CREDENTIALS**。
 - v. 可选：在 **Cloud Object Storage Credentials Secret** 字段中，更新包含存储用户名和密码的 **secret**（如果适用）。此 **secret** 在相关用户命名空间中定义。您必须在托管管道运行的集群中保存 **secret**。
 - vi. 可选：在 **Cloud Object Storage Username** 字段中，更新用户名以连接到与 **S3** 兼容的云存储（如果适用）。如果使用 **AWS S3** 存储桶，请更新 **AWS Secret Access Key ID**。

vii.

可选：在 **Cloud Object Storage Password** 字段中，更新密码以连接到与 S3 兼容云存储（如果适用）。如果使用 AWS S3 存储桶，请更新 **AWS Secret Access Key**。

f.

点 **Save & Close**。

验证

- 您更新的运行时配置显示在 JupyterLab 左侧边栏中的 **Runtimes** 选项卡()中。



3.5.5. 删除运行时配置

使用运行时配置完成后，您可以从 JupyterLab 接口中删除它。删除运行时配置后，您无法在 JupyterLab 中运行管道，直到创建另一个运行时配置为止。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了一个包含工作台的数据科学项目。
- 您已在包含工作台的数据 Science 项目中创建和配置了管道服务器。
- 之前创建的运行时配置在 JupyterLab 界面中可见。
- 您已从包含 Elyra 扩展（标准数据科学、TensorAI、PyAITor 或 HabanaAI）的笔记本镜像创建并启动 Jupyter 服务器。

流程

1. 在 JupyterLab 的左侧边栏中，点 **Runtimes** ()。
2. 将光标悬停在您要删除的运行时配置上，然后点 **Delete Item** 按钮()。

此时会出现一个对话框，提示您确认删除运行时配置。

3. 点击 **确定**。

验证

- 您删除的运行时配置不再显示在 JupyterLab 左侧边栏中的 **Runtimes** 选项卡()中。

3.5.6. 复制运行时配置



要防止您整个使用类似值重新创建运行时配置，您可以在 JupyterLab 界面中复制现有的运行时配置。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了一个包含工作台的数据科学项目。
- 您已在包含工作台的数据 Science 项目中创建和配置了管道服务器。

- 之前创建的运行时配置在 JupyterLab 界面中可见。
- 您已从包含 Elyra 扩展（标准数据科学、TensorAI、PyAITor 或 HabanaAI）的笔记本镜像创建并启动 Jupyter 服务器。

流程

1. 在 JupyterLab 的左侧边栏中，点 Runtimes ()。
2. 将光标悬停在您要重复的运行时配置上，并点 Duplicate 按钮()。

验证

- 您重复的运行时配置显示在 JupyterLab 左侧边栏中的 Runtimes 选项卡()中。

3.5.7. 在 JupyterLab 中运行管道


您可以从 Pipeline Editor 用户界面运行在 JupyterLab 中创建的管道。在运行管道前，您必须创建一个数据科学项目和管道服务器。创建管道服务器后，您必须在与管道服务器相同的项目中创建工作台。JupyterLab 中的管道实例必须包含运行时配置。如果您创建工作台作为数据科学项目的一部分，则会自动创建默认的运行时配置。但是，如果您在 OpenShift AI 仪表板中从 Jupyter 标题创建笔记本，则必须在 JupyterLab 中运行管道前创建一个运行时配置。运行时配置定义了管道实例的连接信息，以及 S3 兼容云存储。

先决条件

- 您已登陆到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。

- 您可以访问 **S3 兼容存储**。
- 您已在 **JupyterLab** 中创建管道。
- 您已在 **JupyterLab** 中的 **Pipeline Editor** 中打开管道。
- 管道实例包含运行时配置。
- 您已在包含工作台的 **data Science** 项目中创建和配置了管道服务器。
- 您已从包含 **Elyra** 扩展（标准数据科学、TensorAI、PyAITor 或 HabanaAI）的笔记本镜像创建并启动 **Jupyter** 服务器。

流程

1. 在 **Pipeline Editor** 用户界面中，点 **Run Pipeline** ()。

此时会出现 **Run Pipeline** 对话框。**Pipeline Name** 字段自动填充管道文件名。
2. 定义管道运行的设置。
 - a. 在 **Runtime Configuration** 列表中，选择要运行管道的相关运行时配置。
 - b. 可选：配置管道参数（如果适用）。如果您的管道包含引用管道参数的节点，您可以更改默认参数值。如果需要参数且没有默认值，则必须输入一个值。
3. 点击 **确定**。

验证

- 您可以查看管道运行的输出工件。工件存储在您指定的对象存储桶中。

3.5.8. 在 JupyterLab 中导出管道

您可以导出在 JupyterLab 中创建的管道。当您导出管道时，管道已准备好进行后续执行，但不会立即上传或执行。在导出过程中，任何软件包依赖项都会上传到兼容 S3 的存储中。另外，为目标运行时生成管道代码。

在导出管道前，您必须创建一个数据科学项目和管道服务器。创建管道服务器后，您必须在与管道服务器相同的项目中创建工作台。另外，JupyterLab 中的管道实例必须包含运行时配置。如果您创建工作台作为数据科学项目的一部分，则会自动创建默认的运行时代配置。但是，如果您从 OpenShift AI 仪表板中的 Jupyter 标题创建笔记本，则必须创建一个运行时配置，然后才能在 JupyterLab 中导出管道。运行时配置定义了管道实例的连接信息，以及 S3 兼容云存储。

先决条件

- 您已登录到 Red Hat OpenShift AI。
- 如果您使用专用的 OpenShift AI 组，则作为 OpenShift 中的用户组或 admin 组的一部分（例如，rh oai-users 或 rhoai-admins）。
- 您已创建了一个包含工作台的数据科学项目。
- 您已在包含工作台的数据 Science 项目中创建和配置了管道服务器。
- 您可以访问 S3 兼容存储。
- 您已在 JupyterLab 中创建了一个管道。
- 您已在 JupyterLab 中的 Pipeline Editor 中打开管道。
- 管道实例包含运行时配置。
-

您已从包含 Elyra 扩展（标准数据科学、TensorAI、PyAITor 或 HabanaAI）的笔记本镜像创建并启动 Jupyter 服务器。

流程

1. 在 Pipeline Editor 用户界面中，点 Export Pipeline (



)。

此时会出现 Export Pipeline 对话框。Pipeline Name 字段自动填充管道文件名。

2. 定义导出管道的设置。
 - a. 在 Runtime Configuration 列表中，选择要导出管道的相关运行时配置。
 - b. 在 Export Pipeline 中选择一个适当的文件格式
 - c. 在 Export Filename 字段中，为导出的管道输入文件名。
 - d. 选择 Replace if file already exists 复选框，以替换与您要导出的管道相同的现有文件。
 - e. 可选：配置管道参数（如果适用）。如果您的管道包含引用管道参数的节点，您可以更改默认参数值。如果需要参数且没有默认值，则必须输入一个值。
3. 点击 确定。

验证

- 您可以查看包含您在指定对象存储存储桶中导出的管道的文件。

3.6. 其他资源

- [Kubeflow Pipelines 2.0 文档](#)
- [在 JupyterLab 中使用管道。](#)