



Red Hat OpenShift AI Cloud Service 1

发行注记

与本发行版本相关的功能、增强功能、已解决的问题

Red Hat OpenShift AI Cloud Service 1 发行注记

与本发行版本相关的功能、增强功能、已解决的问题

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

本发行注记概述了此 Red Hat OpenShift AI 发行版本中的新功能、增强功能、解决的问题以及已知问题。OpenShift AI 目前在 Red Hat OpenShift Dedicated 和 Red Hat OpenShift Service on Amazon Web Services (ROSA) 中提供。

目录

第 1 章 OPENSIFT AI 概述	3
第 2 章 新功能及功能增强	4
2.1. 新功能	4
2.2. 功能增强	4
第 3 章 技术预览功能	5
第 4 章 开发人员预览功能	7
第 5 章 有限的可用性功能	8
第 6 章 支持删除	9
6.1. DATA SCIENCE PIPELINES V1 升级到 V2	9
6.2. 删除 BIAS 检测(TRUSTYAI)	9
6.3. 不再支持用于工作台的版本 1.2 笔记本镜像	9
6.4. NVIDIA GPU OPERATOR 替换 NVIDIA GPU 附加组件	9
6.5. KUBEFLOW NOTEBOOK CONTROLLER 替换 JUPYTERHUB	9
第 7 章 已解决的问题	11
第 8 章 已知问题	35
第 9 章 产品特性	63

第1章 OPENSIFT AI 概述

Red Hat OpenShift AI 是人工智能和机器学习(AI/ML)应用程序的数据科学家和开发人员的平台。

OpenShift AI 提供了一个环境，以便在内部或云中开发、培训、服务、测试和监控 AI/ML 模型和应用程序。

对于数据科学家，OpenShift AI 包括 Jupyter 和默认笔记本镜像集合，使用模型开发所需的工具和库优化，以及 TensorFlow 和 PyTorch 框架。部署并托管您的模型、将模型集成到外部应用程序中，并在任何混合云环境中导出模型以托管它们。您可以使用 Docker 容器构建带有数据科学项目的可移植机器学习(ML)工作流来增强 OpenShift AI 上的数据科学项目。您还可以使用图形处理单元(GPU)和 Habana Gaudi 设备来加快数据科学试验。

对于管理员，OpenShift AI 在现有 Red Hat OpenShift 或 ROSA 环境中启用数据科学工作负载。使用您现有的 OpenShift 身份提供程序来管理用户，并管理可供笔记本服务器使用的资源，以确保数据科学家具有创建、培训和主机模型所需的内容。使用加速器降低成本并允许数据科学家使用图形处理单元(GPU)和 Habana Gaudi 设备提高其端到端数据科学工作流的性能。

OpenShift AI 有两个部署选项：

- 您可以在内部或云中安装 **自我管理的软件**。您可以在自我管理的环境中（如 OpenShift Container Platform）或红帽管理的云环境（如 Red Hat OpenShift Dedicated (AWS 或 GCP 客户云订阅)中安装 OpenShift AI Self-Managed, Red Hat OpenShift Service on Amazon Web Services (ROSA Classic)或带有托管的 control plane (ROSA HCP)的 Red Hat OpenShift Service on Amazon Web Services (ROSA HCP)。有关在连接的或断开连接的环境中的 OpenShift 集群上作为自我管理的软件的信息，[请参阅 Red Hat OpenShift AI Self-Managed 产品文档](#)。
- **受管云服务**，作为 Red Hat OpenShift Dedicated 中的附加组件安装（具有 AWS 或 GCP 的客户云订阅）或 Red Hat OpenShift Service on Amazon Web Services (ROSA Classic)。有关 OpenShift AI Cloud Service 的详情，请参考 [Red Hat OpenShift AI 产品文档](#)。

有关 OpenShift AI 支持的软件平台、组件和依赖项的详情，请参考 [支持的配置](#)。

第 2 章 新功能及功能增强

本节介绍了 Red Hat OpenShift AI 中的新功能和增强。

2.1. 新功能

KServe 的 vLLM 运行时

单型号服务平台（使用 KServe）现在包括 vLLM，它是一个用于大型语言模型的高性能、专用运行时，使用名为 paged 的特殊方法来加快在某些加速器上的多个 LLM 的执行。

vLLM 运行时与 OpenAI API 兼容。vLLM 的初始版本需要 GPU，且没有针对 CPU 推测进行了优化。

如果您为单型号服务平台配置了监控，您也可以查看 vLLM 运行时指标。

改进了仪表板体验

此发行版本引入了一个新的主页，旨在方便地访问数据科学项目、学习资源以及主要解决方案区域概述，从而更轻松的开始使用 OpenShift AI。新的主页还为 OpenShift AI 管理员提供了对关键功能的访问权限，从而简化产品配置。之前的主页仍可从 **Applications > Enabled** 下的左侧导航窗格中访问。

2.2. 功能增强

删除内部镜像 registry 依赖项

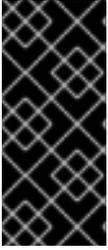
现在，您可以在 OpenShift AI 中创建和使用工作台，而无需启用内部 OpenShift 镜像 registry。如果更新集群以启用或禁用内部镜像 registry，则必须重新创建现有的工作台才能使 registry 生效。

Intel AI 工具集成的新工作台镜像

Intel AI Tools（以前称为 Intel oneAPI AI Analytics Toolkit）集成已通过三个新的工作台镜像进行了改进，其中包括用于流行框架和库（如 PyTorch 和 TensorFlow）的优化。这些优化功能提高了 Intel 硬件的性能，帮助数据科学家加快 Intel 架构上的端到端数据科学和分析管道。

您必须在集群中安装 Intel AI Tools Operator 才能选择新的工作台镜像。

第 3 章 技术预览功能



重要

本节论述了 Red Hat OpenShift AI 中的技术预览功能。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](#)。

RStudio 服务器笔记本镜像

使用 **RStudio 服务器** 笔记本镜像，您可以访问 RStudio IDE，这是 R 的集成开发环境。R 编程语言用于统计计算和图形来支持数据分析和预测。

要使用 **RStudio 服务器** 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](#)。



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 服务器通过 [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 服务器笔记本镜像

使用 **CUDA - RStudio 服务器** 笔记本镜像，您可以访问 RStudio IDE 和 NVIDIA CUDA Toolkit。RStudio IDE 是 R 编程语言用于统计计算和图形的集成开发环境。使用 NVIDIA CUDA 工具包，您可以使用 GPU 加速的库和优化工具提高工作。

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 **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](#)。



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 服务器通过 [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 服务器 笔记本镜像包含 NVIDIA CUDA 技术。CUDA 许可信息包括在 [CUDA 工具包文档](#) 中。在使用此示例工作台前，您应该查看其许可条款。

code-server workbench 镜像

Red Hat OpenShift AI 现在包括 **code-server** workbench 镜像。如需更多信息，[请参阅 GitHub 中的 code-server](#)。

使用 **code-server** 工作台镜像，您可以使用各种扩展来添加新的语言、主题、调试器并连接到其他服务来自定义工作台环境。您还可以提高数据科学工作的效率，语法突出显示、自动和括号匹配。



注意

基于 Elyra 的管道无法使用 **code-server** workbench 镜像。

code-server workbench 镜像目前在 Red Hat OpenShift AI 中作为技术预览功能提供。

第 4 章 开发人员预览功能



重要

本节介绍 Red Hat OpenShift AI 中的开发人员预览功能。

Developer Preview (开发人员预览) 功能不被红帽支持, 其功能可能并不完善且不是生产环境就绪。不要将开发人员预览功能用于生产环境或业务关键型工作负载。开发人员预览功能在红帽产品中包括早期对功能的访问。客户可以使用这些功能在开发过程中测试并提供反馈。开发人员预览功能可能没有任何文档, 可以随时更改或删除, 并且已获得有限的测试。红帽可能会提供在没有关联 SLA 的情况下提交对开发人员预览功能的反馈。

有关红帽开发人员预览功能的支持范围的更多信息, 请参阅 [开发人员预览支持范围](#)。

KServe Modelcars

OpenShift AI 的 KServe 组件包括 Modelcars 作为开发者技术预览功能。Modelcars 功能通过使用包含模型数据的开放容器项目(OCI)镜像来简化模型获取。这个行为可能会减少大型模型的启动时间, 减少磁盘空间的使用, 并提高性能。

默认情况下不启用 Modelcars 功能。您必须修改 KServe 配置才能使用该功能。

如需更多信息, 请参阅 KServe 文档中的 [OCI 镜像的 Serving 模型](#), 以及 [增强 KServe Model Fetching with Modelcars](#) 设计文档中的 Serving 模型。

支持 Kueue 中的 AppWrapper

Kue 中的 AppWrapper 支持作为开发者技术预览功能提供。实验性 API 支持使用带有分布式工作负载功能的基于 AppWrapper 的工作负载。

第 5 章 有限的可用性功能



重要

本节介绍了 Red Hat OpenShift AI 中的有限可用性功能。有限的可用性意味着您只能通过特定红帽批准的方式安装和接收对该功能的支持。如果没有这样的批准，则不支持该功能。这适用于本节中描述的所有功能。

在 OpenShift AI 中调优

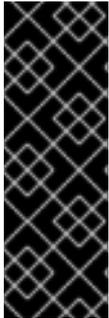
OpenShift AI 中的调优作为有限可用性功能提供。Kubeflow Training Operator 和 Hugging Face Supervised Fine-tuning Trainer (SFT Trainer) 允许用户在分布式环境中轻松微调和培训其模型。在本发行版本中，您可以将此功能用于基于 PyTorch 机器学习框架的型号。

第 6 章 支持删除

本节介绍了对 Red Hat OpenShift AI 中面向用户的功能的支持的主要变化。有关 OpenShift AI 支持的软件平台、组件和依赖项的详情，请参考 [支持的配置](#)。

6.1. DATA SCIENCE PIPELINES V1 升级到 V2

在以前的版本中，OpenShift AI 中的数据科学管道基于 KubeFlow Pipelines v1。数据科学管道现在基于 KubeFlow Pipelines v2，它使用不同的工作流引擎。在 OpenShift AI 中默认启用并部署了 Data Science Pipelines (DSP) 2.0。从仪表板中，无法再部署、查看或编辑基于 DSP 1.0 的管道详情。如需更多信息，请[参阅启用数据科学管道 2.0](#)。



重要

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，请确保集群中没有现有的 Argo 工作流安装。

如果要在升级 OpenShift AI 后使用带有 DSP 2.0 的现有管道和工作台，您必须更新工作台以使用 2024.1 笔记本镜像版本，然后手动将管道从 DSP 1.0 迁移到 2.0。如需更多信息，请[参阅升级到 DSP 2.0](#)。

6.2. 删除 BIAS 检测(TrustyAI)

从 OpenShift AI 2.7 开始，bias 检测(TrustyAI)功能已被删除。如果您之前启用了这个功能，升级到 OpenShift AI 2.7 或更高版本会删除该功能。默认 TrustyAI 笔记本镜像仍被支持。

6.3. 不再支持用于工作台的版本 1.2 笔记本镜像

在创建工作台时，您可以指定一个与工作台搭配使用的笔记本镜像。从 OpenShift AI 2.5 开始，当您创建新的工作台时，无法选择版本 1.2 笔记本容器镜像。已使用版本 1.2 笔记本镜像运行的工作台继续正常工作。但是，红帽建议您更新工作台以使用最新的笔记本容器镜像。

6.4. NVIDIA GPU OPERATOR 替换 NVIDIA GPU 附加组件

在以前的版本中，要启用图形处理单元(GPU)来帮助计算密集型工作负载，您要安装 NVIDIA GPU 附加组件。OpenShift AI 不再支持此附加组件。

现在，要启用 GPU 支持，您必须安装 NVIDIA GPU Operator。要了解如何安装 GPU Operator，请[参阅 Red Hat OpenShift Container Platform \(external\)上的 NVIDIA GPU Operator](#)。

6.5. KUBEFLOW NOTEBOOK CONTROLLER 替换 JUPYTERHUB

在 OpenShift AI 1.15 及更早版本中，JupyterHub 用于创建和启动笔记本服务器环境。在 OpenShift AI 1.16 及更高版本中，JupyterHub 不再包含，其功能由 KubeFlow Notebook Controller 替代。

这个更改有以下优点：

- 用户现在可以立即取消请求，进行更改并重试请求，而不是等待 5+ 分钟，以便初始请求超时。这意味着，当请求失败时，用户不会再等待，例如当笔记本服务器没有正确启动时。
- 架构不再阻止单个用户有多个笔记本服务器会话，从而扩展了将来的功能可能。

- 删除 PostgreSQL 数据库要求允许在 OpenShift AI 中未来的扩展环境支持。

但是，此次更新还会创建以下行为更改：

- 对于 IT Operations 管理员，笔记本服务器管理界面目前还不允许登录数据科学家用户笔记本服务器。计划在以后的版本中添加。
- 对于数据科学家，JupyterHub 接口 URL 不再有效。更新书签以指向 OpenShift AI 仪表板。

JupyterLab 接口保持不变，数据科学家可以继续使用 JupyterLab 与笔记本文件一起工作。

第 7 章 已解决的问题

在 Red Hat OpenShift AI 中解决了以下显著的问题。

有关更新的完整列表，请参阅 [RHBA-2024:4069](#) 公告。

rhoAIENG-7312 - 模型服务在带有 KServe 中的令牌身份验证查询过程中失败

在以前的版本中，如果您在 **DataScienceCluster** 对象中同时启用了 ModelMesh 和 KServe 组件，并添加了 Authorino 作为授权供应商，则可能会出现导致 **odh-model-controller** pod 处于 ModelMesh 状态但不适用于 KServe 和 Authorino 的竞争条件。在这种情况下，如果您向使用 KServe 部署的运行模型发出推测请求，您会看到 **404 - Not Found** 错误。另外，**odh-model-controller** 部署对象的日志会显示 **Reconciler** 错误消息。这个问题现已解决。

RHOAIENG-7181 (以前称为 **RHOAIENG-6343**) - 安装 OpenShift AI 后一些组件被设置为 **Removed**

在以前的版本中，安装 OpenShift AI 后，**codeflare** 的 **managementState** 字段、**kueue** 和 **ray** 组件被错误地设置为 **Removed**，在 **DataScienceCluster** 自定义资源中被错误地设置为 **Removed**。这个问题现已解决。

RHOAIENG-7079 (以前称为 **RHOAIENG-6317**) - Pipeline 任务状态和日志有时不会在 OpenShift AI 仪表板中显示

在以前的版本中，当使用 **Elyra** 运行管道时，**OpenShift AI** 仪表板可能无法显示管道任务状态和日志，即使相关的 pod 没有被修剪，且信息仍在 **OpenShift** 控制台中可用。这个问题现已解决。

RHOAIENG-7070 (以前称为 **RHOAIENG-6709**) - 当指定不同的环境变量时，Jupyterbook 创建可能会失败

在以前的版本中，如果您启动并停止 **Jupyter** 笔记本，并在 **OpenShift AI workbench** 中编辑其环境变量，笔记本无法重启。这个问题现已解决。

rhoAIENG-6853 - 在 **Elyra** 管道 pod 中无法设置 pod 容限

在以前的版本中，如果您为 **Elyra pipeline pod** 设置 pod 容限，则容限不会生效。这个问题现已解决。

RHOAIENG-5314 - 由于网络策略，数据科学管道服务器无法在全新的集群中部署

在以前的版本中，如果您在新集群中创建了数据科学管道服务器，用户界面会保持在加载状态，且管道服务器没有启动。这个问题现已解决。

RHOAIENG-4252 - 数据科学管道服务器删除过程无法删除 ScheduledWorkflow 资源

在以前的版本中，管道服务器删除过程不会删除 ScheduledWorkflow 资源。因此，新的 DataSciencePipelinesApplications (DSPAs)无法识别冗余的 ScheduledWorkflow 资源。这个问题现已解决

RHOAIENG-3411 (以前记录的为 **RHOAIENG-3378**) - 内部镜像 Registry 是 Jupyter 笔记本生成的未决定性依赖项

在以前的版本中，在启动 OpenShift AI 笔记本和工作台前，您必须在 OpenShift 中启用了内部集成的容器镜像 registry。在不首先启用镜像 registry 时尝试启动笔记本或工作台会失败，并显示 "InvalidImageName" 错误。现在，您可以在 OpenShift AI 中创建和使用工作台，而无需启用内部 OpenShift 镜像 registry。如果更新集群以启用或禁用内部镜像 registry，则必须重新创建现有的工作台才能使 registry 生效。

RHOAIENG-2541 - KServe 控制器 pod 遇到 OOM，因为集群中的 secret 数量太多

在以前的版本中，如果您的 OpenShift 集群有大量 secret，KServe 控制器 pod 可能会因为内存不足 (OOM)错误而持续崩溃。这个问题现已解决。

RHOAIENG-1452 - Red Hat OpenShift AI Add-on 会卡住

在以前的版本中，当安装通过 OCM API 触发时，Red Hat OpenShift AI Add-on 卸载不会删除 OpenShift AI 组件。这个问题现已解决。

RHOAIENG-307 - 删除 DataScienceCluster 会删除所有 OpenShift Serverless CR

在以前的版本中，如果您删除了 DataScienceCluster 自定义资源(CR)，则所有 OpenShift Serverless CR (包括 knative-serving、Deployment、gateway 和 pod) 也会被删除。这个问题现已解决。

RHOAIENG-6709 - 当指定不同的环境变量时，Jupyter 笔记本创建可能会失败

在以前的版本中，如果您启动并停止 Jupyter 笔记本，并在 OpenShift AI workbench 中编辑其环境变量，笔记本无法重启。这个问题现已解决。

RHOAIENG-6701 - 没有集群管理员特权的用户无法访问 Ray 仪表板的作业提交端点

在以前的版本中，没有 OpenShift 的集群管理员特权的分布式工作负载功能用户可能无法访问或使用 Ray 仪表板的作业提交端点。这个问题现已解决。

RHOAIENG-6578 - 在默认情况下没有令牌到受保护的 inference 点的请求

在以前的版本中，如果您将 Authorino 添加为单型号服务平台的授权供应商，并且为部署的模型启用了令牌授权，那么仍然可以在没有指定令牌的情况下查询模型。这个问题现已解决。

RHOAIENG-6343 - 安装 OpenShift AI 后，一些组件被设置为 Removed

在以前的版本中，安装 OpenShift AI 后，codeflare 的 managementState 字段、kueue 和 ray 组件被错误地设置为 Removed，在 DataScienceCluster 自定义资源中被错误地设置为 Removed。这个问题现已解决。

RHOAIENG-5067 - 基于 ModelMesh 组件的模型服务器加载模型服务器

在以前的版本中，包含大写字母或空格的数据科学项目名称可能会导致基于 ModelMesh 组件的模型服务器指标页面出现问题。指标页面可能无法正确接收数据，从而导致 400 Bad Request 错误，并阻止页面加载。这个问题现已解决。

rhoAIENG-4966 - 自定义 CA 捆绑包中的自签名证书可能缺少在 odh-trusted-ca-bundle 配置映射中

在以前的版本中，如果您添加了自定义证书颁发机构(CA)捆绑包以使用自签名证书，有时 odh-trusted-ca-bundle ConfigMap 中缺少自定义证书，或者在 ConfigMap 设置为受管时不包含 odh-trusted-ca-bundle ConfigMap。这个问题现已解决。

RHOAIENG-4938 (之前记录的为 **RHOAIENG-4327**) - Workbenches 不使用集中配置的捆绑包中的自签名证书

在 OpenShift AI、`ca-bundle.crt` 和 `odh-ca-bundle.crt` 中包含自签名证书有两个捆绑包选项。在以前的版本中，工作台不会自动使用来自集中配置的捆绑包的自签名证书，您必须定义指向证书路径的环境变量。这个问题现已解决。

[RHOAIENG-4572](#) - 在某些情况下无法安装和升级后运行数据科学管道

在以前的版本中，在以下情况下，在安装或升级 OpenShift AI 后您无法运行数据科学管道：

- 已安装 OpenShift AI，并且具有有效的 CA 证书。在 `default-dsci` 对象中，您将 `trustedCABundle` 字段的 `managementState` 字段更改为 `Removed` 安装后。
- 您已将 OpenShift AI 从 2.6 升级到 2.8 版本，并且具有有效的 CA 证书。
- 您已将 OpenShift AI 从 2.7 升级到 2.8 版本，并且具有有效的 CA 证书。

这个问题现已解决。

[RHOAIENG-4524](#) - RStudio 镜像的 BuildConfig 定义包含错误的分支

在以前的版本中，RStudio 和 CUDA - RStudio 工作台镜像的 BuildConfig 定义指向 OpenShift AI 中错误的分支。这个问题现已解决。

[rhoAIENG-3963](#) - Unnecessary 受管资源警告

在以前的版本中，当编辑并保存 `redhat-ods-applications` 项目的 `OdhDashboardConfig` 自定义资源时，系统会错误地显示 `Managed 资源` 警告消息。这个问题现已解决。

[rhoAIENG-2542](#) - Inference 服务 pod 并不总是获得 Istio sidecar

在以前的版本中，当使用单模式服务平台（它使用 KServe）部署模型时，生成的 pod 中可能会缺少 `istio-proxy` 容器，即使 `inference` 服务具有 `sidecar.istio.io/inject=true` 注解。这个问题现已解决。

RHOAIENG-1666 - 导入管道按钮被预先访问

在以前的版本中，当您将在管道导入到属于数据科学项目的工作台时，可以在管道服务器完全可用前访问 **Import Pipeline** 按钮。这个问题现已解决。

RHOAIENG-673（之前记录的为 **RHODS-12946**）- 无法在断开连接的环境中从 PyPI 镜像安装或使用私有证书

在断开连接的环境中，Red Hat OpenShift AI 无法连接到面向公共的 PyPI 存储库，因此您必须在网络中指定存储库。在以前的版本中，如果您使用私有 TLS 证书，且数据科学管道被配置为安装 Python 软件包，则管道运行会失败。这个问题现已解决。

RHOAIENG-3355 - KServe 上的 OVMS 无法正确使用加速器

在以前的版本中，当使用单模式服务平台部署模型并选择了 **OpenVINO Model Server serving** 运行时，如果您请求了加速器来附加到模型服务器，则会检测到加速器硬件，但在响应查询时不会被模型使用。这个问题现已解决。

RHOAIENG-2869 - 无法编辑现有模型框架和模型路径

在以前的版本中，当您尝试使用 **Deploy model** 对话框编辑多型号项目中的模型时，模型框架和路径值不会更新。这个问题现已解决。

RHOAIENG-2724 - 型号部署失败，因为字段会在对话框中自动重置

在以前的版本中，当部署模型或编辑部署模型时，"**Deploy model**"对话框中的 **Model servers** 和 **Model 框架** 字段可能会重置为默认状态。**Deploy** 按钮可能会保持启用状态，即使这些强制字段不再包含有效的值。这个问题现已解决。

RHOAIENG-2099 - 数据科学管道服务器无法在新集群中部署

在以前的版本中，当您在新集群中创建数据科学项目管道服务器时，用户界面会保持在加载状态，管道服务器不会启动。这个问题现已解决。

RHOAIENG-1199 (之前记录的为 **ODH-DASHBOARD-1928**) - 自定义服务运行时创建错误消息不方便

在以前的版本中，当尝试创建或编辑自定义 **model-serving** 运行时并出现错误时，错误消息没有指定错误。改进了错误消息。

RHOAIENG-556 - 无论错误是什么，都为 **KServe** 模型创建 **ServingRuntime**

在以前的版本中，当您试图部署 **KServe** 模型并出现错误时，**InferenceService** 自定义资源(CR)仍然被创建，且在 **Data Science Projects** 页面中显示模型，但状态始终会保持未知。**KServe** 部署过程已更新，以便在出现错误时不会创建 **ServingRuntime**。

RHOAIENG-548 (之前记录的为 **ODH-DASHBOARD-1776**) - 用户没有项目管理员权限时的错误消息

在以前的版本中，如果您没有项目的管理员权限，则无法访问一些功能，且错误消息没有解释原因。例如，当您在只能访问单个命名空间的环境中创建了模型服务器时，会出现 **Error create model server** 错误消息。但是，模型服务器仍然成功创建。这个问题现已解决。

RHOAIENG-66 - 由 **CodeFlare SDK** 部署 **Ray** 仪表盘路由公开自签名证书，而不是集群证书

在以前的版本中，当您使用带有 **openshift_oauth=True** 选项的 **CodeFlare SDK** 部署 **Ray** 集群时，**Ray** 集群生成的路由会使用 **passthrough** 方法进行保护，因此会公开 **OAuth** 代理使用的自签名证书。这个问题现已解决。

RHOAIENG-12 - 从一些浏览器无法访问 **Ray** 仪表盘

在某些浏览器中，分布式工作负载功能的用户可能无法访问 **Ray** 仪表盘，因为浏览器会自动将仪表盘 URL 的前缀从 **http** 改为 **https**。这个问题现已解决。

rhoDS-6216 - **ModelMesh oauth-proxy** 容器不稳定

在以前的版本中，由于 **ModelMesh oauth-proxy** 容器失败，**ModelMesh pod** 无法正确部署。此问题会间歇性发生，只有在 **ModelMesh** 运行时环境中启用了身份验证时才发生。这个问题现已解决。

RHOAIENG-535 - 如果没有 HTTP 请求，显示已部署模型的 HTTP 请求的指标图不正确

在以前的版本中，如果部署的模型没有为每个数据类型（成功和失败）都至少收到一个 HTTP 请求，则显示 HTTP 请求性能指标（用于模型服务器或特定模型的所有模型）的图表错误地呈现，直接代表失败请求的数量。这个问题现已解决。

RHOAIENG-1467 - Serverless net-istio 控制器 pod 可能达到 OOM

在以前的版本中，Knative net-istio-controller pod（这是 KServe 的依赖项）可能会因为内存不足 (OOM) 错误而持续崩溃。这个问题现已解决。

RHOAIENG-1899（之前记录的为 **RHODS-6539**）- Anaconda 专业版无法验证并启用

在以前的版本中，您无法启用 Anaconda 专业版，因为仪表板的密钥验证是不可操作的。这个问题现已解决。

RHOAIENG-2269 -(Single-model) Dashboard 无法显示正确的模型副本数

在以前的版本中，在单型号服务平台上，数据科学项目的 Models 和 model servers 部分没有显示正确的模型副本数量。这个问题现已解决。

RHOAIENG-2270 -(Single-model)用户无法更新模型部署设置

在以前的版本中，您无法编辑使用单模式服务平台部署的模型的部署设置（如副本数）。这个问题现已解决。

rhoDS-8865 - 管道服务器无法启动，除非您指定了 Amazon Web Services (AWS) Simple Storage Service (S3) 存储桶资源

在以前的版本中，当您为数据科学项目创建数据连接时，AWS_S3_BUCKET 字段不会被指定为必需的字段。但是，如果您试图使用 AWS_S3_BUCKET 字段没有填充的数据连接配置管道服务器，管道服务器将无法成功启动。这个问题现已解决。Configure pipeline server 对话框已更新，将 Bucket 字段包含为必填字段。

rhoDS-12899 - OpenVINO 运行时缺少 NVIDIA GPU 注解

在以前的版本中，如果用户选择了 **OpenVINO 模型服务器（支持 GPU）** 运行时并在模型服务器用户界面中选择了 **NVIDIA GPU 加速器**，则系统可能会显示与所选运行时不兼容的不必要的警告。警告将不再显示。

RHOAIENG-84 - 无法使用带有 KServe 的自签名证书

在以前的版本中，单模式服务平台不支持自签名证书。这个问题现已解决。要将自签名证书与 KServe 搭配使用，请按照使用 [证书](#) 中描述的步骤操作。

RHOAIENG-164 - 从仪表板中无法正确应用 Kserve 的模型服务器副本数

在以前的版本中，当您设置与默认的(1)不同的多个模型服务器副本时，模型(server)仍然使用 1 个副本部署。这个问题现已解决。

RHOAIENG-288 - 有两个版本显示了工作台的建议镜像版本标签

OpenShift AI 中提供的大多数工作台镜像都在多个版本中都提供了。唯一推荐的版本是最新版本。在 Red Hat OpenShift AI 2.4 和 2.5 中，对于镜像的多个版本，推荐的标签被错误地显示。这个问题现已解决。

RHOAIENG-293 - 从 2.4 升级到 2.5 后不会删除已弃用的 ModelMesh 监控堆栈

在 Red Hat OpenShift AI 2.5 中，以前的 ModelMesh 监控堆栈不再被用户工作负载监控替代。但是，在升级到 OpenShift AI 2.5 的过程中不会删除以前的监控堆栈。有些组件保留并使用集群资源。这个问题现已解决。

RHOAIENG-343 - OpenShift Service Mesh 的手动配置，OpenShift Serverless 无法用于 KServe

如果安装了 OpenShift Serverless 和 OpenShift Service Mesh，然后安装了启用了 KServe 的 Red Hat OpenShift AI，则不会部署 KServe。这个问题现已解决。

RHOAIENG-517 - 具有编辑权限的用户无法看到创建模型

具有编辑权限的用户无法看到任何创建的模型，除非他们是项目所有者或具有项目的 **admin** 权限。这个问题现已解决。

RHOAIENG-804 - 无法在启用了 FIPS 的集群上使用 KServe 部署大型语言模型

在以前的版本中，Red Hat OpenShift AI 还没有完全为 FIPS 设计。您不能在启用了 FIPS 的集群中使用 KServe 部署大语言模型(LLMs)。这个问题现已解决。

rhOAIENG-908 - 如果之前启用了 KServe，则无法使用 ModelMesh，然后删除

在以前的版本中，当在 DataScienceCluster 对象中启用 ModelMesh 和 KServe 时，您随后删除了 KServe，则无法使用 ModelMesh 部署新模型。您可以继续使用之前使用 ModelMesh 部署的模型。这个问题现已解决。

RHOAIENG-2184 - 无法创建 Ray 集群或分布式工作负载

在以前的版本中，用户无法在具有 **admin** 或 **edit** 权限的命名空间中创建 Ray 集群或分布式工作负载。这个问题现已解决。

ODH-DASHBOARD-1991 - ovms-gpu-ootb 缺少推荐的加速器注解

在以前的版本中，当您在项目中添加模型服务器时，**Serving** 运行时 列表不会显示 NVIDIA GPU 的推荐服务运行时 标签。这个问题现已解决。

RHOAIENG-807 - 重启工作台时删除加速器配置集容限

在以前的版本中，如果您创建了一个工作台，它使用一个包括容限的加速器配置集，重启工作台会删除容限信息，这意味着重启无法完成。新创建的启用了 GPU 的工作台可能会首次启动，但永远不会成功重启，因为生成的 pod 会一直处于待处理状态。这个问题现已解决。

DATA-SCIENCE-PIPELINES-OPERATOR-294 - Scheduled 管道运行（使用 **data-passing**）可能无法在步骤间传递数据，或者完全失败步骤

使用 **S3** 对象存储来存储管道工件的调度管道运行可能会失败，并显示以下错误：

```
Bad value for --endpoint-url "cp": scheme is missing. Must be of the form http://<hostname>/ or https://<hostname>/
```

出现这个问题的原因是 **S3** 对象存储端点没有成功传递给调度的管道运行的 **pod**。这个问题现已解决。

RHODS-4769 - 具有不支持污点的节点上的 GPU 无法分配给笔记本服务器

在创建笔记本服务器时，无法选择带有除 **supported nvidia.com/gpu** 污点的节点上的 **GPU**。这个问题现已解决。

RHODS-6346 - 使用无效字符创建数据科学项目时会显示 Unclear 错误消息

当使用无效特殊字符创建数据科学项目的数据连接、工作台或存储连接时，会显示以下出错信息：

```
the object provided is unrecognized (must be of type Secret): couldn't get version/kind; json parse error: unexpected end of JSON input ({"apiVersion":"v1","kind":"Sec ...)
```

错误消息未能明确指示问题。错误消息现在表示输入无效字符。

RHODS-6950 - 使用集群中的所有 GPU 时无法缩减工作台 GPU

在早期版本中，如果使用集群中的所有 **GPU**，则无法缩减工作台 **GPU**。这个问题适用于一个工作台使用的 **GPU**，以及由多个工作台使用的 **GPU**。现在，您可以通过从 **Accelerators** 列表中选择 **None** 来缩减 **GPU**。

rhODS-8939 - 上一发行版本中创建的 Jupyter 笔记本的默认共享内存会导致运行时错误

从版本 **1.31** 开始，这个问题已被解决，任何新笔记本的共享内存被设置为节点的大小。

对于在早于 1.31 的发行版本中创建的 Jupyter 笔记本，Jupyter 笔记本的默认共享内存被设置为 64 MB，您无法在笔记本配置中更改此默认值。

要解决这个问题，您必须重新创建笔记本，或按照知识库文章 [如何更改 Red Hat OpenShift AI 中的 Jupyter 笔记本的共享内存](#)。

[rhoDS-9030](#) - 删除 kfdefs 资源时 OpenShift AI 的卸载过程可能会卡住

卸载 OpenShift AI 管理的服务的步骤请参考 [卸载 OpenShift AI](#)。

但是，即使您遵循本指南，您可能已经看到卸载过程没有成功完成。相反，进程会保留在删除 Kubeflow Operator 使用的 kfdefs 资源的步骤。如以下示例所示，kfdefs 资源可能存在于 redhat-ods-applications, redhat-ods-monitoring, 和 rhods-notebooks 命名空间中：

```
$ oc get kfdefs.kfdef.apps.kubeflow.org -A
```

NAMESPACE	NAME	AGE
redhat-ods-applications	rhods-anaconda	3h6m
redhat-ods-applications	rhods-dashboard	3h6m
redhat-ods-applications	rhods-data-science-pipelines-operator	3h6m
redhat-ods-applications	rhods-model-mesh	3h6m
redhat-ods-applications	rhods-nbc	3h6m
redhat-ods-applications	rhods-osd-config	3h6m
redhat-ods-monitoring	modelmesh-monitoring	3h6m
redhat-ods-monitoring	monitoring	3h6m
rhods-notebooks	rhods-notebooks	3h6m
rhods-notebooks	rhods-osd-config	3h5m

删除 kfdefs 资源失败可能会阻止以后安装 OpenShift AI。这个问题不再发生。

[rhoDS-9764](#) - 编辑工作台时数据连接详情会被重置

当您编辑了具有现有数据连接的工作台，然后选择 **Create new data connection** 选项时，编辑页面可能会在指定新的连接详情前恢复到 **Use existing data connection** 选项。这个问题现已解决。

[rhoDS-9583](#) - Data Science 仪表板没有检测到现有的 OpenShift Pipelines 安装

当 OpenShift Pipelines Operator 作为全局 Operator 安装在集群中时，OpenShift AI 仪表板不会检测到它。OpenShift Pipelines Operator 现在可以成功检测到。

[ODH-DASHBOARD-1639](#) - 在仪表板路由中 Wrong TLS 值

在以前的版本中，当在 OpenShift 上为 OpenShift AI 仪表板创建路由时，`tls.termination` 字段具有无效的默认值 `Reencrypt`。这个问题现已解决。新值为 `reencrypt`。

[ODH-DASHBOARD-1638](#) - Triggered Runs 选项卡中的 Name placeholder shows Scheduled run name

在以前的版本中，当点 `Pipelines > Runs`，然后选择 `Triggered` 选项卡来配置触发的运行时，`Name` 字段中显示的示例值为 `Scheduled run` 名称。这个问题现已解决。

[ODH-DASHBOARD-1547](#) - "We couldn't find that page" 消息在后台安装管道 operator 时显示在仪表板中

在以前的版本中，当使用仪表板的 `Data Science Pipelines` 页面安装 OpenShift Pipelines Operator 时，当 Operator 安装完成后，刷新的页面以显示 我们无法找到该页面 信息。这个问题现已解决。Operator 安装完成后，仪表板会将您重定向到 `Pipelines` 页面，您可以在其中创建管道服务器。

[ODH-DASHBOARD-1545](#) - 当 Models 选项卡扩展时，仪表板会保持滚动到项目底部

在以前的版本中，在仪表板的 `Data Science Projects` 页面中，如果您点击 `Deployed models` 选项卡来扩展它，然后尝试对页面执行其他操作，页面会自动滚动到 `Deployed models` 部分。这会影响到您执行其他操作的能力。这个问题现已解决。

[注意BOOKS-156](#) - Elyra 包括一个称为 Test 的示例运行时

在以前的版本中，Elyra 包括一个示例运行时配置，称为 `Test`。如果在运行数据科学项目时选择了此配置，您可能会看到错误。`Test` 配置现已被删除。

[rhODS-9622](#) - Duplicating a scheduled 管道运行不会复制现有的 period 和 pipeline 输入参数值

在以前的版本中，当复制带有定期触发器的调度的管道运行时，重复过程不会为重复运行或指定的管道输入参数复制配置的执行频率。这个问题现已解决。

rhoDS-8932 - 在调度重复管道运行时默认会显示不正确的 cron 格式

当您通过配置 cron 任务调度周期性管道运行时，OpenShift AI 接口默认显示不正确的格式。它现在显示正确的格式。

RHODS-9374 - 带有非唯一名称的管道不会出现在数据科学项目用户界面中

如果您从支持 Elyra 的 Jupyter 应用程序启动笔记本，或者提交一个工作台时，带有非唯一名称的管道不会出现在相关数据科学项目页面的 Pipelines 部分或 data Science 管道标题的 Pipelines 标题中。这个问题现已解决。

RHODS-9329 - 部署自定义 model-serving 运行时可能会导致错误消息

在以前的版本中，如果您使用 OpenShift AI 仪表板部署自定义 model-serving 运行时，部署过程可能会失败，并显示 Error retrieve Serving Runtime 信息。这个问题现已解决。

rhODS-9064 - 升级后，OpenShift AI 仪表板中没有启用 Data Science Pipelines 选项卡

当您从 OpenShift AI 1.26 升级到 OpenShift AI 1.28 时，OpenShift AI 仪表板中没有启用 Data Science Pipelines 选项卡。这个问题已在 OpenShift AI 1.29 中解决。

RHODS-9443 - 导出 Elyra 管道以纯文本形式公开 S3 存储凭证

在 OpenShift AI 1.28.0 中，当您以 Python DSL 格式或 YAML 格式从 JupyterLab 导出 Elyra 管道时，生成的输出以纯文本形式包含 S3 存储凭证。这个问题已在 OpenShift AI 1.28.1 中解决。但是，在升级到 OpenShift AI 1.28.1 后，如果您的部署包含带有管道服务器和数据连接的数据科学项目，您必须执行以下附加操作才能使修复生效：

1. 刷新浏览器页面。

2. 停止部署中运行的工作台，然后重新启动它们。

另外，要确认您的 Elyra 运行时配置包含该修复，请执行以下操作：

1. 在 JupyterLab 的左侧边栏中，点 Runtimes ()。
)。
2. 将光标悬停在您要查看的运行时配置上，并点击 Edit 按钮()。
)。

Data Science Pipelines 运行时配置页面将打开。
3. 确认 KUBERNETES_SECRET 定义为 Cloud Object Storage Authentication Type 字段中的值。
4. 关闭运行时配置而不更改它。

RHODS-8460 - 编辑共享项目详情时，用户界面会一直处于加载状态，而不会报告错误

当具有编辑项目权限的用户试图编辑其详情时，用户界面会保持在加载状态，且没有显示适当的错误消息。具有编辑项目权限的用户无法编辑项目中的任何字段，如描述。这些用户只能编辑属于项目的组件，如其工作台、数据连接和存储。

用户界面现在显示适当的错误消息，且不会尝试更新项目描述。

rhoDS-8482 - Data Science pipeline 图没有显示运行管道的节点边缘

如果您在其 YAML 代码中运行不包含 Tekton 格式的参数或 when 表达式的管道，OpenShift AI 用户界面不会显示与图形节点的连接边缘。例如，如果您使用包含 runAfter 属性或 Workspaces 的管道，用

户界面在没有边缘连接的情况下显示所执行管道的图形。OpenShift AI 用户界面现在显示连接到图形节点的边缘。

RHODS-8923 - 当您试图创建管道服务器时不会检测到新创建的数据连接

如果您在 Data Science 项目中创建了数据连接，然后尝试创建管道服务器，则 **Configure a pipeline server** 对话框不会检测到您创建的数据连接。这个问题现已解决。

rhoDS-8461 - 当与其他用户共享项目时，OpenShift AI 用户界面文本是误导

当您试图与另一个用户共享 Data Science 项目时，用户界面文本会错误地表示用户可以编辑其所有详情，如描述。但是，用户只能编辑属于某一项目的组件，如其工作台、数据连接和存储。这个问题现已解决，用户界面文本不再有误导，这意味着用户可以编辑所有详情。

rhoDS-8462 - 具有 "Edit" 权限的用户无法创建模型服务器

具有 "Edit" 权限的用户现在可以创建一个没有令牌授权的 Model Server。用户必须具有 "Admin" 权限，才能创建具有令牌授权的 Model Server。

rhODS-8796 - OpenVINO Model Server 运行时没有强制 GPU 使用所需的标记

OpenShift AI 默认包括 OpenVINO Model Server (OVMS) 模型运行时。当您配置新的模型服务器并选择这个运行时，配置模型服务器 对话框可让您指定要与模型服务器一起使用的 GPU 数。但是，当您完成配置模型服务器并从中部署模型时，模型服务器实际上没有使用任何 GPU。这个问题现已解决，模型服务器使用 GPU。

rhoDS-8861 - 在创建管道运行时更改主机项目会导致可用管道列表不准确

如果您在创建管道运行时更改了主机项目，接口将无法使新主机项目的管道可用。相反，接口会显示属于您最初在 Data Science Pipelines > Runs 页面中选择的项目的管道。这个问题现已解决。您不再从 Create run 页面中选择一个管道。当您点 Create run 按钮（基于当前项目及其管道）时，管道选择会自动更新。

RHODS-8249 - 作为 ConfigMap 上传的环境变量被存储在 Secret 中

在以前的版本中，在 OpenShift AI 界面中，当您通过上传 ConfigMap 配置将环境变量添加到工作台时，变量会改为存储在 Secret 对象中。这个问题现已解决。

RHODS-7975 - Workbenches 可以有多个数据连接

在以前的版本中，如果您更改了工作台的数据连接，则现有数据连接不会被释放。因此，工作台可能会保持连接到多个数据源。这个问题现已解决。

rhoDS-7948 - 上传包含环境变量的 secret 文件会导致双编码值

在以前的版本中，当在数据科学项目中创建工作台时，如果您上传了一个包含环境变量的基于 YAML 的 secret 文件，则环境变量值不会被解码。然后，在由此过程创建的生成的 OpenShift 机密中，编码的值会再次编码。这个问题现已解决。

RHODS-6429 - 使用 Intel OpenVINO 或 Anaconda 专业版镜像创建工作台时会显示一个错误

在以前的版本中，当使用 Intel OpenVINO 或 Anaconda 专业版镜像创建工作台时，创建过程中会出现一个错误。但是，工作台仍然成功创建。这个问题现已解决。

rhODS-6372 - Idle notebook culler 没有考虑活跃的终端

在以前的版本中，如果笔记本镜像有一个正在运行的终端，但没有活跃的、运行内核，空闲的 notebook culler 会将笔记本检测到为不活跃并停止终端。这个问题现已解决。

RHODS-5700 - 创建工作台时无法创建或连接到数据连接

在创建工作台时，用户无法创建新数据连接，或者连接到现有数据连接。

RHODS-6281 - 如果从集群中删除了 admin 组，则 OpenShift AI 管理员无法访问 Settings 页面

在以前的版本中，如果从集群中删除 Red Hat OpenShift AI 管理员组，OpenShift AI 管理员用户无法访问 OpenShift AI 仪表板中的 Settings 页面。特别是，可以看到以下行为：

- 当 OpenShift AI 管理员用户试图访问 **Settings** → **User management** 页面时，会出现 "Page Not Found" 错误。
- 集群管理员 不会丢失对 OpenShift AI 仪表板上的 **Settings** 页面的访问。当集群管理员访问 **Settings** → **User Management** 页面时，会出现警告消息，表示 OpenShift 中不再存在已删除的 OpenShift AI 管理员组。然后，删除的管理员组已从 **OdhDashboardConfig** 中删除，并且恢复了管理员访问权限。

这个问题现已解决。

rhODS-1968 - 删除的用户会一直登录，直到刷新仪表板为止

在以前的版本中，当用户撤销 Red Hat OpenShift AI 仪表板的权限时，用户只有在仪表板页面刷新后才会注意到更改。

这个问题现已解决。当用户的权限被撤销时，OpenShift AI 仪表板会在 30 秒内锁定用户，而无需刷新。

RHODS-6384 - 创建重复数据连接时，工作台数据连接被错误地更新

在创建包含与现有数据连接相同的数据连接时，数据连接创建会失败，但相关的工作台仍然会重启并连接到错误的数据连接。这个问题已解决。现在，工作台连接到正确的数据连接。

rhoDS-6370 - Workbenches 无法接收最新的容限

在以前的版本中，若要获取最新的容限，用户必须尝试编辑相关的工作台，不进行任何更改，然后再次保存工作台。现在，用户可以通过停止，然后重启其数据科学项目的工作台来应用最新的容限更改。

RHODS-6779 - 从 OpenShift AI 1.20 升级到 OpenShift AI 1.21 后模型无法提供

当从 OpenShift AI 1.20 升级到 OpenShift AI 1.21 时，**modemesh-serving pod** 会尝试拉取不存在的镜像，从而导致镜像拉取错误。因此，无法使用 OpenShift AI 中的模型服务功能提供模型。**odh-**

openvino-servingruntime-container-v1.21.0-15 镜像现在可以成功部署。

RHODS-5945 - OpenShift AI 中无法启用 Anaconda 专业版

在 OpenShift AI 中无法启用 Anaconda 专业版。相反，关联的 pod 的 Events 页面中会显示 InvalidImageName 错误。Anaconda 专业版现在可以成功启用。

RHODS-5822 - 当数据科学项目创建的 PVC 超过 90% 和 100% 时，管理员用户不会被警告。

当 PVC 超过其容量的 90% 和 100% 时，未能向 admin 用户显示由数据科学项目创建的 PVC 的警告。管理员用户现在可以查看当 PVC 超过 90% 和从仪表板中其容量的 100% 时的警告。

RHODS-5889 - 如果数据科学笔记本处于 "pending" 状态，则不会显示 Error 信息

如果无法创建笔记本 pod，OpenShift AI 接口不会显示错误消息。现在，如果无法生成数据科学笔记本，则会显示错误消息。

rhODS-5886 - 从数据科学工作台返回 Hub Control Panel 仪表板失败

如果您试图通过点 **File** → **Log Out**，从工作台 Jupyter 笔记本返回仪表板，您将被重定向到仪表板并保持在 "Logging out" 页面。同样，如果您尝试通过点 **File** → **Hub Control Panel** 以返回仪表板，则您会错误地重定向到 启动笔记本服务器 页面。从数据科学工作台返回 Hub Control Panel 仪表板现在可以按预期工作。

rhoDS-6101 - 管理员无法停止所有笔记本服务器

OpenShift AI 管理员无法同时停止所有笔记本服务器。管理员现在可以使用 **Stop all servers** 按钮停止所有笔记本服务器，并通过从相关用户旁的操作菜单中选择 **Stop server** 来停止单个笔记本。

RHODS-5891 - Workbench 事件日志没有明确可见

在创建工作台时，用户无法在 OpenShift AI 界面中轻松找到事件日志窗口。现在，当您悬停在其中时，**Status** 列下的 **Starting** 标签是下划线的，这表示您可以点它来查看笔记本状态和事件日志。

rhODS-6296 - 使用 Google Chrome 以外的浏览器时, ISV 图标不会被显示

当使用 Google Chrome 以外的浏览器时, 不会呈现 Explore 和 Resources 页面下的所有 ISV 图标。ISV 图标现在在所有支持的浏览器中正确显示。

rhODS-3182 - Jupyter 中显示了可用 GPU 的数量

当用户试图在 Jupyter 中创建 notebook 实例时, 不会更新用于调度的最大 GPU 数, 因为分配了 GPU。Jupyter 现在显示可用 GPU 的正确数量。

RHODS-5890 - 当多个持久性卷挂载到同一目录中时, 工作台无法启动

在同一工作台将多个持久性卷 (PV) 挂载到同一工作台中的同一挂载文件夹时, 创建笔记本 pod 会失败且没有显示错误来指示问题。

RHODS-5768 - Data Science 项目对 Red Hat OpenShift AI 中的用户不可见

在项目的 Display Name 属性末尾删除 [DSP] 后缀会导致相关的数据科学项目不再可见。用户无法再删除此后缀。

rhoDS-5701 - 数据连接配置详情被覆盖

当数据连接添加到工作台时, 这些数据连接的配置详情会保存在环境变量中。添加第二个数据连接时, 配置详情使用相同的环境变量保存, 这意味着覆盖第一个数据连接的配置。目前, 用户可以为每个工作台添加一个数据连接。

RHODS-5252 - 笔记本管理页面没有为用户提供笔记本服务器的管理员访问权限

从 OpenShift AI 仪表板访问的笔记本管理页面没有提供管理员访问用户笔记本服务器的方法。管理员仅限于启动或停止用户笔记本服务器。

RHODS-2438 - 升级时PyT 和 TensorFlow 镜像不可用

当从 OpenShift AI 1.3 升级到更新的版本时，PyTorch 和 TensorFlow 镜像在大约 30 分钟不可用。因此，在升级过程中，用户无法在 Jupyter 中启动 PyTorch 和 TensorFlow 笔记本。这个问题现已解决。

rhODS-5354 - 启动笔记本服务器时环境变量名称不会被验证

在启动笔记本服务器页面中不会验证环境变量名称。如果添加了无效的环境变量，用户无法成功启动笔记本。环境变量名称现在会实时检查。如果输入了无效的环境变量名称，则会显示错误消息表示有效环境变量名称必须包含字母字符、数字、_、-、或 .，且不得以数字开头。

rhoDS -4617 - 只有在 GPU 可用时才会看到 GPU 数

在以前的版本中，只有 GPU 节点可用时，在启动笔记本服务器页面中才会显示 GPU 数。现在，当一个自动扩展集群池在集群中定义，即使当前还没有 GPU 节点可用（这可能会在集群中置备新的 GPU 节点），GPU 的数量下拉菜单也会正确显示，。

RHODS-5420 - 如果集群管理员是集群中唯一用户，则不会获得管理员访问权限

在以前的版本中，当集群管理员是集群中唯一存在的用户时，它不会自动获得 Red Hat OpenShift 管理员访问权限。管理员访问权限现在可以正确地分配给管理员用户。

rhODS-4321 - 在笔记本选择过程中显示正确的软件包版本

启动一个笔记本服务器页面为 CUDA 笔记本镜像显示了一个不正确的版本号(11.4 而不是 11.7)。在此页面中不再指定安装的 CUDA 版本。

rhODS-5001 - 管理员用户可以向笔记本 pod 添加无效的容限

admin 用户可以在 Cluster settings 页面中添加非合规容限，而无需触发错误。如果添加了非合规容限，用户无法成功启动笔记本。现在，容限键会实时检查。如果输入了无效的容限名称，则会显示错误消息表示有效容限名称由字母数字字符（-, _、或 .）组成，必须以字母数字字符开头并以字母数字字符结尾。

RHODS-5100 - 组角色绑定没有应用到集群管理员

在以前的版本中，如果您为组而非特定用户分配了集群管理特权，控制面板无法识别管理组中的用户管理特权。现在，组角色绑定可以正确地应用到集群管理员。

rhoDS-4947 - Old Minimal Python 笔记本镜像在升级后保留

从 OpenShift AI 1.14 升级到 1.15 后，Minimal Python 笔记本的旧版本会保留，包括所有关联的软件包版本。升级后，Minimal Python 笔记本的旧版本不再保留。

rhoDS-4935 - Excessive "missing x-forwarded-access-token header" 错误信息显示在仪表板日志中

因为就绪度探测达到 /status 端点，rhods-dashboard pod 的日志中会包含大量的 "missing x-forwarded-access-token header" 错误消息。这个问题现已解决。

rhoDS-2653 - 在获取示例 Pachyderm 笔记本时出现错误

当用户尝试使用 Jupyter 中的示例 Pachyderm 笔记本获取镜像时会出现一个错误。指示镜像无法找到错误。Pachyderm 解决了这个问题。

rhoDS-4584 - Jupyter 无法使用 OpenVINO 笔记本镜像启动笔记本服务器

Jupyter 的启动一个笔记本服务器页无法使用 OpenVINO 笔记本镜像启动笔记本服务器。Intel 已经为 OpenVINO 操作器提供了一个更新来更正此问题。

rhoDS-4923 - 禁用使用数据收集后显示的非标准复选框

在 Cluster settings 页面中禁用使用数据收集后，当用户访问 OpenShift AI 仪表板的另一个区域，然后返回到 Cluster settings 页面，允许收集使用数据复选框应用了非标准类型，因此在选择或清除时不会与其他复选框相同。

[rhoDS-4938](#) - 在 Notebook Images 页面中显示不正确的标题

在 OpenShift AI 仪表板上的 Settings 页面中访问 Notebook Images 页面，在用户界面中显示不正确的标题。Notebook image settings 的标题显示 BYON image settings，Import Notebook images 标题显示 Import BYON images。现在，正确的标题会显示如预期。

[rhODS-4818](#) - Jupyter 在安装 NVIDIA GPU 附加组件时无法显示镜像

在安装 NVIDIA GPU 附加组件后，启动笔记本服务器页面不会显示笔记本镜像。现在，镜像会被正确显示，可以从启动笔记本服务器页面中启动。

[rhoDS-4797](#) - 当用量超过 90% 和 100% 时，PVC 使用限制警报不会被发送

当 PVC 超过容量的 90% 和 100% 时，未能触发和发送相关的警报。现在，这些警告会按预期触发并发送。

[rhODS-4366](#) - 集群设置在 operator 重启时被重置

当 OpenShift AI operator pod 重启时，集群设置有时会重置为默认值，删除任何自定义配置。当发布新版本的 OpenShift AI 以及运行 Operator 的节点失败时，OpenShift AI operator 会被重启。出现这个问题的原因是 Operator 部署的 ConfigMap 不正确。Operator 部署已被更新，这个问题不再会发生。

[rhoDS-4318](#) - OpenVINO 笔记本镜像无法成功构建

OpenVINO notebook 镜像未能成功构建并显示错误消息。这个问题现已解决。

[RHODS-3743](#) - Starburst Galaxy 快速启动没有在指令步骤中提供下载链接

Starburst Galaxy 快速启动（位于仪表板上的 Resources 页面中），要求用户打开 explore-data.ipynb notebook，但无法提供指令步骤中的链接。相反，链接是在快速启动的介绍中提供的。

[rhoDS-1974](#) - 更改警报通知电子邮件所需的 pod 重启

在 rhods-operator pod 和 prometheus114 pod 重启后，不会应用 Red Hat OpenShift AI Add-On 中的通知电子邮件 地址列表的更改。

[RHODS-2738](#) - Red Hat OpenShift API Management 1.15.2 附加组件安装无法成功完成

对于与 Red Hat OpenShift API Management 1.15.2 附加组件集成的 OpenShift AI 安装，Red Hat OpenShift API Management 安装过程无法成功获取 SMTP 凭证 secret。因此，安装不会完成。

[RHODS-3237](#) - 仪表板中没有显示 GPU 教程

位于 [Gtc2018-numba](#) 的 "GPU 计算"教程不会显示在仪表板上的 Resources 页面中。

[RHODS-3069](#) - 当 GPU 节点不可用时 GPU 选择会保留

如果用户使用 GPU 支持置备笔记本服务器，而使用的 GPU 节点随后会从集群中移除，用户就无法创建笔记本服务器。这是因为最近一次用于附加 GPU 数的设置被默认使用。

[RHODS-3181](#) - Pachyderm 现在与 OpenShift Dedicated 4.10 集群兼容

Pachyderm 最初与 OpenShift Dedicated 4.10 不兼容，因此在 OpenShift Dedicated 4.10 集群中运行的 OpenShift AI 中不可用。Pachyderm 现在提供与 OpenShift Dedicated 4.10 兼容的版本。

[RHODS-2160](#) - 安装 OpenShift AI 和 OpenShift API Management 时卸载过程无法完成

当 OpenShift AI 和 OpenShift API 管理在同一集群中安装时，它们使用相同的虚拟私有集群(VPC)。这些附加组件的卸载过程会尝试删除 VPC。在以前的版本中，当同时安装 Add-ons 时，一个服务的卸载过程会被阻断，因为其他服务仍有 VPC 中的资源。清理过程已更新，不会发生此冲突。

[RHODS-2747](#) - 升级 OpenShift AI 后镜像被错误地更新

升级 OpenShift AI 后，Jupyter 无法更新其笔记本镜像。这是因为镜像缓存机制存在问题。现在，在升级后镜像会被正确更新。

RHODS-2425 - 在笔记本选择过程中显示正确 TensorFlow 和 TensorBoard 版本

Start a notebook 服务器 页面显示 TensorFlow 和 TensorBoard 在 TensorFlow 笔记本镜像中显示不正确的版本号(2.4.0)。这些版本已被修正为 TensorFlow 2.7.0 和 TensorBoard 2.6.0。

rhoDS-24339 - 启用的应用程序没有显示快速启动链接

对于一些应用程序，**Open quick start** 链接无法在 **Enabled** 页面中的应用程序标题中显示。因此，用户无法直接访问相关应用程序的快速入门。

rhODS-2215 - 在笔记本选择过程中显示不正确的 Python 版本

Start a notebook server 页面显示 TensorFlow 和 PyTorch 笔记本电脑镜像的错误版本 Python。另外，现在不再显示软件包版本号的第三个整数。

rhoDS-1977 - 在笔记本服务器启动失败后等待十分钟

在启动笔记本服务器时，如果 Jupyter leader pod 失败，该用户将无法访问其笔记本服务器，直到 pod 重启前，这需要大约 10 分钟。这个进程已被改进，以便在选择新 leader pod 时将用户重定向到其服务器。如果此过程超时，用户会看到 504 网关超时错误，并且可以刷新以访问其服务器。

第 8 章 已知问题

这部分论述了 Red Hat OpenShift AI 中已知的问题，以及这些问题的已知方法。

[RHOAIENG-8819](#) - `ibm-granite/granite-3b-code-instruct` 模型无法在单模式服务平台上部署

如果您试图使用 `vLLM ServingRuntime for KServe` 运行时在单模式服务平台上部署 `ibm-granite/granite-3b-code-instruct` 模型，则模型会失败，并显示的错误。

```
ValueError: Head size 80 is not supported by FlashAttention. Supported head sizes are: [32, 64, 96, 128, 160, 192, 224, 256]
```

临时解决方案

将以下参数添加到 `vLLM ServingRuntime for KServe` 运行时的 `ServingRuntime` 规格的 `env` 部分。

```
VLLM_ATTENTION_BACKEND=XFORMERS
```

使用自定义镜像创建的 [RHOAIENG-855 3- Workbench](#) 显示 `!Deleted` 标记

如果您在 OpenShift 集群上禁用内部镜像 `registry`，然后使用镜像标签导入的自定义镜像工作台，例如：`quay.io/my-wb-images/my-image:tag`，则会在 `Data Science Projects` 页面的 `Workbenches` 标签页中显示一个 `!Deleted` 标志。如果停止工作台，则无法重启它。

- OpenShift 集群管理员可确认集群中是否启用了内部镜像 `registry`。
- OpenShift AI admin 用户可以使用标签表示法确认自定义镜像是否已导入。

临时解决方案

使用 SHA 摘要导入自定义镜像，如 `quay.io/my-repo/my-image@sha256:xxxxxxxxxxxxxx`，然后使用自定义镜像创建工作台。

当将 OpenShift AI 2.8 升级到 2.10 时，[RHOAIENG-8294](#) - CodeFlare 错误

如果您试图将 **OpenShift AI 2.8** 升级到 **2.10**，则 **CodeFlare** 组件会显示以下出错信息，因为 **AppWrapper** 自定义资源定义(CRD)版本不匹配。

```
ReconcileCompletedWithComponentErrors DataScienceCluster resource reconciled with component
errors: 1 error occurred: * CustomResourceDefinition.apiextensions.k8s.io
"appwrappers.workload.codeflare.dev" is invalid: status.storedVersions[0]: Invalid value: "v1beta1":
must appear in spec.versions
```

临时解决方案

1. 删除现有的 **AppWrapper CRD**。

```
$ oc delete crd appwrappers.workload.codeflare.dev
```

2. 安装 **AppWrapper CRD** 的最新版本。

```
$ oc apply -f https://raw.githubusercontent.com/project-codeflare/codeflare-
operator/main/config/crd/crd-appwrapper.yml
```

RHOAIENG-8218 - 无法登录到在没有 OCP 内部镜像 registry 的 OpenShift 4.15 集群上创建的工作台

当您在没有启用 **OpenShift Container Platform** 内部镜像 registry 的 **OpenShift 4.15** 集群上创建工作台时，工作台可以成功启动，但您无法登录到它。

授权服务器遇到意外条件，阻止它满足请求 错误。

临时解决方案

为您创建的每个工作台在 **OCP** 中手动创建服务帐户令牌 **secret**。

1. 在 **OpenShift CLI** 中，使用以下命令来创建服务帐户令牌 **secret**。将 **<workbench-name >** 替换为您的工作台的名称，**< project-name >** 替换为数据科学项目的名称。

```
-----
cat <<'EOF' | oc apply -f -
---
kind: Secret
apiVersion: v1
metadata:
```

```

name: <workbench-name>-token
namespace: <project-name>
annotations:
  kubernetes.io/service-account.name: <workbench-name>
type: kubernetes.io/service-account-token
EOF
-----

```

2.

刷新浏览器中的页面并登录到您的工作台。

有关使用 Web 控制台创建服务帐户令牌 `secret` 的详情，请参考 [创建服务帐户令牌 secret](#)。

在 KServe 中的查询过程中，[RHOAIENG-7947- Model serving](#) 会失败

如果您最初安装 ModelMesh 组件并启用多模式服务平台，但稍后安装 KServe 组件并启用单模式服务平台，则对在单模式服务平台上部署的模型的统计请求可能会失败。在这些情况下，inference 请求会返回 404 - Not Found 错误，odh-model-controller 部署对象的日志会显示 Reconciler 错误消息。

临时解决方案

在 OpenShift 中，重启 odh-model-controller 部署对象。

[RHOAIENG-788](#) 7- Kue 无法监控 RayCluster 或 PyTorchJob 资源

当您创建启用了所有组件的 DataScienceCluster CR 时，Kueue 组件会在 Ray 组件和 training Operator 组件之前安装。因此，Kueue 组件不会监控 RayCluster 或 PyTorchJob 资源。

临时解决方案

执行以下操作之一：

- 安装 Ray 组件和 Training Operator 组件后，重启 redhat-ods-applications 命名空间中的 Kueue 控制器 pod。
- 或者，编辑 DataScienceCluster CR，将 kueue 组件标记为 Removed，等待 Kueue 被卸载，然后再次将 kueue 组件标记为 Managed。

RHOAIENG-7716 - Pipeline 条件组状态不会更新

当您运行具有条件组的管道（如 `dsl.if`）时，UI 会显示组的 **Running** 状态，即使管道执行完成后也是如此。

临时解决方案

您可以通过检查没有子任务保持活动状态来确认管道是否仍在运行。

1. 在 OpenShift AI 仪表板中点 **Data Science Pipelines** → **Runs**。
2. 在 **Project** 下拉菜单中选择您的数据科学项目。
3. 在 **Runs** 选项卡中，点您要检查状态的管道运行。
4. 展开 **condition** 组，再单击子任务。

此时会显示包含子任务信息的面板
5. 在面板中点 **Task details** 选项卡。

Status 字段显示子任务的正确状态。

升级后，RHOAIENG-7346 - 分布式工作负载不再从现有管道运行

升级到 OpenShift AI 2.10 后，如果仅在管道中创建集群，则分布式工作负载不再从现有管道运行。管道停止工作，Ray 集群不会启动。

临时解决方案

在分布式工作负载的每个现有管道的 Python 代码中，将 `cluster.wait_ready` 替换为 `time.sleep(180)`，并重新编译代码。导入管道并调度管道运行。

RHOAIENG-1197 - 因为管道运行创建页面中的 **End date picker** 在 Linux 上使用 **Firefox** 时，无法创建管道

当您尝试使用 Linux 上的 **Firefox** 创建带有调度的重复运行的管道时，启用 **End Date** 参数会导致日期和时间不是 **Number (NaN)** 值，您无法创建管道。

临时解决方案

在 Linux 上，使用不同的受支持的浏览器来创建调度的管道运行。如需有关 **OpenShift AI** 支持的浏览器的更多信息，请参阅 [Red Hat OpenShift AI 支持的配置](#)。

RHOAIENG-7209 - 设置默认管道 **root** 时显示错误

当使用 **Data Science Pipelines SDK** 或 **OpenShift AI** 用户界面设置默认管道 **root** 时，会出现类似如下的错误：

```
F0513 18:25:25.155794 28 main.go:49] failed to execute component: Failed to open bucket
"mlpipeline": Failed to retrieve credentials for bucket mlpipeline: Failed to get Bucket credentials from
secret name="" namespace="dspa1": resource name may not be empty44
```

临时解决方案

无。

rhoAIENG-6711 - **ODH-model-controller** 覆盖 **ServiceMeshMemberRoll** 对象中的 **spec.memberSelectors** 字段

如果您使用 **ServiceMeshMemberRoll** 资源的 **spec.memberSelectors** 字段将项目或命名空间添加到 **ServiceMeshMemberRoll** 资源中，**ODH-model-controller** 会覆盖字段。

临时解决方案

使用 **spec.members** 字段明确将命名空间添加到 **ServiceMeshMemberRoll** 资源中，如下例所示：

```
spec:
  members:
  - <namespace 1>
  - <namespace 2>
```

RHOAIENG-6649 - 在没有定义外部路由的模型上查看模型时会显示错误

如果您使用仪表板在未启用外部路由的模型服务器上部署模型，则在模型创建过程中可能会显示 **t.components is undefined** 错误消息。

临时解决方案

无。

RHOAIENG-6646 - 升级过程中查看 **Model Serving** 页面时会显示一个错误

如果您在 OpenShift AI 升级过程中尝试使用仪表板来部署模型，可能会显示 **t.status is undefined** 错误消息。

临时解决方案

等待升级的 OpenShift AI Operator 就绪，然后在浏览器中刷新页面。

RHOAIENG-6486 - 使用带有 TensorFlow 2024.1 笔记本镜像的 Elyra JupyterLab 扩展时，无法配置 Pod 标签、注解和容限

当在 TensorFlow 2024.1 笔记本镜像中使用 Elyra JupyterLab 扩展时，您无法从执行的管道配置 pod 标签、注解或容限。这是因为，依赖项与 kfp 和 tf2onnx 软件包冲突。

临时解决方案

如果您使用 TensorFlow 2024.1 笔记本镜像，请在完成工作后将分配的工作台笔记本镜像改为 Standard Data Science 2024.1 笔记本镜像。

在 Elyra JupyterLab 扩展中的 Pipeline 属性 选项卡中，将 Tensorflow 运行时镜像单独设置为管道节点的默认运行时镜像，以及每个管道节点的相关 pod 标签、注解或容限。

RHOAIENG-6435 - 分布式工作负载资源不包括在项目指标中

当您点 **Distributed Workloads Metrics > Project metrics** 并查看 **Requested resources** 部分时，所有项目 值都当前排除了尚未接受到队列的分布式工作负载的资源。

临时解决方案

无。

RHOAIENG-6409 - 在管道日志中出现无法保存参数 错误，以便成功运行

当您使用 Data Science Pipelines 2.0 运行管道多次时，无法在管道日志中出现 **Cannot save parameter error for successful pipeline run**。您可以安全地忽略这些错误。

临时解决方案

无。

RHOAIENG-6376 - 在管道组件中将 `pip_index_urls` 设置为包含端口号和路径的 URL 后，管道运行创建会失败

当您创建管道并将组件的 `pip_index_urls` 值设置为包含端口号和路径的 URL 时，编译管道代码，然后创建管道运行会导致以下错误：

```
ValueError: Invalid IPv6 URL
```

临时解决方案

1. 仅使用 `protocol://hostname` 创建一个新的 pip 服务器，并使用新服务器更新组件的 `pip_index_urls` 值。
2. 重新编译管道代码。
3. 创建新管道运行。

RHOAIENG-4812 - 分布式工作负载指标排除 GPU 指标

在这个 OpenShift AI 发行版本中，分布式工作负载指标排除 GPU 指标。

临时解决方案

无。

RHOAIENG-4570 - 现有 Argo 工作流安装与安装或升级冲突

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 OpenShift AI，请确保集群中没有现有的 Argo 工作流安装。如需更多信息，[请参阅启用数据科学管道 2.0](#)。

临时解决方案

删除现有的 Argo 工作流安装或将 `datasciencepipelines` 设置为 `Removed`，然后继续安装或升级。

RHOAIENG-3913 - Red Hat OpenShift AI Operator 错误地显示 Degraded 条件 False，并显示错误

如果您在 OpenShift AI Operator 使用的 DataScienceCluster (DSC)对象中启用了 KServe 组件，但没有安装依赖的 Red Hat OpenShift Service Mesh 和 Red Hat OpenShift Serverless Operator，则 DSC 对象中的 `kserveReady` 条件可以正确地显示 KServe 未就绪。但是，`Degraded` 条件会错误地显示 `False` 值。

临时解决方案

安装 Red Hat OpenShift Serverless 和 Red Hat OpenShift Service Mesh Operator，然后重新创建 DSC。

RHOAIENG-4240 - 作业无法在未安全的环境中提交到 Ray 集群

当从一个不安全的 OpenShift 集群中从笔记本运行分布式数据科学工作负载时，可能会显示 `ConnectionError: Failed to connect to Ray` 错误消息。

临时解决方案

在笔记本的 `ClusterConfiguration` 部分中，将 `openshift_oauth` 选项设置为 `True`。

RHOAIENG-3981 - 在不安全的环境中，等待 Ray 集群处于卡住状态的功能

当从未安全的 OpenShift 集群中的笔记本运行分布式数据科学工作负载时，在继续 (`cluster.wait_ready ()`) 前，在 Ray 集群就绪前等待 Ray 集群就绪的功能，即使 Ray 集群就绪。

临时解决方案

执行以下操作之一：

- 在笔记本的 ClusterConfiguration 部分中，将 `openshift_oauth` 选项设置为 `True`。
- 您可以通过打开 Ray 集群路由 URL 来手动检查 `Ray_ready ()` 功能，而不必使用 `cluster.wait_ready ()` 功能。当 URL 上有 Ray 仪表板时，集群就已就绪。

RHOAIENG-3025 - OVMS 预期目录布局与 KServe StoragePuller 布局冲突

当您使用 OpenVINO Model Server (OVMS)运行时在单一模型服务平台（使用 KServe）上部署模型时，OVMS 预期的目录布局和 KServe 使用模型逻辑之间存在不匹配。具体来说，OVMS 要求模型文件位于 `/mnt/models/1/` 目录中，而 KServe 将它们放在 `/mnt/models/` 目录中。

临时解决方案

执行以下操作：

1. 在 S3 兼容存储桶中，将模型文件放在名为 `1/` 的目录中，例如：`/mnt/s3_storage_bucket/models/1/<model_files >`。
2. 要使用 OVMS 运行时在单型号服务平台上部署模型，请选择以下选项之一来指定模型文件的路径：
 - 如果您使用 OpenShift AI 仪表板来部署模型，请在数据连接的 Path 字段中使用 `/mnt/s3_storage_bucket/models/` 格式指定模型文件的路径。不要将 `1/` 目录指定为路径的一部分。
 - 如果您要创建自己的 InferenceService 自定义资源来部署模型，请将 `storageURI` 字段的值配置为 `/mnt/s3_storage_bucket/models/`。不要将 `1/` 目录指定为路径的一部分。

KServe 从您指定的路径的子目录中拉取模型文件。在这种情况下，KServe 可以正确地从 S3 兼容存储中的 `<s3_storage_bucket>/models/1/` 目录中拉取模型文件。

RHOAIENG-3018 - OVMS on KServe 不会在仪表板中公开正确的端点

当您使用 OpenVINO Model Server (OVMS)运行时在单型号服务平台上部署模型时，部署模型的 Inference 端点 字段中的 URL 不会被完成。

临时解决方案

要将查询发送到模型，您必须将 `/v2/models/<model-name>/infer` 字符串添加到 URL 的末尾。将 `<model-name>` 替换为部署模型的名称。

RHOAIENG-2759 - 当项目中存在安全和常规模型服务器时，模型部署会失败

当您在使用令牌身份验证的项目中创建第二模型服务器时，而其他服务器不使用身份验证，则第二个模型的部署可能无法启动。

临时解决方案

无。

RHOAIENG-2602 - "Average 响应时间"服务器指标图因为 ModelMesh pod 重启而显示多行

如果 ModelMesh pod 被重启，Average 响应时间 服务器指标图会显示多行。

临时解决方案

无。

RHOAIENG-2585 - 当集群中未启用 UWM 时 UI 不会显示错误/警告

如果集群中禁用了 User Workload Monitoring (UWM)，Red Hat OpenShift AI 无法正确警告用户。UWM 对于模型指标的正确功能是必需的。

临时解决方案

手动确保在集群中启用了 UWM，如 [为用户定义的项目启用监控](#) 中所述。

RHOAIENG-2555 - 在更改 Serving 运行时时，模型框架选择器不会重置

当您使用 **Deploy 模型** 对话框在单型号服务平台上部署模型时，如果您选择了运行时和支持的框架，但切换到不同的运行时，不会重置现有的框架选择。这意味着，可以使用不支持所选运行时的框架部署模型。

临时解决方案

在部署模型时，如果您更改了所选运行时，请再次点 **Select a framework** 列表并选择支持的框架。

RHOAIENG-2468 - 与 KServe 位于同一项目中的服务可能会在 OpenShift 中无法访问

如果您在包含在单一型号服务平台上部署的模型的数据科学项目中部署非 OpenShift AI 服务（使用 KServe），则服务可访问性可能受 OpenShift 集群的网络配置的影响。特别是当您将 [OVN-Kubernetes 网络插件](#) 与主机网络命名空间结合使用时，这特别有用。

临时解决方案

执行以下操作之一：

- 在另一个数据科学项目中部署服务，其中不包含在单模型服务平台上部署的模型。或者，将服务部署到另一个 OpenShift 项目中。
- 在服务的科学项目中，添加一个 [网络策略](#) 来接受应用程序 pod 的入口流量，如下例所示：

```
apiVersion: networking.k8s.io/v1
kind: NetworkPolicy
metadata:
  name: allow-ingress-to-myapp
spec:
  podSelector:
    matchLabels:
      app: myapp
  ingress:
    - {}
```

■

RHOAIENG-2312 - 导入 numpy 在 code-server 工作台失败

在 code-server 工作台导入 numpy 会失败。

临时解决方案

1.

在 code-server workbench 中，从 Activity bar 中，选择菜单图标



) > View > Command Palette 以打开命令面板。

在 Firefox 中，您可以使用 F1 键盘快捷键打开命令面板。

2.

输入 `python: s`。

3.

从下拉列表中选择 **Python: Select interpreter** 操作。

4.

在 **Select Interpreter** 对话框中，选择 **Enter interpreter path...**。

5.

输入 `/opt/app-root/bin/python3` 作为解释器路径，然后按 **Enter**。

6.

从下拉列表中选择新的 **Python** 解释器。

7.

确认新解释器(`app-root`)显示在 **Status bar** 上。如果工作台停止并再次启动，所选解释器会保留，因此每个工作台只需要执行一次临时解决方案。

RHOAIENG-2228 - 性能指标图形在间隔设置为 15 秒时持续更改

在模型指标屏幕的 **Endpoint performance** 选项卡中，如果您将 **Refresh interval** 设为 15 秒，并且 **时间范围** 设为 1 小时，图形结果会持续变化。

临时解决方案

无。

RHOAIENG-2183 - 端点性能图可能会显示不正确的标签

在模型指标屏幕的 **Endpoint performance** 选项卡中，图形工具提示可能会显示不正确的标签。

临时解决方案

无。

RHOAIENG-1919 - Model Serving 页面在部署后无法立即获取或报告模型路由 URL

从 OpenShift AI 仪表板部署模型时，系统会显示以下警告信息，而模型的 **Status** 列表示成功并带有 OK/green checkmark。

```
Failed to get endpoint for this deployed model. routes.rout.openshift.io"<model_name>" not found
```

临时解决方案

刷新浏览器页面。

RHOAIENG-880 - 默认管道服务帐户无法创建 Ray 集群

您不能使用默认管道服务帐户创建 Ray 集群。

临时解决方案

通过在管道代码中添加以下行，使用 **CodeFlare SDK** 进行身份验证：

```
from codeflare_sdk.cluster.auth import TokenAuthentication
auth = TokenAuthentication(
    token=openshift_token, server=openshift_server
)
auth_return = auth.login()
```

**注意**

如果您的集群使用自签名证书，请在 `TokenAuthentication` 参数列表中包含 `ca-cert-path=<path >`。

RHOAIENG-404 - 在 OpenShift AI 仪表板中随机显示没有组件发现页面而不是 **Enabled** 页面

当您访问 Red Hat OpenShift AI 仪表板时，可能会出现 **No Components Found** 页面。

临时解决方案

刷新浏览器页面。

RHOAIENG-1128 - 当尝试增加未连接到工作台的持久性卷(PV)的大小时，会显示 **Unclear** 错误消息

当尝试增大没有连接到工作台的持久性卷(PV)的大小时，会显示不明确的错误消息。

临时解决方案

在尝试增大大小前，验证您的 PV 已连接到工作台。

RHOAIENG-545 - 在 JupyterLab 管道编辑器中无法指定通用默认节点运行时镜像

当您在 JupyterLab IDE 管道编辑器中编辑 Elyra 管道时，您单击 **PIPELINE PROPERTIES** 选项卡，并滚动到 **Generic Node Defaults** 部分并编辑 **Runtime Image** 字段，您的更改不会保存。

临时解决方案

为每个节点明确定义所需的运行时镜像。单击 **NODE PROPERTIES** 选项卡，然后在 **Runtime Image** 字段中指定所需的镜像。

RHOAIENG-497 - 在 OpenShift Service Mesh CR 中删除 DSCI 结果，在没有用户通知的情况下删除

如果您删除 `DSCInitialization` 资源，`OpenShift Service Mesh CR` 也会被删除。不显示警告消息。

临时解决方案

无。

RHOAIENG-282 - 如果所需资源不可用，则不应分配工作负载

有时，即使单个机器实例没有足够的资源来成功置备 `RayCluster`，则有时也会分配工作负载。`AppWrapper CRD` 处于 `Running` 状态，相关的 `pod` 无限期处于 `Pending` 状态。

临时解决方案

在集群中添加额外资源。

RHOAIENG-131 - 在 `InferenceService` 报告为 `Loaded` 后的 `gRPC` 端点没有正确响应

当生成大量 `InferenceService` 实例时，`Service Mesh Control Plane (SMCP)` 会变得无响应。`InferenceService` 实例的状态为 `Loaded`，但对 `gRPC` 端点的调用会返回错误。

临时解决方案

编辑 `ServiceMeshControlPlane` 自定义资源(CR)以增加 `Istio` 出口和入口 `pod` 的内存限值。

RHOAIENG-130 - 刚刚启动模型时同步问题

当 `KServe` 容器的状态为 `Ready` 时，即使 `TGIS` 容器未就绪，也会接受请求。

临时解决方案

等待几秒钟，以确保所有初始化都已完成，并且 `TGIS` 容器实际就绪，然后查看请求输出。

RHOAIENG-3115 - 在模型显示为 `ready` 后无法查询模型

使用多模型服务平台部署的模型可能会无响应查询，尽管在仪表板中显示 `Ready`。在查询模型端点时，

您可能会看到 "Application is not available" 响应。

临时解决方案

等待 30-40 秒，然后在浏览器中刷新页面。

RHOAIENG-1619 (之前记录的是 **DATA-SCIENCE-PIPELINES-165**) - S3 存储桶无法写入时的 Poor 错误消息

当您设置数据连接时，S3 存储桶无法写入，并且您尝试上传管道，错误消息 **Failed to store pipelines is not helpful**。

临时解决方案

验证您的数据连接凭证是否正确，并且您对指定的存储桶有写入权限。

RHOAIENG-1207 (之前记录的为 **ODH-DASHBOARD-1758**) - Error duplicating OOTB 自定义运行时几次

如果您复制了 **model-serving** 运行时多次，则重复会失败，并显示 **Serving** 运行时名称 "**<name>**" **already exists** 错误信息。

临时解决方案

将 **metadata.name** 字段更改为唯一值。

RHOAIENG-1204 (之前记录在 **ODH-DASHBOARD-1771**) - JavaScript 错误，在 Pipeline 步骤初始化过程中

有时，管道运行 详情页面 会在运行启动时停止工作。

临时解决方案

刷新页面。

RHOAIENG-1203 (之前记录的为 **ODH-DASHBOARD-1781**) - Missing tooltip for Started Run status

Data Science 管道运行有时不会显示显示的状态图标的工具提示文本。

临时解决方案

如需更多信息，查看管道运行 详情页面 并查看运行输出。

RHOAIENG-1201 (以前称为 ODH-DASHBOARD-1908) - 无法创建带有空环境变量的工作台

在创建工作台时，如果您点击 **Add** 变量，但没有从列表中选择环境变量类型，则无法创建工作台。该字段未标记为必需，不显示任何错误消息。

临时解决方案

无。

RHOAIENG-1196 (之前记录的为 ODH-DASHBOARD-2140) - 仪表板中显示的软件包版本与已安装的版本不匹配

仪表板可能会显示不准确的软件包版本号，如 JupyterLab 和 Notebook。如果手动更新软件包，则镜像中的软件包版本号可能会有所不同。

临时解决方案

要查找软件包的真正版本号，请运行 `pip list` 命令并搜索软件包名称，如下例所示：

```
$ pip list | grep jupyterlab
jupyterlab          3.5.3
$ pip list | grep notebook
notebook            6.5.3
```

RHOAIENG-582 (之前记录的为 ODH-DASHBOARD-1335) - Rename Edit permissions to Contributor

术语 *Edit is not accurate*:

- 对于大多数资源，具有 **Edit** 权限的用户不能编辑资源，他们也可以创建和删除资源。
- 具有 **Edit** 权限的用户无法编辑项目。

术语 *Contributor* 可以更准确地描述此权限授予的操作。

临时解决方案

无。

RHOAIENG-432 (以前称为 **RHODS-12928**) - 使用不支持的字符可以生成带有多个短划线的 Kubernetes 资源名称

当您创建资源并在名称中指定不支持的字符时，每个空格都会被替换为一个短划线，其他不支持的字符会被删除，这可能会导致无效资源名称。

临时解决方案

无。

RHOAIENG-226 (之前记录的为 **RHODS-12432**) - 删除 notebook-culler ConfigMap 会导致仪表板中出现 **Permission Denied**

如果您在 `redhat-ods-applications` 命名空间中删除 `notebook-controller-culler-config` ConfigMap，则无法将更改保存到 OpenShift AI 仪表板上的 **Cluster Settings** 页面。保存操作失败并显示 HTTP 请求失败。

临时解决方案

以具有 `cluster-admin` 权限的用户完成以下步骤：

1. 使用 `oc` 客户端登录到集群。
2. 输入以下命令更新 `redhat-ods-applications` 应用程序命名空间中的 `OdhDashboardConfig` 自定义资源：

```
$ oc patch OdhDashboardConfig odh-dashboard-config -n redhat-ods-applications --  
type=merge -p '{"spec": {"dashboardConfig": {"notebookController.enabled": true}}}'
```

RHOAIENG-133 - 在笔记本重启后无法运行 Elyra 管道

如果您使用 Elyra JupyterLab 扩展在 JupyterLab 中创建并运行数据科学管道，且您在工作台中创建工作台并在工作台中指定笔记本镜像后，您无法执行管道，即使在重启笔记本后也是如此。

临时解决方案

1. 停止正在运行的笔记本。
2. 编辑工作台以做一个小的修改。例如，添加新的 dummy 环境变量，或删除现有的不必要的环境变量。保存您的更改。
3. 重启笔记本。
4. 在 JupyterLab 的左侧边栏中，单击 Runtimes。
5. 确认选择了默认运行时。

RHOAIENG-52 - Token 身份验证在带有自签名证书的集群中失败

如果您使用自签名证书，且您在笔记本中使用 Python codeflare-sdk，或者在管道的一部分中使用 Python codeflare-sdk，则令牌身份验证将失败。

临时解决方案

无。

RHOAIENG-11 - 9 月安装的 CodeFlare Operator 实例不被支持

在 Red Hat OpenShift AI 中，codeFlare Operator 包含在基本产品中，而不包含在一个单独的 Operator 中。不支持独立于红帽或社区安装 CodeFlare Operator 实例。

临时解决方案

删除任何已安装的 CodeFlare Operator，并安装并配置 Red Hat OpenShift AI，如红帽知识库解决方案 [如何从数据科学集群中单独安装的 CodeFlare Operator 迁移](#) 中所述。

RHODS-12798 - Pod 失败"unable to init seccomp"错误

Pod 无法显示 CreateContainerError 状态或 Pending 状态，而不是 Running 状态，因为已知的内核错误引入了 seccomp 内存泄漏。当您检查 pod 失败的命名空间中的事件或运行 `oc describe pod` 命令时，会出现以下错误：

```
runc create failed: unable to start container process: unable to init seccomp: error loading seccomp filter into kernel: error loading seccomp filter: errno 524
```

临时解决方案

增加 `net.core.bpf_jit_limit` 的值，如红帽知识库解决方案 [Pod 失败](#)，并将 [error seccomp 过滤器](#) 加载到 [kernel: errno 524 in OpenShift 4](#)。

KUBEFLOW-177 - 来自应用程序的 Bearer 令牌不是由 OAuth-proxy 转发

如果应用程序的内部身份验证机制基于 bearer 令牌，则无法将应用程序用作自定义工作台镜像。OAuth-proxy 配置从标头中删除 bearer 令牌，应用程序无法正常工作。

临时解决方案

无。

注意BOOKS-210 - 笔记本无法在 Jupyter 中导出为 PDF 文件

当您笔记本导出为 Jupyter 中的 PDF 文件时，导出过程会失败并显示错误。

临时解决方案

无。

RHOAIENG-1210 (之前记录的为 **ODH-DASHBOARD-1699**) - Workbench 不会为所有配置更改自动重启

当您编辑工作台的配置设置时，会出现一条警告消息，表示当您对配置设置进行任何更改时，工作台将重新启动。这个警告是误导，因为在以下情况下，工作台不会自动重启：

- 编辑名称
- 编辑描述
- 编辑、添加或删除现有环境变量的键和值

临时解决方案

手动重启工作台。

RHOAIENG-1208 (之前记录的为 **ODH-DASHBOARD-1741**) - 无法创建一个以数字开头的工作台

如果您试图创建名称以数字开头的工作台，则工作台不会启动。

临时解决方案

删除工作台，创建一个以字母开头的名称的新工作台。

RHOAIENG-1205 (之前记录的为 **RHODS-11791**) - 升级后会启用使用数据收集

如果您之前有 **Allow collection usage data** 选项（即禁用），则在升级 **OpenShift AI** 时此选项将变为选择（启用）。

临时解决方案

手动重置 **Allow collection usage data** 选项。要做到这一点，请执行以下操作：

1. 在 OpenShift AI 仪表板中，在左侧菜单中点击 **Settings** → **Cluster settings**。

Cluster Settings 页面将打开。
2. 在 **Usage data collection** 部分中，取消选择 **Allow collection usage data**。
3. 点 **Save Changes**。

KUBEFLOW-157 - 如果您已从 OpenShift AI 仪表板注销 JupyterLab 时无法正常工作

如果在退出 JupyterLab 前退出 OpenShift AI 仪表板，则从 JupyterLab 注销将无法成功。例如，如果您知道 Jupyter 笔记本的 URL，您可以在浏览器中再次打开它。

临时解决方案

从 OpenShift AI 仪表板注销前，先从 JupyterLab 注销。

RHODS-9789 - 如果 Pipeline 服务器包含数据库名或用户名字段中包含短划线的自定义数据库，则它们无法启动

当您创建一个使用自定义数据库的管道服务器时，如果您为 **dbname** 字段或 **username** 字段设置的值中包含短划线，则管道服务器无法启动。

临时解决方案

编辑管道服务器，从受影响的字段中省略横线。

RHOAIENG-580 (之前记录的为 **RHODS-9412**) - 如果具有编辑权限的用户创建工作台，Slyra 管道将无法运行

如果被授予项目的编辑权限的用户创建了项目工作台，该用户会看到以下行为：

- 在工作台创建过程中，用户会看到与创建 Kubernetes 角色绑定相关的 **Error create**

workbench 消息。

- 虽然前面的错误消息，OpenShift AI 仍然会创建工作台。但是，错误消息意味着用户无法使用工作台来运行 Elyra 数据科学管道。
- 如果用户尝试使用工作台运行 Elyra 管道，Jupyter 会显示一个 Error making request 消息。

临时解决方案

具有管理员权限（如项目所有者）的用户必须代表具有编辑权限的用户创建工作台。然后，用户可以使用工作台运行 Elyra 管道。

RHOAIENG-583（之前记录的为 **RHODS-8921** 和 **RHODS-6373**）- 您无法创建管道服务器，或者在超过累积字符限制时启动工作台

当数据科学项目名称和管道服务器名称的累积字符限制超过 62 个字符时，您无法成功创建管道服务器。同样，当数据科学项目名称的累积字符限制和工作台名称超过 62 个字符时，工作台无法启动。

临时解决方案

重命名您的数据科学项目，使其不超过 30 个字符。

RHODS-7718 - 没有仪表板权限的用户可以无限期地继续使用其正在运行的笔记本和工作台

当 Red Hat OpenShift AI 管理员撤销用户权限时，用户可以无限期地使用其运行的笔记本和工作台。

临时解决方案

当 OpenShift AI 管理员撤销用户的权限时，管理员也应停止该用户的任何正在运行的笔记本和工作台。

RHOAIENG-1157（之前记录的为 **RHODS-6955**）- 尝试编辑工作台时可能会出现错误

在编辑工作台时，可能会出现类似如下的错误：

Error creating workbench

Operation cannot be fulfilled on notebooks.kubeflow.org "workbench-name": the object has been modified; please apply your changes to the latest version and try again

临时解决方案

无。

RHOAIENG-1132（之前记录的为 **RHODS-6383**） - 在工作台创建过程中不需要显示 ImagePullBackOff 错误消息

Pod 可能会遇到从容器 registry 中拉取容器镜像的问题。如果发生错误，相关的 pod 会进入 ImagePullBackOff 状态。在工作台创建过程中，如果发生 ImagePullBackOff 错误，则不会显示适当的信息。

临时解决方案

检查事件日志以了解更多有关 ImagePullBackOff 错误的信息。为此，可在其启动时点工作台状态。

RHOAIENG-1152（之前记录的为 **RHODS-6356**） - 笔记本创建过程对于从未登录到仪表板的用户失败

仪表板的笔记本管理页面显示属于 OpenShift 中用户组和 admin 组的用户。但是，如果管理员尝试代表从未登录到仪表板的用户启动笔记本服务器，服务器创建过程会失败，并显示以下出错信息：

Request invalid against a username that does not exist.

临时解决方案

请求相关用户登录到仪表板。

rhODS-5763 - 在笔记本选择过程中显示正确的软件包版本

启动一个笔记本服务器 页面显示 Anaconda 笔记本镜像的版本号。

临时解决方案

无。

rhODS-5543 - 使用 NVIDIA GPU Operator 时，超过 Node Autoscaler 创建的节点数量

当因为可用资源不足而无法调度 pod 时，Node Autoscaler 将创建一个新节点。在新创建的节点接收相关 GPU 工作负载前会有一个延迟。因此，pod 无法调度，Node Autoscaler 会不断创建额外的新节点，直到其中一个节点准备好接收 GPU 工作负载。有关此问题的更多信息，请参阅红帽知识库解决方案 [使用 NVIDIA GPU Operator 时，超过 Node Autoscaler 创建的节点数量](#)。

临时解决方案

在 `machineset.spec.template.spec.metadata` 中应用 `cluster-api/accelerator` 标签。这会导致自动扩展将这些节点视为未就绪，直到部署了 GPU 驱动程序。

RHOAIENG-1137（之前记录的为 **RHODS-5251**）- 笔记本服务器管理页面显示已丢失权限的用户

如果之前在 Jupyter 中启动笔记本服务器的用户丢失了其权限（例如，如果 OpenShift AI 管理员更改了用户组设置或从允许组中删除用户，则管理员将继续在服务器管理页面中查看用户的笔记本服务器。因此，管理员可以重启属于撤销权限的用户的笔记本服务器。

临时解决方案

无。

rhoDS-4799 - Tensorboard 需要手动步骤来查看

当用户有 TensorFlow 或 PyTorchbook 镜像，并希望使用 TensorBoard 显示数据，需要手动步骤在笔记本环境中包含环境变量，并在您的代码中导入这些变量。

临时解决方案

当您启动笔记本服务器时，使用以下代码来设置 `TENSORBOARD_PROXY_URL` 环境变量的值，以使用您的 OpenShift AI 用户 ID。

```
import os
os.environ["TENSORBOARD_PROXY_URL"] = os.environ["NB_PREFIX"] + "/proxy/6006/"
```

rhoDS-4718 - Intel® oneAPI AI Analytics Toolkits 快速启动引用不存在的示例笔记本

Intel® OneAPI AI Analytics Toolkits 快速开始（位于仪表板上的 **Resources** 页面中），要求用户以指令步骤的一部分加载示例笔记本，但引用相关存储库中不存在的笔记本。

临时解决方案

无。

rhODS-4627 - 负责验证 **Anaconda** 专业版许可证的 **CronJob** 已暂停，且不会每天运行

负责验证 **Anaconda** 专业版许可证的 **CronJob** 由 **OpenShift AI** 操作器自动暂停。因此，**CronJob** 不会每日运行。另外，当 **Anaconda** 专业版的许可证过期时，**Anaconda** 专业版在 **OpenShift AI** 仪表板中没有被禁用。

临时解决方案

无。

RHOAIENG-1141（之前记录的为 **RHODS-4502**）- 仪表板中的 **NVIDIA GPU Operator** 标题显示不必要的按钮

安装 **NVIDIA GPU Operator** 后，**Jupyter** 中会自动提供 **GPU**。因此，在 **Explore** 页面中的 **NVIDIA GPU Operator** 标题中的 **Enable** 按钮是多余的。另外，点 **Enable** 按钮会将 **NVIDIA GPU Operator** 标题移到 **Enabled** 页面，即使 **Operator** 没有被安装。

临时解决方案

无。

RHOAIENG-1135（之前记录的为 **RHODS-3985**）- 仪表板不会在 **ISV operator** 卸载后显示已启用的页面内容

卸载 **ISV** 操作器后，仪表板的 **Enabled** 页没有显示任何内容。相反，会显示以下错误：

```
Error loading components
HTTP request failed
```

临时解决方案

等待 30-40 秒，然后在浏览器中刷新页面。

rhODS-3984 - 在笔记本选择过程中显示正确的软件包版本

在 OpenShift AI 界面中，启动笔记本服务器页面 会显示 oneAPI AI Analytics Toolkit 笔记本镜像中包含的 JupyterLab 和 Notebook 软件包的版本号。该页面还可能显示此镜像使用的 Python 版本的错误值。

临时解决方案

当您启动 oneAPI AI Analytics Toolkit 笔记本服务器时，您可以在笔记本服务器上安装了哪些 Python 软件包，以及在笔记本单元中运行 `!pip list` 命令的软件包的版本。

RHODS-2956 - 创建笔记本实例时可能会出现错误

在 Jupyter 中创建 notebook 实例时，有时会出现未找到目录错误。单击 **Dismiss** 可忽略此错误消息。

临时解决方案

无。

RHOAING-1147（之前记录的为 RHODS-2881） - 对仪表板的操作没有明确可见

重新验证禁用的应用程序许可证的仪表板操作，删除禁用的应用程序标题对用户没有明确可见。当用户点击应用程序标题的 **Disabled** 标签时，会出现这些操作。因此，预期的工作流对于用户可能并不明确。

临时解决方案

无。

RHOAIENG-1134（之前记录的为 RHODS-2879） - 许可证重新验证操作不必要

对于没有许可证验证或激活系统的应用程序，用于重新验证禁用的应用程序许可证的仪表板操作并不必要。另外，当用户尝试重新验证无法重新验证的许可证时，不会显示反馈以说明无法完成该操作的原因。

临时解决方案

无。

RHOAIENG-2305 (之前记录的为 **RHODS-2650**) - Pachyderm 部署期间可能会出现错误

在创建 Pachyderm operator 的实例时，webhook 错误会出现间歇性错误，从而导致创建过程成功启动。webhook 错误表明，Pachyderm operator 无法进行健康检查，从而导致它重启，或者 Operator 进程超过其容器分配的内存限值，可触发内存不足(OOM)终止。

临时解决方案

重复 Pachyderm 实例创建过程，直到不再显示错误。

RHODS-2096 - IBM Watson Studio 不在 OpenShift AI 中

当在 OpenShift Dedicated 4.9 或更高版本上安装 OpenShift AI 时，IBM Watson Studio 不可用，因为它与这些版本的 OpenShift Dedicated 不兼容。

临时解决方案

联系 [Marketplace 支持](#)，以获取在 OpenShift Dedicated 4.9 及更高版本上手动配置 Watson Studio 的帮助。

RHODS-1888 - OpenShift AI hyperlink 在卸载后仍然可见

从 OpenShift Dedicated 集群卸载 OpenShift AI Add-on 时，应用程序启动程序菜单中可以看到 OpenShift AI 界面的链接。点击此链接会导致 "Page Not Found" 错误，因为 OpenShift AI 不再可用。

临时解决方案

无。

第 9 章 产品特性

Red Hat OpenShift AI 为数据科学家和 IT 操作管理员提供一组丰富的功能。如需更多信息，请参阅 [Red Hat OpenShift AI 简介](#)。