



Red Hat OpenShift Data Science 1

Red Hat OpenShift Data Science 简介

法律通告

Copyright © 2023 Red Hat, Inc.

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

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

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

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

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

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

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

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

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

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

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

All other trademarks are the property of their respective owners.

摘要

Red Hat OpenShift Data Science 是人工智能和机器学习应用程序的数据科学家和开发人员的平台。

目录

第 1 章 OPENSIFT DATA SCIENCE 概述	3
第 2 章 支持的配置	4
第 3 章 产品特性	5
3.1. 数据科学家的功能	5
3.2. IT OPERATIONS 管理员的功能	5
第 4 章 试用	7
第 5 章 获取它	8

第 1 章 OPENSIFT DATA SCIENCE 概述

通过使用 Red Hat OpenShift Data Science，用户可以集成数据、智能和机器学习软件，以执行端到端机器学习工作流。OpenShift Data Science 有两个支持的配置：

1. 在一个红帽管理的环境（如 Red Hat OpenShift Dedicated 和 Red Hat OpenShift Service on Amazon Web Services (ROSA)）中作为一个附加组件安装。
2. 在自助管理环境中（如 Red Hat OpenShift Container Platform）作为自我管理的 Operator 安装。

对于数据科学家，OpenShift Data Science 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以及 TensorFlow 和 PyTorch 框架进行优化。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。您还可以使用图形处理单元(GPU)加快数据科学试验。

对于管理员，OpenShift Data Science 为现有的 Red Hat OpenShift Dedicated 或 ROSA 环境提供了数据科学工作负载功能。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。

第 2 章 支持的配置

OpenShift Data Science 有两个支持的配置：

- Red Hat OpenShift Dedicated **的受管云服务附加组件**（具有 AWS 或 GCP 的客户云订阅）或 Red Hat OpenShift Service on Amazon Web Services (ROSA)。有关 Red Hat 受管环境中的 OpenShift Data Science 的详情，请参考 [Red Hat OpenShift Data Science 产品文档](#)。
- 您可以在自我管理的环境中（如 OpenShift Container Platform）安装内部或公有云中的**自我管理**软件。有关在连接或断开连接的环境中的 OpenShift 集群上将 OpenShift Data Science 作为自我管理的软件的信息，请参阅 [Red Hat OpenShift Data Science 自我管理的产品文档](#)。

第 3 章 产品特性

Red Hat OpenShift Data Science 为数据科学家和 IT 操作管理员提供了一些功能。

3.1. 数据科学家的功能

容器

虽然 JupyterLab 等工具已经为数据科学家在其机器上开发模型提供了直观的方法，但存在与协作和共享工作相关的固有复杂性。此外，当您需要自行购买和维护时，使用专用的硬件（如强大的 GPU）可能会非常昂贵。OpenShift Data Science 中包含的 Jupyter 环境可让您在任意地方获取开发环境。由于所有工作负载都作为容器运行，协作也会像与团队成员共享镜像一样容易，甚至将其添加到他们可以使用的默认容器列表中。GPU 和大量内存可突然成为更多可访问，因为您的笔记本不再受到支持的限制。

与第三方机器学习工具集成

我们在我们最喜欢的工具或服务之间无法很好地运行。在设计 OpenShift Data Science 时，实现灵活性的非常重要的因素。您可以在 OpenShift Data Science 中使用广泛的开源和第三方工具。这些工具支持完整的机器学习生命周期，从数据工程和功能提取到模型部署和管理。

在笔记本中使用 Git

使用 Jupyter 的 Git 界面与他人协作，并跟踪代码的更改。

安全构建笔记本镜像

从预定的默认笔记本镜像中选择与模型开发所需的工具和库进行预配置。软件堆栈，特别是机器学习中涉及的软件堆栈，往往是复杂的系统。Python 生态系统中有多种模块和库可以使用，因此确定要使用库的哪个版本可能非常困难。OpenShift Data Science 附带许多打包的笔记本镜像，它们构建了来自数据科学家和建议引擎的见解。您可以在右面上快速启动新项目，而无需担心从随机上游存储库下载未验证且可能不安全的镜像。

自定义笔记本

除了由红帽和独立软件供应商 (ISV) 提供和支持的笔记本镜像外，您还可以配置针对项目特定要求的自定义笔记本镜像。

数据科学项目(data Science)管道

OpenShift Data Science 支持数据科学管道，以实现运行数据科学项目工作负载的成熟和高效方法。您可以标准化和自动化机器学习 workflow，以便您开发和部署数据科学模型。

模型服务

作为数据科学家，您现在可以部署经过培训的机器学习模型，为生产环境中的智能应用程序提供服务。模型的功能作为服务端点提供，可用于测试或集成到应用程序中。您对此服务的执行方式有更多控制。

3.2. IT OPERATIONS 管理员的功能

使用身份提供程序管理用户

OpenShift Data Science 支持与 OpenShift 集群相同的身份验证系统。默认情况下，OpenShift Data Science 可以被身份提供程序中列出的所有用户访问，这些用户不需要单独的凭证集来访问 OpenShift Data Science。另外，您可以通过创建指定用户子集的 OpenShift 组来限制有权访问的用户集合。您还可以创建一个 OpenShift 组，来标识具有 OpenShift Data Science 管理员访问权限的用户列表。

使用 OpenShift 管理资源

使用您现有的 OpenShift 知识来为您的 OpenShift Data Science 用户配置和管理资源。

控制红帽对使用数据的收集

选择是否允许红帽收集有关集群中 OpenShift Data Science 使用的数据。在 OpenShift 集群上安装 OpenShift Data Science 时，会默认启用数据收集。

将自动扩展应用到集群以降低使用量成本

使用集群自动扩展来调整集群的大小，以满足其当前需求并优化成本。

通过停止空闲笔记本来管理资源使用情况

通过自动停止在一段时间内闲置的笔记本服务器，减少 OpenShift Data Science 部署中的资源使用情况。

实施 model-serving 运行时

OpenShift Data Science 支持模型服务运行时。模型运行时提供与指定模型服务器和它支持的模型框架集成。默认情况下，OpenShift Data Science 包含 OpenVINO Model Server 运行时。但是，如果此运行时不满足您的需要（例如，如果它不支持特定模型框架），您可以添加自己的自定义运行时。

在断开连接的环境中安装

OpenShift Data Science 自我管理支持在断开连接的环境中安装。断开连接的集群位于受限网络中，通常在防火墙后面，无法访问互联网。在这种情况下，集群无法访问红帽提供的 OperatorHub 源所在的远程 registry。在这种情况下，您可以通过使用一个私有的 registry，其中包括相关镜像的副本（mirrored 的镜像），将 OpenShift Data Science Operator 部署到断开连接的环境中。

第 4 章 试用

数据科学家和开发人员可在 [Red Hat Developer sandbox](#) 环境中尝试 OpenShift Data Science 和访问教程和活动。

[60 天产品试用](#) 可以让 IT 操作管理员在自己的集群中试用 OpenShift Data Science。

第 5 章 获取它

受管云服务

您有以下选项用于订阅 OpenShift Data Science 作为受管服务：

- 对于 OpenShift Dedicated，请订阅红帽。
- 对于 Red Hat OpenShift Service on Amazon Web Services (ROSA)，通过红帽订阅或通过 AWS Marketplace 订阅。

自我管理的软件

要让 Red Hat OpenShift Data Science 作为自我管理的软件，请与您的红帽客户团队一起注册。