



Red Hat OpenShift Data Science 1

Red Hat OpenShift Data Science の概要

法律上の通知

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.

概要

Red Hat OpenShift Data Science は、人工知能および機械学習アプリケーションのデータサイエンティストおよび開発者向けのプラットフォームです。

目次

第1章 OPENSIFT DATA SCIENCE の概要	3
第2章 サポートされる構成	4
第3章 製品機能	5
3.1. データサイエンティスト向けの機能	5
3.2. IT オペレーション管理者の機能	5
第4章 テスト	7
第5章 取得	8

第1章 OPENSIFT DATA SCIENCE の概要

Red Hat OpenShift Data Science を使用すると、ユーザーはデータ、人工知能、および機械学習ソフトウェアを統合し、エンドツーエンドの機械学習ワークフローを実行できます。OpenShift Data Science は以下の2つの設定でサポートされます。

1. Red Hat OpenShift Dedicated および Red Hat OpenShift Service on Amazon Web Services (ROSA) などの Red Hat 管理環境へのアドオンとしてインストール
2. Red Hat OpenShift Container Platform などのセルフマネージド Operator としてインストール

データサイエンティスト向けとして、OpenShift Data Science には、Jupyter と、モデル開発に必要なツールおよびライブラリーや TensorFlow および PyTorch フレームワークで最適化されたデフォルトのノートブックイメージのコレクションが含まれます。モデルのデプロイおよびホスト、モデルの外部アプリケーションへの統合、任意のハイブリッドクラウド環境でホストするためのモデルのエクスポートを行います。グラフィックスプロセッシングユニット (GPU) を使用して、データサイエンス実験を加速することもできます。

管理者向けとして、OpenShift Data Science は、既存の Red Hat OpenShift Dedicated または ROSA 環境でのデータサイエンスワークロードを有効にします。既存の OpenShift アイデンティティプロバイダーを使用してユーザーを管理し、ノートブックサーバーで利用可能なリソースを管理し、データサイエンティストがモデルの作成、トレーニング、ホストに必要なリソースを確実に入手できるようにします。

第2章 サポートされる構成

OpenShift Data Science は以下の 2 つの設定でサポートされます。

- Red Hat OpenShift Dedicated (AWS または GCP のカスタマークラウドサブスクリプション付き) または Red Hat OpenShift Service on Amazon Web Services (ROSA) 用の **マネージドクラウドサービスアドオン**。
Red Hat 管理環境での OpenShift Data Science の詳細は、[Red Hat OpenShift Data Science の製品ドキュメント](#) を参照してください。
- オンプレミスまたはパブリッククラウドに OpenShift Container Platform などの **セルフマネージド環境でインストールできるセルフマネージドソフトウェア**。
接続環境または非接続環境で OpenShift クラスターでセルフマネージドソフトウェアとして OpenShift Data Science を使用する情報は、[Red Hat OpenShift Data Science self-managed のドキュメント](#) を参照してください。

第3章 製品機能

Red Hat OpenShift Data Science は、データサイエンティストおよび IT オペレーション管理者に多くの機能を提供します。

3.1. データサイエンティスト向けの機能

コンテナ

JupyterLab のようなツールは、データサイエンティストがマシン上でモデルを直感的に開発する方法をすでに提供していますが、コラボレーションや作業の共有には常に固有の複雑さが伴います。さらに、強力な GPU などの特殊なハードウェアを使用すると、独自のハードウェアを購入して保守する必要があるため、非常に高価になる可能性があります。OpenShift Data Science に含まれる Jupyter 環境を使用すると、開発環境を必要な場所にどこにでも持ち運ぶことができます。すべてのワークロードはコンテナとして実行されるため、チームメンバーとイメージを共有したり、メンバーが使用できるデフォルトコンテナのリストにイメージを追加したりするだけでコラボレーションが簡単になります。お使いのノートパソコンでサポートできる内容に限定されなくなったため、GPU や大容量メモリーもはるかに利用しやすくなりました。

サードパーティーのマシン学習ツールとの統合

お気に入りのツールやサービスが互いにうまく連携できない状況に遭遇したことがあるでしょう。OpenShift Data Science は、柔軟性を念頭に置いて設計されています。OpenShift Data Science では、幅広いオープンソースおよびサードパーティーツールを使用できます。これらのツールは、データエンジニアリングや機能の抽出からモデルのデプロイと管理に至るまで、機械学習のライフサイクル全体をサポートします。

Git を使用したノートブックでのコラボレーション

Jupyter の Git インターフェイスを使用して、他のインターフェイスと連携し、コードの変更を適切に追跡します。

セキュアにビルドされたノートブックイメージ

モデル開発に必要なツールとライブラリーで事前設定されたノートブックイメージのデフォルトのセットから選択します。ソフトウェアスタック、特に機械学習に関係するものは、システムが複雑になる傾向があります。Python エコシステムには多数のモジュールとライブラリーがあるため、使用するライブラリーのバージョンを決定することが非常に困難になります。OpenShift Data Science には、データサイエンティストと推奨エンジンからの見解に基づいて構築され、パッケージ化されたノートブックのイメージが多数同梱されています。ランダムなアップストリームリポジトリから、証明されておらず、セキュアでない可能性のあるイメージをダウンロードするといった懸念を抱くことなく、迅速に滞りなく新規プロジェクトを開始できます。

カスタムノートブック

Red Hat および独立系ソフトウェアベンダー (ISV) によって提供およびサポートされているノートブックイメージに加えて、プロジェクトの特定の要件に対応するカスタムノートブックイメージを設定できます。

データサイエンスパイプライン

OpenShift Data Science は、データサイエンスワークロードを実行する成熟した効率的な方法のためのデータサイエンスパイプラインをサポートします。データサイエンスモデルの開発とデプロイを可能にする機械学習ワークフローを標準化および自動化できます。

モデルサービス

データサイエンティストとして、トレーニングされた機械学習モデルをデプロイし、インテリジェントなアプリケーションを実稼働環境で提供できます。モデルをデプロイまたは提供すると、モデルの関数が、テストやアプリケーションへの統合に使用できるサービスエンドポイントとして利用できるようになります。このサービスがどのように実行されるかを細かく制御できます。

3.2. IT オペレーション管理者の機能

ID プロバイダーを使用したユーザーの管理

OpenShift Data Science は、OpenShift クラスターと同じ認証システムをサポートします。デフォルトでは、アイデンティティプロバイダーにリストされているすべてのユーザーが OpenShift Data Science にアクセスできるので、OpenShift Data Science にアクセスする際に別の認証情報セットを必要としません。オプションで、ユーザーのサブセットを指定して OpenShift グループを作成し、アクセス権のあるユーザーを制限することができます。OpenShift Data Science への管理者アクセス権が割り当てられたユーザーのリストを特定する OpenShift グループを作成することもできます。

OpenShift でのリソース管理

既存の OpenShift ナレッジを使用して、OpenShift Data Science ユーザーのリソースを設定および管理します。

Red Hat 使用状況データ収集の制御

Red Hat がクラスターでの OpenShift Data Science の使用に関するデータを収集することを許可するかどうかを選択します。OpenShift クラスターに OpenShift Data Science をインストールすると、使用状況データの収集はデフォルトで有効になります。

クラスターに自動スケーリングを適用した使用コストの削減

クラスターオートスケーラーを使用してクラスターのサイズを調整し、現在のニーズを満たし、クラウドコストを最適化します。

アイドルノートブックの停止によるリソース使用状況の管理

一定期間アイドル状態だったノートブックサーバーを自動的に停止することで、OpenShift Data Science デプロイメントでのリソース使用量を削減します。

モデル監視ランタイムの実装

OpenShift Data Science は、モデル提供ランタイムをサポートします。モデル提供ランタイムは、指定されたモデルサーバーおよびそれがサポートするモデルフレームワークとの統合を提供します。デフォルトでは、OpenShift Data Science には OpenVINO Model Server ランタイムが含まれています。ただし、このランタイムがニーズを満たさない場合（たとえば、特定のモデルフレームワークをサポートしていない場合）、独自のカスタムランタイムを追加できます。

非接続環境でのインストール

OpenShift Data Science self-managed は、非接続環境でのインストールをサポートします。非接続クラスターは、通常はファイアウォールの背後などの制限されたネットワーク上にあり、インターネットに接続できません。この場合、クラスターは、Red Hat が提供する OperatorHub ソースが存在するリモートレジストリーにアクセスできません。この場合、関連するイメージをミラーリング（コピー）したプライベートレジストリーを使用して、OpenShift Data Science Operator をネットワークに接続されていない環境にデプロイします。

第4章 テスト

データサイエンティストと開発者は、[Red Hat Developer sandbox](#) 環境で OpenShift Data Science を試し、チュートリアルおよびアクティビティーにアクセスできます。

IT 運用管理者は、[60 日間の製品トライアル](#) を使用して、独自のクラスターで OpenShift Data Science を試すことができます。

第5章 取得

マネージドクラウドサービス

以下のオプションを利用して、マネージドサービスとして OpenShift Data Science をサブスクライブできます。

- OpenShift Dedicated の場合は、Red Hat 経由でサブスクライブします。
- Red Hat OpenShift Service on Amazon Web Services (ROSA) の場合は、Red Hat を通じてサブスクライブするか、AWS Marketplace を通じてサブスクライブします。

セルフマネージドソフトウェア

セルフマネージドソフトウェアとして Red Hat OpenShift Data Science を取得するには、Red Hat アカウントチームと確認して登録してください。