



1. イントロダクション

Azure Machine Learning – 構築・運用編

Keita Onabuta

FastTrack for Azure
Senior Customer Engineer for AI/ML

Agenda

機械学習サービスの選択
Azure Machine Learning 基本

機械学習サービスの選択

基礎知識

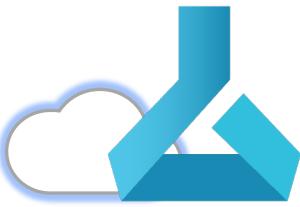
- Azure Machine Learning
- Azure Databricks
- Azure Data Science VM

ガイドライン・実装手順

- Azure Machine Learning と Azure Databricks

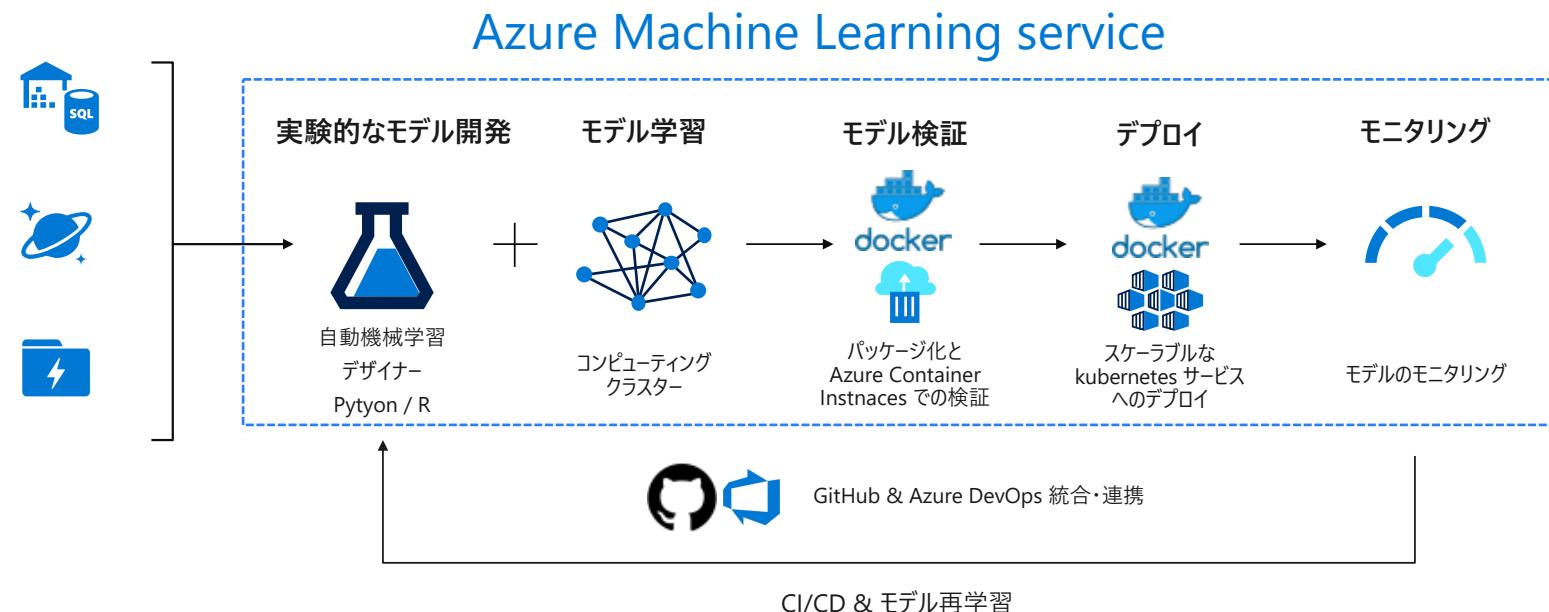
検討事項

- 利用する機械学習サービスの検討



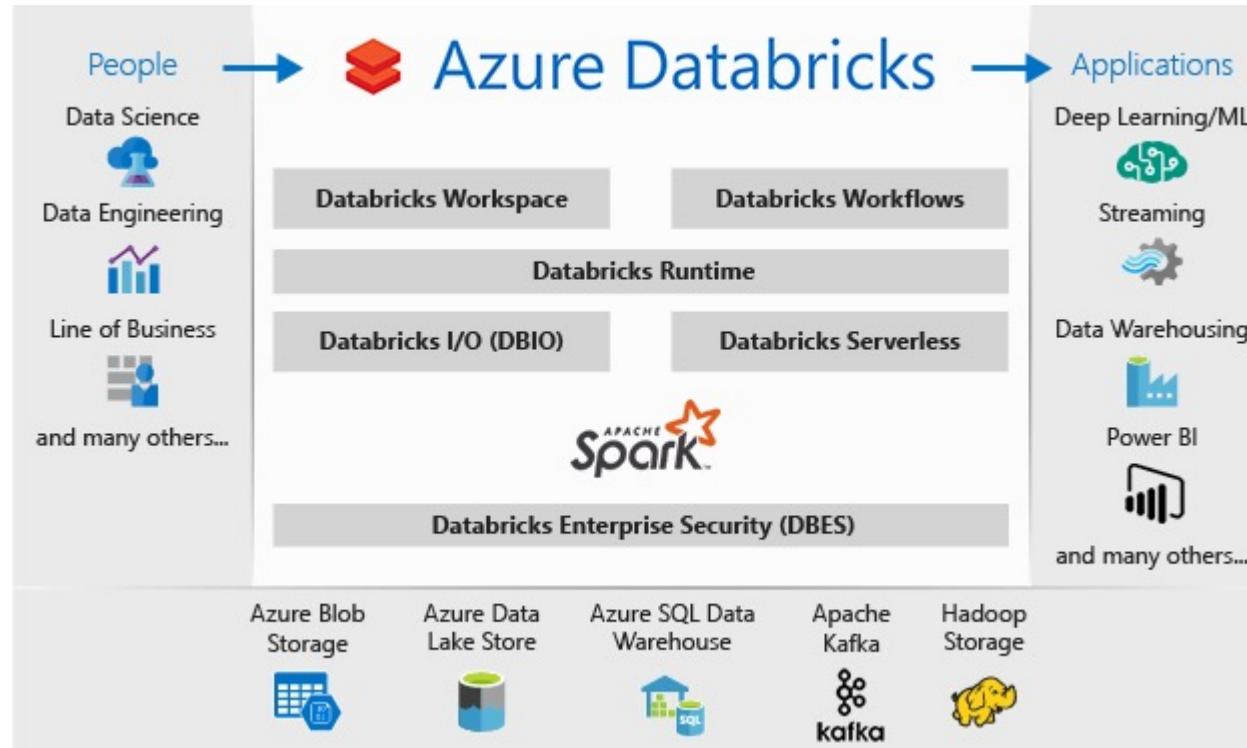
Azure Machine Learning

- ・機械学習プロセスをエンドツーエンドでサポートするマネージドサービス
 - ・必要なシステムモジュールをあらかじめビルトインしている
- ・自動機械学習やパラメータチューニング機能による効率的なモデル開発
- ・継続的なモデルのデプロイ & 運用管理をサポート
- ・スケーラブルな計算環境による並列分散処理 etc



Azure Databricks

迅速、簡単で協調的な Apache Spark ベースの分析プラットフォーム



できること

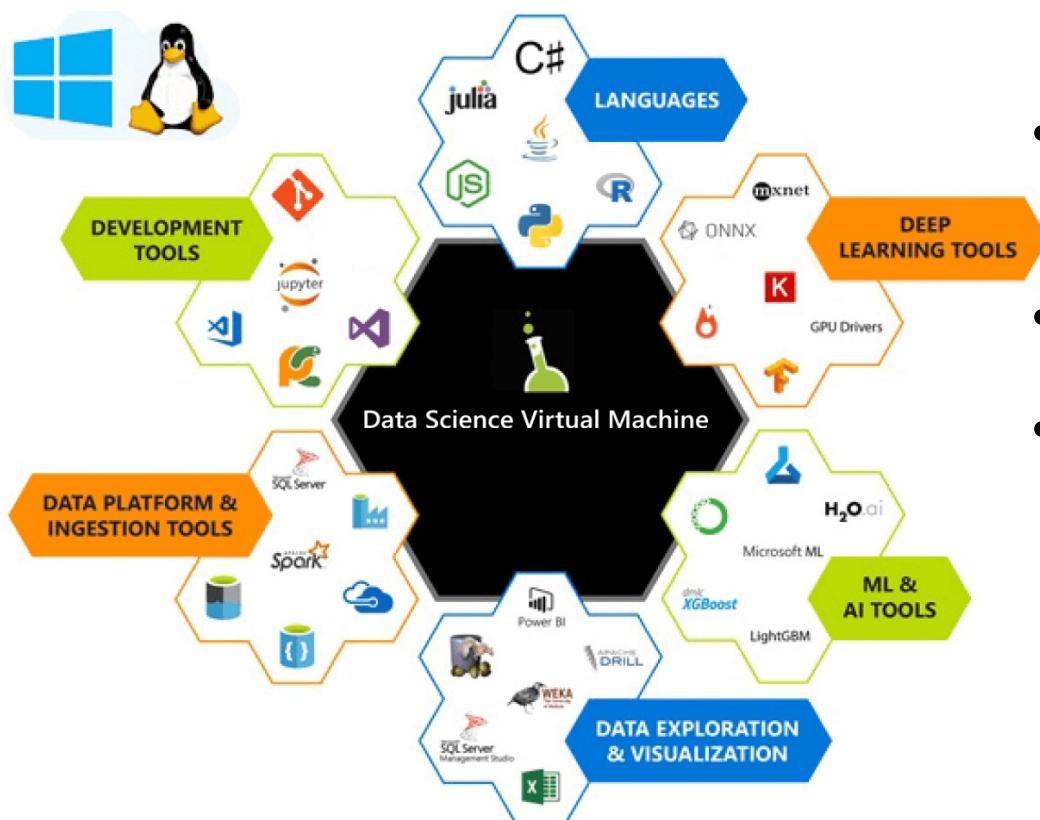
Azure Active Directoryと連携した認証・アクセス制御により、セキュアな環境での機械学習をインメモリ分散テクノロジー Spark をベースに高速に実現できます

お客様のメリット

Azure Active Directory と連携したクラウド内の完全に管理された Apache Spark クラスターを迅速に立ち上げ、インフラを意識することなく分析者は共同分析作業に集中できます。また、処理に応じて柔軟にスケールを変更することができ、これまで分析に要していた時間を短縮することにより様々なアプローチに取り組むことができます。

Azure Data Science VM

データ分析、機械学習、AI トレーニングでよく使用されるツールを事前インストール、構成、テストされた Azure 仮想マシンイメージ

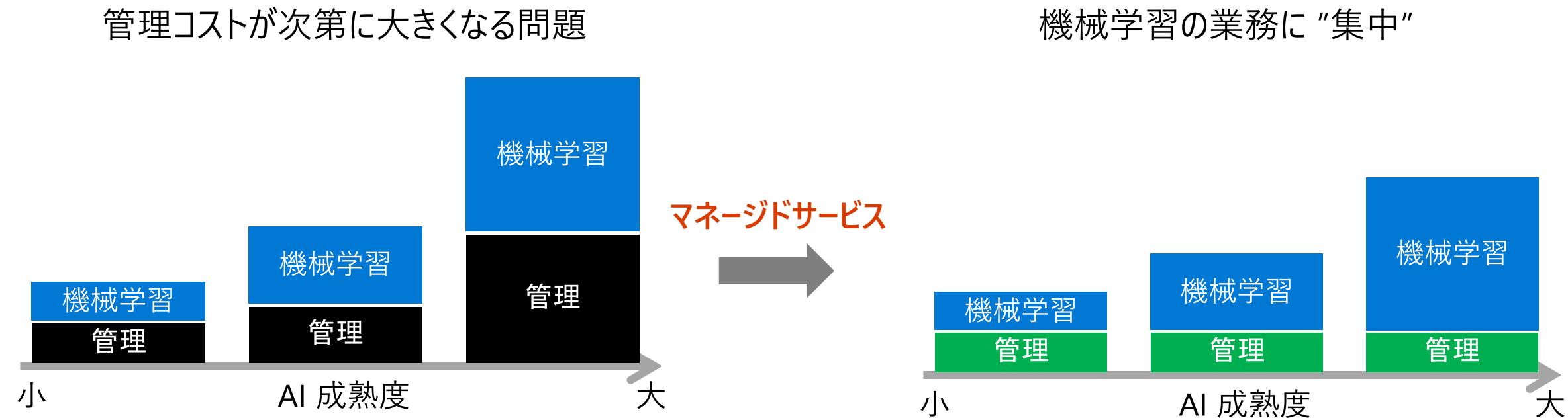


- 任意の VM インスタンスで稼働
 - ✓ GPU のドライバも同梱
- OS は Windows / Linux 選択から可能
- PyTorch / TensorFlow などの主要 Deep Learning フレームワークも事前インストール済み

<https://docs.microsoft.com/ja-jp/azure/machine-learning/data-science-virtual-machine/overview>

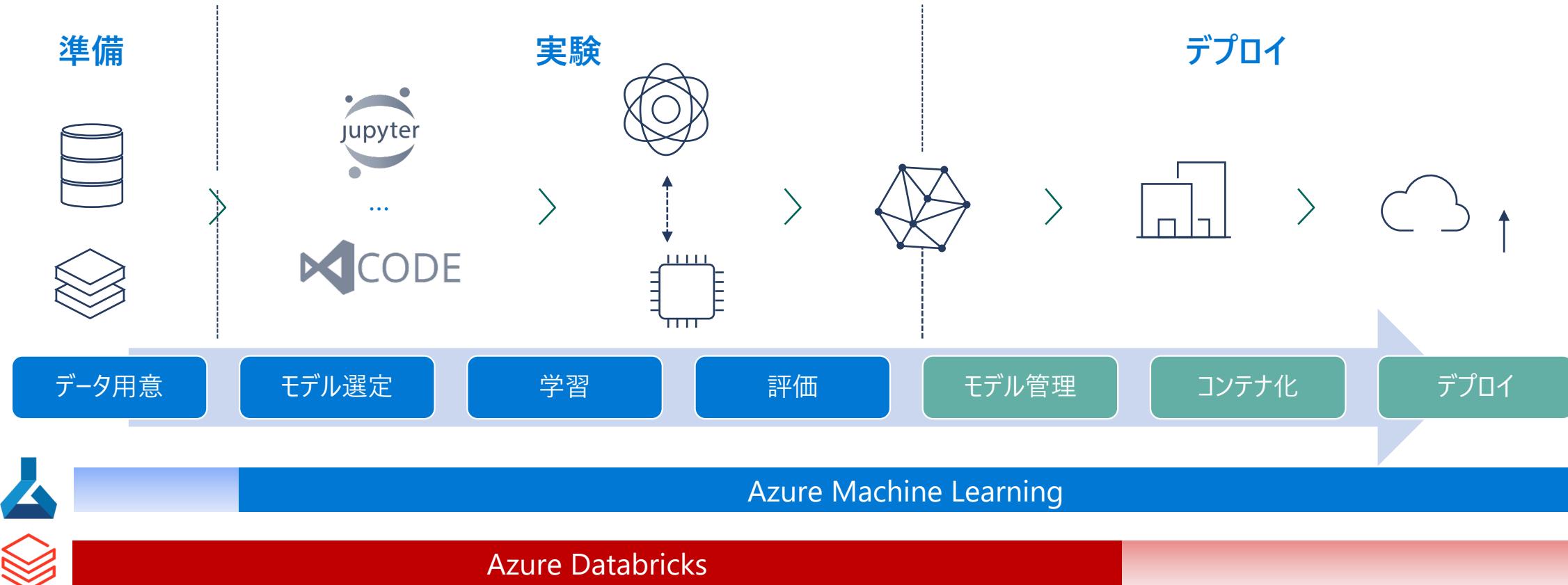
利用する機械学習サービスの検討

機械学習プラットフォームを構成する方法は多数あります。社内のエンジニアリソースが豊富にある場合はスクラッチで開発することも選択肢に上がりますが、機械学習を構成する要素は多岐に渡っており簡単ではありません。極力マネージドサービス (PaaS) を利用することで管理コストを軽減することができます。



Azure Machine Learning と Azure Databricks

Azure Databricks は Spark を利用した強力なデータの前処理をスムーズに行うことができます。Azure Machine Learning はデプロイに強みを持っており、様々なリソースとの連携が可能です。



Azure Machine Learning と Azure Databricks (cont'd)

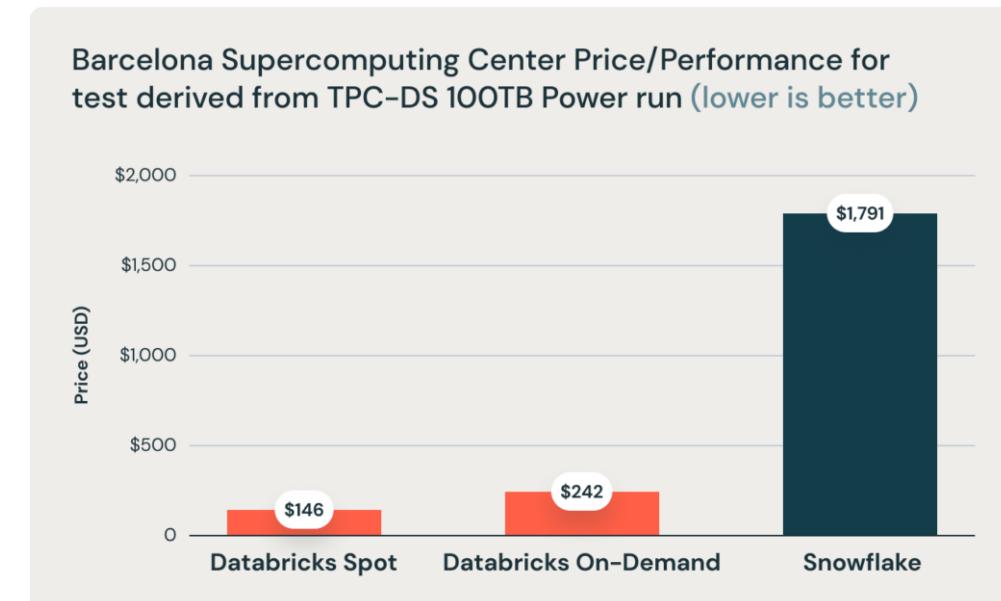
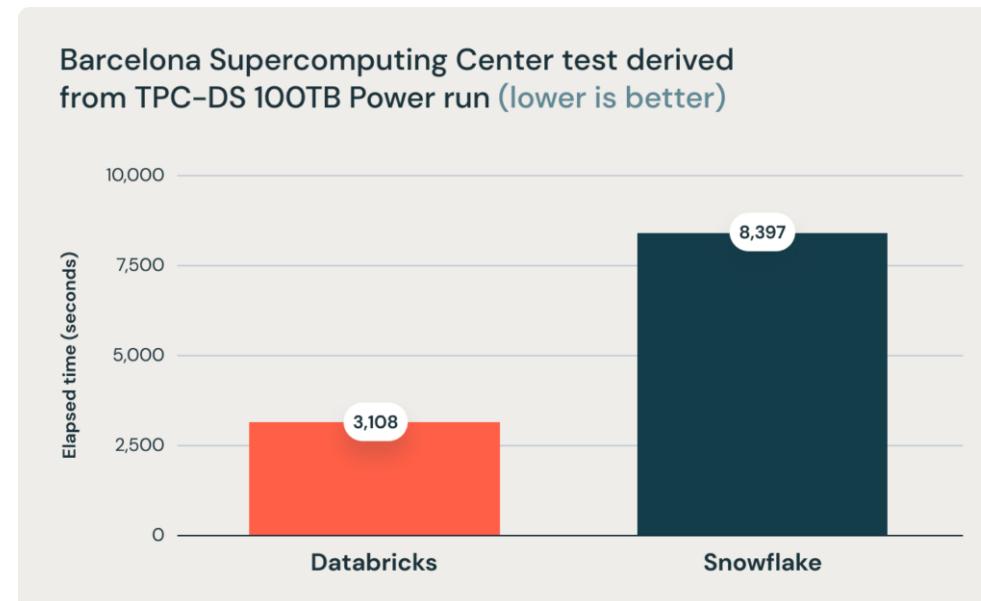
Azure ML の強み

- 開発環境の充実
 - 市場での認知度が高く実績が豊富な Jupyter, JupyterLab, R Studio, Visual Studio Code がシームレスに利用可能
 - オープンソーステクノロジーとの親和性が高くカスタマイズが可能
 - ローカル環境をスムーズに移行
- Azure Kubernetes Service との連携
 - Auto ML モデルを直接本番水準のクラスターにデプロイして API として利用可能
 - エントリスクリプトや設定ファイルの用意のみで任意のモデルを API としてデプロイ可
- パイプライン
 - Azure Data Factory パイプラインから Azure ML パイプラインを呼び出し可能
 - GUI を使用したパイプライン構築のサポート
 - デプロイまで含めたモデル作成の全工程のカバー
 - Python スクリプトやノートブックをパイプラインに組み込み可能

Azure Machine Learning と Azure Databricks (cont'd)

Azure Databricks の強み

- 大規模なデータに対する処理能力
 - Spark ベースのチューニングされた分散処理の仕組みによる圧倒的処理能力
 - ノートブックを使用してインタラクティブに分散処理の実行が可能



Azure Machine Learning 基本

基礎知識

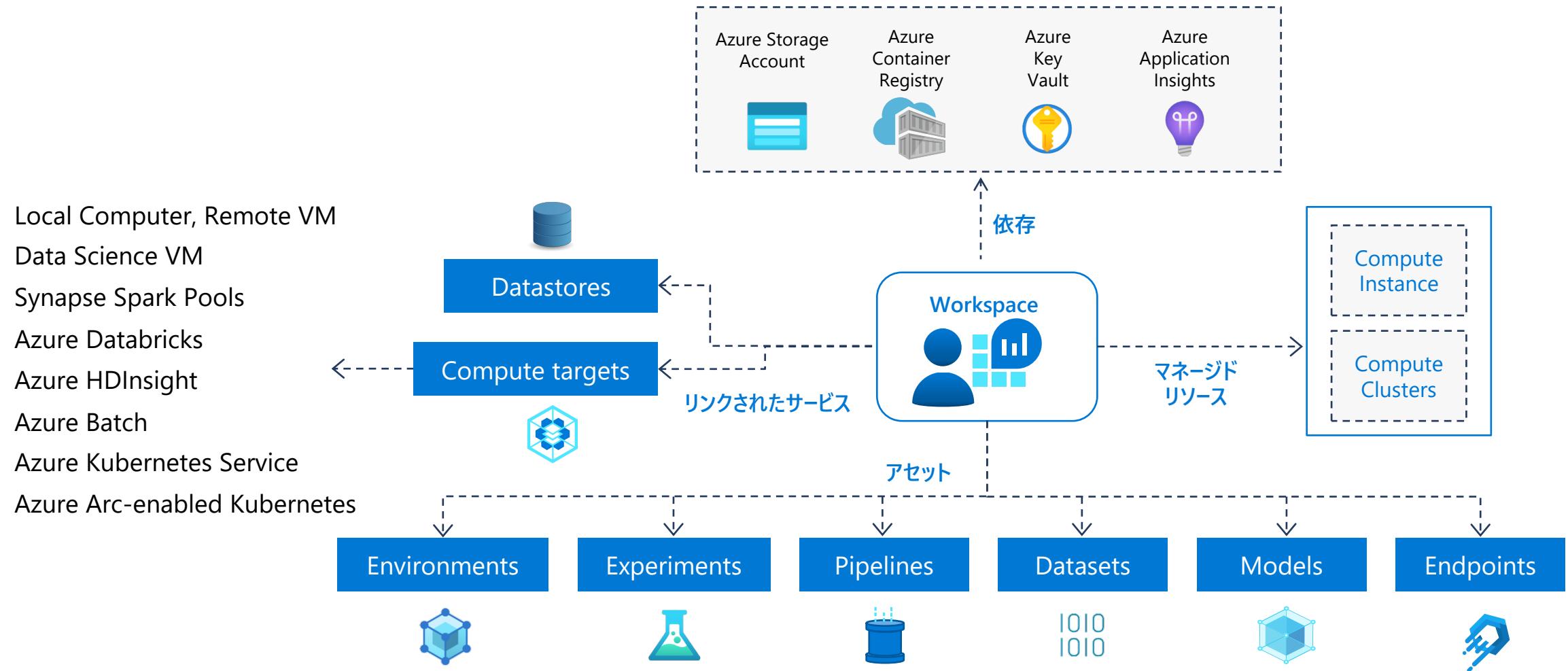
- Azure Machine Learning 基本構成

ガイドライン・実装手順

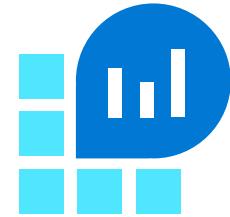
- ワークスペース構成のガイドライン
- クイックスタートテンプレート

Azure Machine Learning 基本構成

Azure Machine Learning Workspae は、最上位のリソースであり、あらゆるリソースを操作・運用管理するための一元的な環境を提供します。



Azure Machine Learning 基本構成 (cont'd)

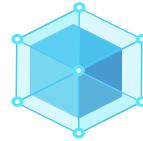


Workspace



データストア & データセット (Datastores & Datasets)

モデル学習で利用するデータの運用管理



モデル (Models)

学習済み機械学習モデルを運用管理



コンピューティング (Computes)

モデル学習や推論で利用する計算環境



パイプライン (Pipelines)

モデル学習や推論のプロセスをパイプライン化



環境 (Environments)

Python パッケージ、Docker イメージを管理



エンドポイント (Endpoints)

クラウドやエッジデバイスにモデルをデプロイ



実験と実行 (Experiments & Runs)

モデル学習時のメトリックやログを管理



データラベリング (Data Labelling)

画像やテキストへラベルを付与

ワークスペース構成のガイドライン

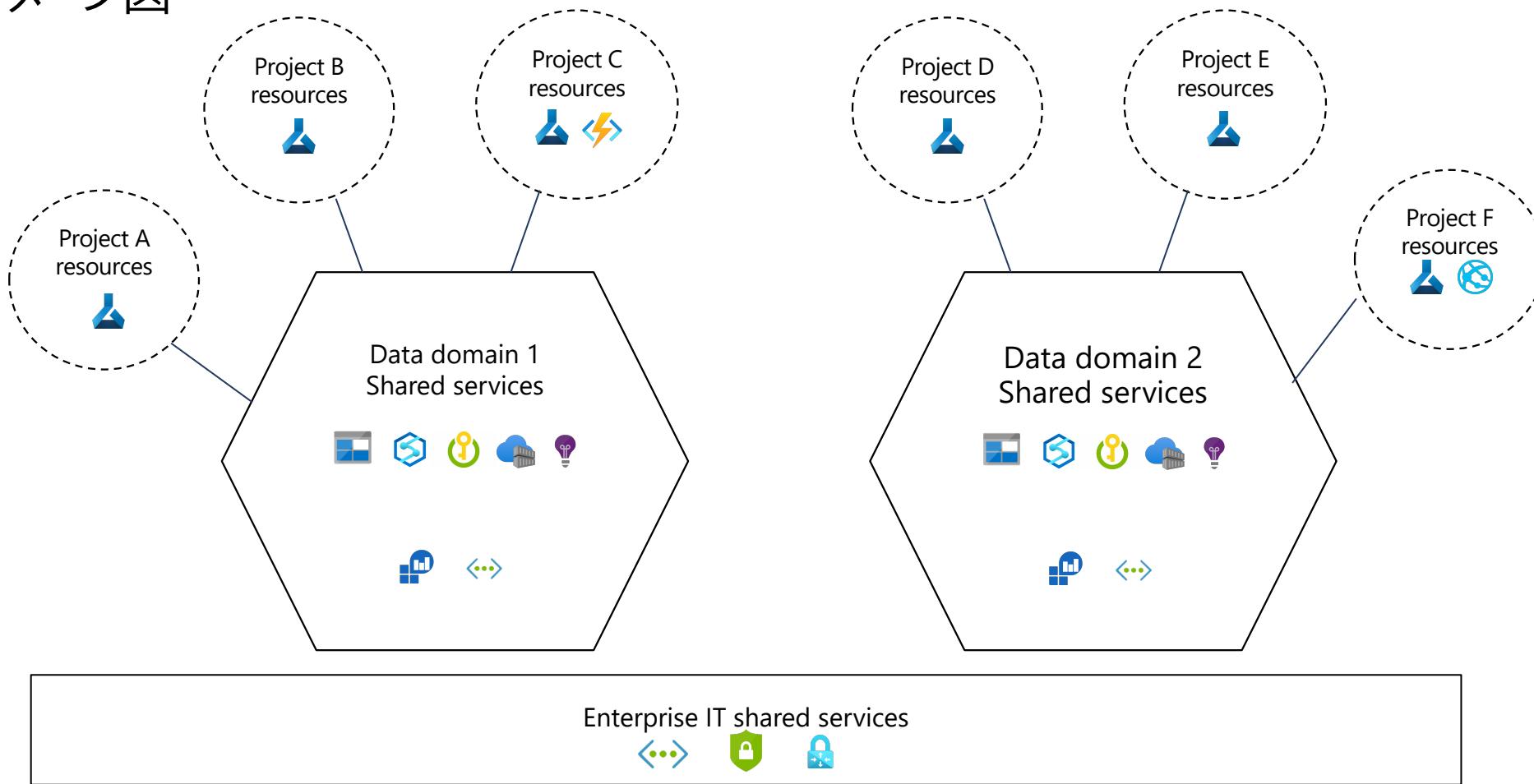
チームやプロジェクトで Azure ML を利用する場合、企業で 1 つの Azure ML Workspace では無く、下記の観点から複数の Workspace を構築・運用管理することが推奨です。

ポイント

- セキュリティ
 - Azure ML Workspace 内部で細かいアクセス制御ができない (2022年4月時点)。
 - データソースに対しては Identity ベースのアクセス設定が可能。
 - Notebooks は File Share にコード・ノートブックを保持しており、アクセス制御に対応していない。
- リージョン
 - チームやプロジェクトの拠点が地理的に散在しているケースにおいてはアクセス速度や法的な理由から Workspace をリージョン毎に構築することが望ましいケースがある。
- ワークロード
 - 開発環境、テスト環境、本番環境などワーカロードに応じて Workspace を分ける。

ワークスペース構成のガイドライン (cont'd)

構成のイメージ図



クイックスタートテンプレートの活用

Bicep もしくは Terraform のテンプレートが準備されており、
クイックにセキュアなリソースを作成することが可能です。

- ベストプラクティスが実装されたパイロット環境がクイックに作成可能
- テンプレートは要件に応じてパラメータ変更が可能
- 利用しているクラウド環境が Azure のみの場合は、Bicep テンプレートの利用を推奨
- マルチクラウド環境の場合は、Terraform テンプレートを利用するケースが多い

Bicep → [Bicep/ARM quickstart](#)

Terraform → [Terraform quickstart](#)

The screenshot shows the Microsoft Azure (Preview) portal interface. At the top, there's a navigation bar with 'Microsoft Azure (Preview)', a search bar, and a user profile icon. Below the header, the page title is 'Azure Machine Learning end-to-end secure setup' under 'Azure quickstart template'. There are two tabs: 'Basics' (which is selected) and 'Review + create'. Under the 'Template' section, there's a preview icon showing 'machine-learning-end-to-end-secure' with '9 resources'. To the right of the preview are three buttons: 'Edit template', 'Edit parameters', and 'Visualize'. The 'Project details' section contains fields for 'Subscription' (set to 'FTA keonabut - Azure CXP Internal') and 'Resource group' (set to 'Create new'). The 'Instance details' section includes fields for 'Region' (set to 'East Asia'), 'Prefix' (empty), 'Location' (set to '[resourceGroup().location]'), 'Tags' (empty), 'Vnet Address Prefix' (set to '192.168.0.0/16'), 'Training Subnet Prefix' (set to '192.168.0.0/24'), 'Scoring Subnet Prefix' (set to '192.168.1.0/24'), 'Azure Bastion Subnet Prefix' (set to '192.168.250.0/27'), 'Deploy Jumphost' (set to 'true'), 'Dsvm Jumpbox Username' (empty), 'Dsvm Jumpbox Password' (redacted), and 'Aml Compute Public Ip' (set to 'true').

[テンプレートを使用してセキュリティで保護されたワークスペースを作成する - Azure Machine Learning | Microsoft Docs](#)



Microsoft AI





© Copyright Microsoft Corporation. All rights reserved.