



# 本番利用に適した AI/ML 環境構築

のための最重要事項

# 目次

## 1 データからより多くのビジネス価値を引き出す

## 2 本番利用に適した AI/ML 環境を構築する

- 2.1 コンテナ
- 2.2 コンテナ・オーケストレーション
- 2.3 アプリケーション・ライフサイクル管理
- 2.4 MLOps プラクティス
- 2.5 ハイブリッドクラウド・プラットフォーム
- 2.6 エッジデプロイメント

## 3 オープンで柔軟な基盤から始める

## 4 成功事例を見る

## 5 AI/ML を開始する準備は整っていますか？



# データからより多くの ビジネス価値を引き出す

作成されるデータ量は、2026年までに221,000エクサバイト以上に達すると予想されています。<sup>1</sup> デジタル世界においてデータは重要な競争力となり得ますが、データ収集は出発点に過ぎず、データをどのように活用するかが真の差別化要因となります。

人工知能 (AI)、機械学習 (ML)、ディープラーニング (DL) は、データを活用してビジネス上の知見を提供し、タスクを自動化し、システム機能を向上させます。これらのテクノロジーは、顧客や従業員から開発、運用に至るまで、ビジネスのあらゆる側面を変革する可能性を秘めています。AI/ML をアプリケーションに組み込むと、次のよ

うな測定可能なビジネス成果を達成する上で役立ちます。

- ▶ 顧客満足度を向上させる
- ▶ 差別化されたデジタルサービスを提供する
- ▶ 既存のビジネスサービスを最適化する
- ▶ ビジネス運用を自動化する
- ▶ 収益を増加させる
- ▶ 意思決定を向上させる
- ▶ 効率を向上させてコストを削減する

## キーテクノロジー

この e ブックでは、実用的なデータ分析に必要なテクノロジーについて説明します。

- ▶ **人工知能 (AI)** では、機械が人間の行動を模倣し、通常であれば人間の介入を必要とするタイプのタスクを実行します。
- ▶ **機械学習 (ML)** は AI の一部であり、アルゴリズムと統計モデルを使用して、明示的な指示なしにタスクを実行できるようにします。
- ▶ **ディープラーニング (DL)** は ML の一部であり、人間の脳と同様に、レイヤーを使用して生の情報から高レベルの特徴を段階的に抽出します。例えば、生成 AI は、訓練された DL モデルに基づいて、高品質のテキスト、画像、その他のコンテンツを作成することができます。
- ▶ **機械学習運用 (MLOps)** には、クラウドネイティブ・アプリケーションで使用する AI/ML モデルの作成、トレーニング、デプロイ、モニタリング、継続的な改善に必要なすべてのツール、プラットフォーム、プロセスが含まれます。

<sup>1</sup> Dell Technologies および NVIDIA 後援の IDC ホワイトペーパー「High Data Growth and Modern Applications Drive New Storage Requirements in Digitally Transformed Enterprises」、Document #US49359722、2022年7月。

## 各業界における AI/ML のユースケース

業界を問わず、AI/ML はビジネス成果をより早く実現するのに役立ちます。



### 金融サービス

- ▶ カスタマーサービスやオファリングのパーソナライズ
- ▶ リスク分析の向上
- ▶ 詐欺やマネーロンダリングの検出



### 通信

- ▶ 顧客の行動に関する知見の取得
- ▶ カスタマーエクスペリエンスの向上
- ▶ 5G ネットワークのパフォーマンス最適化



### 小売

- ▶ サプライチェーンと在庫管理の最適化
- ▶ カスタマーインサイトおよびカスタマーエクスペリエンスの向上



### 自動車

- ▶ 自動運転テクノロジーのサポート
- ▶ 機器のメンテナンスの必要性の予測
- ▶ サプライチェーンの強化



### 医療

- ▶ 病院やクリニックでの効率の向上
- ▶ 診断の速度と精度の向上
- ▶ 患者の治療効果の向上



### エネルギー

- ▶ 現場の運用および保守の最適化
- ▶ 作業員の安全性の向上
- ▶ エネルギー取引の最適化



### 製造

- ▶ 機器の故障予測
- ▶ 予防保守の実施
- ▶ 工場フロアの安全性の向上

# 本番利用に適した AI/ML 環境を構築する

プロダクションへの AI/ML のデプロイは反復型のプロセスで、単に AI/ML モデルを作成するだけではありません。AI/ML のライフサイクルでの主な手順は次のとおりです。

1. AI/ML の取り組みに関するビジネス目標を設定し、それをすべてのステークホルダーと共有する
2. AI/ML の取り組みに必要なデータを収集して準備する
3. 目標に応じてモデルを開発する
4. アプリケーション開発プロセスにモデルをデプロイする
5. インテリジェントな ML ベースのアプリケーションを実装し、推論を開始する
6. モデルの精度を長期にわたって監視および管理する

## AI/ML のライフサイクルと MLOps



## AI/ML のデプロイの課題

AI/ML 環境の構築に取り組む組織は、いくつかの課題に直面します。

- ▶ **人材不足:** AI/ML エキスパートの数は限られているため、データサイエンティストやエンジニア、ML エンジニア、ソフトウェア開発者、その他適切な知識を持つスタッフを見つけ、確保することは困難です。
- ▶ **直ちに使用できるデータの欠如:** 企業は大量のデータを収集していますが、AI/ML プロジェクトごとに適切なデータを見つけ、準備し、保護しなければなりません。
- ▶ **統合されていないチームとテクノロジー:** 運用やインフラストラクチャが遅く、手作業で、統合されていない場合、チーム間のコラボレーションや組織全体の AI/ML デプロイメントの妨げになります。
- ▶ **リソースの可用性の遅延:** インフラストラクチャとツールの提供に時間がかかると、アプリケーション内でのモデルの開発、統合、デプロイに悪影響が及ぼされる場合があります。

しかし、クラウドネイティブのアプリケーション開発アプローチを AI/ML のライフサイクルに適用することで、これらの課題を克服することができます。

**オープンで適応性の高いアーキテクチャは、ビジネス目標を達成するために AI/ML と MLOps をより効果的に導入するのに役立ちます。**

本番利用に適した AI/ML アーキテクチャには、いくつかの主要なテクノロジーと機能が必要です。

- ▶ **AI/ML および MLOps ツール**によって、データサイエンティスト、ML エンジニア、アプリケーション開発者は、ML モデルとアプリケーションを作成、デプロイ、管理できます。
- ▶ **クラウド・プラットフォーム**により、データエンジニア、データサイエンティスト、ML エンジニア、アプリケーション開発者は、迅速に作業するために必要なリソースを利用できます。
- ▶ **コンピューティング、ストレージ、ネットワーク・アクセラレーター**は、データの準備、モデル開発、推論タスクを高速化します。
- ▶ **インフラストラクチャ・エンドポイント**は、オンサイト、仮想、エッジ、プライベート、パブリック、およびハイブリッドの各クラウド環境において、AI/ML 運用のすべての段階でリソースを提供します。
- ▶ **エッジデプロイメント (オプション)** は、デバイスやセンサーからの大量のデータを提供するため、リアルタイムでモデルを訓練し、知見を得るために使用することができます。

この e ブックでは、効果的な AI/ML アーキテクチャを構築するための重要な考慮事項について説明します。

# コンテナ

**コンテナ**はソフトウェアの基本単位であり、あらゆる依存関係とともにアプリケーションをパッケージ化します

データサイエンティスト、ML エンジニア、アプリケーション開発者が最大限に生産性を高めるためには、自分の好みのツールとリソースを利用できることが必要です。同時に、IT 運用チームは、リソースが最新で、コンプライアンスに準拠し、安全に使用されるようにしなくてはなりません。コンテナを使用することでアプリケーションのビルドプロセスは単純化され、変更を加えることなくアプリケーションを異なる環境にデプロイできるようになります。コンテナを使用すると、ハイブリッド環境全体に一貫した方法でさまざまな AI/ML ツールをデプロイできます。チームは、変更を追跡して透明性を実現するバージョン管理機能を使用して、コンテナイメージを繰り返し変更し、共有することができます。同時に、プロセスの分離とリソース制御により、脅威からの保護が向上します。

## コンテナソリューションに関する推奨事項

統合されたセキュリティ機能が含まれ、環境間でのコンテナのデプロイ、管理、移動の複雑性を軽減する、堅牢で可用性の高いコンテナ・プラットフォームを選択しましょう。幅広いテクノロジーと統合できるオープンソース・プラットフォームを選択すれば、柔軟性と選択肢を増やすことができます。

# コンテナ・オーケストレーション

コンテナのオーケストレーションにより、環境全体におけるコンテナの作成、デプロイ、ライフサイクルが管理されます

コンテナを導入したら、それらを効率的にデプロイ、管理、スケーリングする方法が必要です。コンテナ・オーケストレーション・ツールを使用すると、コンテナのライフサイクルを一貫した方法で管理できます。これらのツールは通常、オンサイト、エッジ、およびクラウド環境間で、コンピューティング、ストレージ、ネットワークリソースへのアクセスを一元化します。また、ワークロード・スケジューリング、マルチテナンシー制御、クォータの適用も統合します。

## コンテナ・オーケストレーションに関する推奨事項

優れたオープンソース・テクノロジーを活用するには、**Kubernetes** ベースのオーケストレーション・ツールを選択しましょう。

# アプリケーションのライフサイクル管理

アプリケーションのライフサイクル管理には、コンテナ内で実行されるアプリケーションのデプロイ、スケーリング、管理が含まれます。

AI/ML 環境は本質的に複雑です。コンテナ・オーケストレーション・ツールと統合できるコンテナ・アプリケーションのライフサイクル管理コンポーネントがあれば、AI/ML 開発ツールを含むコンテナ化アプリケーションを直接管理できます。IT 運用チームは、構成、プロビジョニング、更新などの一般的なライフサイクル管理タスクを自動化して、効率、スピード、精度を高めることができます。データサイエンティスト、ML エンジニア、アプリケーション開発者は、IT チームの手を借りることなく、事前に承認されたサービスカタログからツールとアプリケーションを使用できます。自動化により、スタッフは面倒なタスクから解放され、より有益で戦略的な作業に専念できるようになります。

## アプリケーション管理に関する推奨事項

使いやすい自動化機能や好みの AI/ML ツールとの統合機能を備えたコンテナ・アプリケーションのライフサイクル管理ツールを選択しましょう。よく使われるツールには、**Kubernetes Operator** や **Helm Chart** があります。

# MLOps プラクティス

MLOps プラクティスは、AI/ML を大規模に運用するために必要なツール、プラットフォーム、プロセスを統合します。

組織は、AI/ML モデル（およびそれらを使用するアプリケーション）を迅速かつ効率的に開発してデプロイする必要があります。このような取り組みを成功させるためには、チーム間の連携が不可欠です。**DevOps** と同様に、MLOps のアプローチは、AI/ML チーム、アプリケーション開発者、IT 運用間のコラボレーションを促進し、ML モデルと ML ベースのアプリケーションの作成、トレーニング、デプロイ、管理を加速します。自動化（多くの場合、**継続的インテグレーション/継続的デリバリー (CI/CD)** パイプラインの形式）により、モデルおよびアプリケーション開発のライフサイクルを短縮するために、迅速で段階的な反復型の変更を行うことが可能になります。

## MLOps のベストプラクティス

MLOps はテクノロジーだけの問題ではなく、人とプロセスが重要な役割を果たします。AI/ML のライフサイクル全体に **MLOps プラクティス** を適用しましょう。Argo、Kubeflow、Tekton、Jenkins のようなオープンソース・テクノロジーとともに、プラットフォームとツールの自動化を使用して、CI/CD パイプラインとワークフローを構築しましょう。

# ハイブリッドクラウド・プラットフォーム

ハイブリッドクラウド・プラットフォームは、オンサイト、エッジ、およびクラウド環境で、インテリジェント・アプリケーションおよびモデルの開発、デプロイ、管理を行うための基盤になります。

AI/ML モデルとインテリジェント・アプリケーションには、開発とデプロイのためのインフラストラクチャが必要です。一貫性のあるハイブリッドクラウド・プラットフォームを使用すると、インフラストラクチャのあらゆる部分で、モデルとアプリケーションの開発、テスト、デプロイ、管理を同じ方法で行えます。ハイブリッドクラウド・プラットフォームは AI/ML 環境をオンデマンドでプロビジョニングするための可搬性、拡張性、柔軟性を提供します。また、IT 制御を維持しながら、リソース提供を高速化できるセルフサービス機能も得られます。そして、一貫性のあるプラットフォームは、サードパーティベンダー、オープンソース・コミュニティ、および使用する可能性のあるカスタム開発ツールによるテクノロジー統合の基盤を提供します。

## ハイブリッドクラウド・プラットフォームに関する推奨事項

ハードウェア・アクセラレーション、AI/ML とアプリケーション開発ツールの広範なエコシステム、統合された DevOps と管理機能をサポートする、セキュリティ重視のプラットフォームを選択しましょう。オープンソース・プラットフォームは、統合の機会と柔軟性を高めることができます。

# エッジデプロイメント

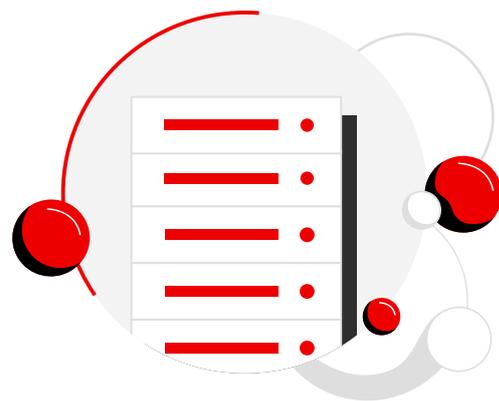
**エッジデプロイメント**は、コアデータセンターの外にある現場でデータを収集し、機能を実行するデバイスを備えた分散型環境です。

**エッジコンピューティング**は、必要とされる場所と瞬間に、知見とエクスペリエンスを提供することができます。センサーやデバイスは、AI/ML ワークフローでモデルのトレーニングやランタイムの推論に使用できる大量のデータを生成します。このようなデータを中央のクラウドにリアルタイムで送信するのは難しく、コストもかかります。例えば、画像認識アルゴリズムはデータソースに近いほど効率的に実行され、処理のために大量のデータを転送する必要がなくなります。

## エッジのベストプラクティス

エッジデプロイメントには、スケーラビリティ、接続性、デバイス管理が不可欠です。データセンターやクラウド・インフラストラクチャと同じツールやプロセスで管理できるソリューションを探しましょう。通信が断続的にしつかつながらない、あるいはネットワーク接続がない環境に対応できるプラットフォームは極めて重要です。最後に、多くのデバイスとハードウェアのフットプリントをサポートするソリューションは、より高い柔軟性とカスタマイズ性をもたらします。

# オープンで柔軟な基盤 から始める



Red Hat は、AI/ML に関する組織の目標達成を支援する、完全なテクノロジー・ポートフォリオ、実証済みの専門知識、戦略的パートナーシップを提供します。また、本番利用に適した AI/ML 環境を構築するための基盤と、迅速な導入を実現するためのサービスおよびトレーニングを提供します。

**Red Hat® OpenShift®** は、クラウドネイティブのイノベーションを実現する、統合されたエンタープライズ向けアプリケーション・プラットフォームです。オンデマンドのコンピュータリソース、ハードウェアおよびグラフィックス・プロセッシング・ユニット (GPU) のアクセラレーションのサポート、オンサイト、パブリッククラウド、およびエッジ環境間の一貫性により、成功するために必要な速度と柔軟性がもたらされます。例えば、セルフサービス MLOps プラットフォームを作成し、そこでデータサイエンティスト、データエンジニア、開発者は迅速にモデルを構築し、そのモデルをアプリケーションに組み込み、推論タスクを実行できます。コラボレーション機能は、チームがコンテナ化されたモデリング結果を作成し、同僚や開発者と一貫した方法で共有することを可能にします。

Red Hat OpenShift AI は、AI/ML モデルとアプリケーションのトレーニング、提供、監視、ライフサイクル管理を支援する製品ポートフォリオです。この製品に含まれる **Red Hat OpenShift Data Science** は、データサイエンティストと開発者に対し、インサイトを集めてインテリジェント・アプリケーションを構築するための強力な AI/ML プラットフォームを提供します。チームは、NVIDIA、インテル、Starburst、Anaconda、IBM、Pachyderm などの主要な認定パートナー製品を統合したコラボレーティブで一貫性のある環境で、実験からプロダクションへと移行することができます。

**Red Hat Application Services** のポートフォリオは、アプリケーションの開発、提供、統合、および自動化のための統一された環境の作成に役立ちます。データ統合サービスは効果的なデータパイプラインの構築に役立ち、ランタイムサービスはアプリケーション開発を単純化します。プロセス自動化ツールおよびサービスにより、インテリジェント・アプリケーションと ML モデルにアクセスして、ビジネス上の意思決定を自動化することができます。

最後に、**Red Hat Enterprise Linux®**、**Red Hat OpenStack® Platform**、および **Red Hat OpenShift Platform Plus** などの Red Hat プラットフォーム製品は、スケーラブルなソフトウェア・デファインド・インフラストラクチャを提供します。

## コミュニティとともに構築

Red Hat は、**Kubeflow** および **Open Data Hub** のオープンソース・コミュニティに積極的に参加しています。Open Data Hub は、一般的なオープンソースの AI/ML ツールを OpenShift 環境に統合するためのブループリントを提供するコミュニティプロジェクトです。リファレンス・アーキテクチャには、一般的なデータ分析ツールと機械学習ツール (Ray、Ceph®、Apache Kafka、Kubeflow、TensorFlow、Jupyter notebooks など) が統合されています。

## 認定パートナー・エコシステムで柔軟性を得る

Red Hat の認定パートナー・エコシステムを利用すれば、一般的な AI/ML、データ分析、管理、ストレージ、セキュリティ、開発ツールをこのアーキテクチャに統合できます。Red Hat はパートナーと緊密に連携して Red Hat プラットフォーム上で各社のソフトウェアを認定しており、優れた管理性、セキュリティ、サポートが実現します。多くのパートナーは、ソフトウェアのライフサイクル管理を単純化するために、認定 Red Hat OpenShift Operator も提供しています。

## 製品とテクノロジーの幅広い選択肢

Red Hat は、認定 AI/ML パートナーのエコシステムの拡大を促進しているため、広く普及している製品やテクノロジーをお客様の環境に取り入れることができます。

**NVIDIA** と Red Hat は、AI を活用したインテリジェント・アプリケーションを環境全体で迅速に提供するソリューションを提供しています。**NVIDIA AI Enterprise with Red Hat OpenShift** は、AI とデータ分析ソフトウェアの完全で最適化されたクラウドネイティブのセットを提供します。Red Hat Enterprise Linux、Red Hat OpenShift、および NVIDIA DGX システムは、AI インフラストラクチャの IT 管理性を実現します。**NVIDIA GPU Operator** は、GPU のプロビジョニングに必要なすべての NVIDIA ソフトウェア・コンポーネントの管理を自動化します。

**インテル** と Red Hat は協力して、データセンターのアジリティと柔軟性を向上させるソフトウェア・デファインド・インフラストラクチャと業界標準のプラットフォームを提供します。インテルの **OpenVINO ツールキット** のディストリビューションは、DL モデルを最適化し、Red Hat OpenShift 上で数千ノードまで自動的に拡張できる高性能推論エンジンに変換します。**Intel AI Analytics Toolkit powered by oneAPI** は、ML ワークフローを高速化し拡張するための相互運用可能な AI ソフトウェア・ツール一式を提供します。

**Starburst** と Red Hat は、分散したデータソースから知見を引き出すお手伝いをします。**Starburst Enterprise** は Red Hat OpenShift と連携し、複数のプラットフォームにわたるデータを迅速に分析します。この組み合わせにより、エンタープライズグレードの自動化、高可用性 (HA)、弾力性、監視機能が実現します。このソリューションにより、データをモダン化し、抽出、変換、ロード (ETL) ワークロードを効率的に実行し、インタラクティブなデータ調査を実行し、ビジネスインテリジェンスツールに情報を提供することができます。

**SAS** と Red Hat は、ビジネスレベルのインテリジェンスを提供するオープン・ハイブリッドクラウド・テクノロジーと分析機能を共同で開発しています。**SAS Viya on Red Hat OpenShift** は、最先端の分析、ML、AI アプリケーションをハイブリッドクラウド・プラットフォームと組み合わせ、一度構築すればどこにでもデプロイできるようにします。インフラストラクチャ全体で一貫した管理は、チームを結びつけ、コラボレーションを促進します。この統一されたプラットフォームにより、好みのインターフェース、言語、インフラストラクチャを使用してモデルを開発し、デプロイすることができます。

# 成功事例を見る



**Banco Galicia** は Red Hat コンサルティングと連携し、Red Hat OpenShift、Red Hat Integration、および Red Hat のシングルサインオン上に AI ベースのインテリジェントな自然言語処理 (NLP) ソリューションを構築しました。

法人顧客のオンボーディング時間を

## 20 日間から数分に短縮

さらにデータ分析の精度 90% を実現

[お客様導入事例](#)を読む



**東日本電信電話株式会社 (NTT 東日本)** は、Red Hat OpenShift を使ってエッジコンピューティングによるデータ分析サービスを構築しました。

「[...] Red Hat OpenShift により、AI 開発者とのコラボレーションによる革新的な映像 AI サービスを安定して開発、運用することが可能になりました。」

**外山 将司氏**

NTT 東日本 ネットワーク事業推進本部 高度化推進部  
クラウドサーバ技術部門 サーバ基盤技術担当 担当課長

[お客様導入事例](#)を読む

## アメリカ合衆国退役軍人省

**アメリカ合衆国退役軍人省の Team Guidehouse** は、Red Hat OpenShift と Red Hat OpenShift Data Science をデプロイして、退役軍人の自殺を防ぐプロトタイプ・ソリューションで機械学習の技法を使用しました。

## フェーズ 2 の勝者

Mission Daybreak の課題

[ブログ記事](#)を読む



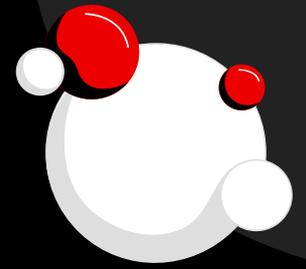
**ボストン大学**では、コンピュータサイエンスおよびコンピュータエンジニアリングシステムコースの主要な授業用プラットフォームとして Red Hat OpenShift Data Science を使用しています。

「このプラットフォームは非常に細やかながらも簡単にアクセスでき、私の教材や方法論に統合できる、豊かで本格的な Linux エクスペリエンスを生徒に提供しています。」

**Jonathan Appavoo 氏**

ボストン大学准教授

[ブログ記事](#)を読む



# AI/ML を開始する 準備は整っていますか？

AI/ML と MLOps は、ビジネスのほぼあらゆる側面を変革しています

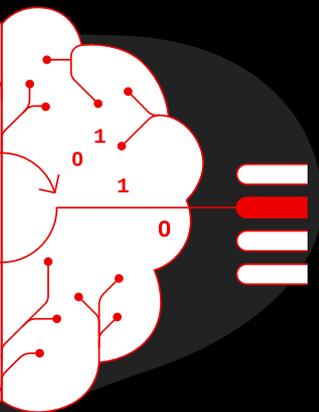
Red Hat は、組織のビジネス目標をサポートするインテリジェント・アプリケーションの開発と提供を高速化する、本番利用に適した AI/ML 環境の構築を支援します。

## Red Hat OpenShift AI

Red Hat OpenShift がどのように AI/ML ワークフローを加速させ、AI ベースのインテリジェント・アプリケーションを提供できるかをご覧ください：[red.ht/OpenShiftAI](https://red.ht/OpenShiftAI)

## Red Hat OpenShift Data Science

Red Hat OpenShift Data Science が MLOps のベストプラクティスの確立にどのように役立つかをご覧ください：[red.ht/datascience](https://red.ht/datascience)



## Red Hat コンサルティングに今すぐご相談ください

Red Hat のエキスパートと連携し、AI/ML プロジェクトをスピーディに立ち上げましょう。Red Hat は、AI/ML 導入の加速を支援するコンサルティングおよびトレーニングサービスを提供しています。

- ▶ AI/ML サービスの詳細：[red.ht/aiml-consulting](https://red.ht/aiml-consulting)
- ▶ 無料のディスカバリー・セッションを予約：[red.ht/consulting-ja](https://red.ht/consulting-ja)

Copyright © 2023 Red Hat, Inc. Red Hat, Red Hat ロゴ, OpenShift, および Ceph は、米国およびその他の国における Red Hat, Inc. またはその子会社の商標または登録商標です。Linux® は、米国およびその他の国における Linus Torvalds 氏の登録商標です。OpenStack® ワードマークと Square O Design は個別に、または一体として米国とその他の国における OpenStack Foundation の商標または登録商標であり、OpenStack Foundation の許諾の下で使用されています。Red Hat は、OpenStack Foundation と OpenStack コミュニティのいずれにも所属しておらず、公認や出資も受けていません。

479615\_0823\_KVM

