



Azure Machine Learning 基礎編

FastTrack for Azure – Live Session

Keita Onabuta

FastTrack for Azure
Senior Customer Engineer for AI/ML

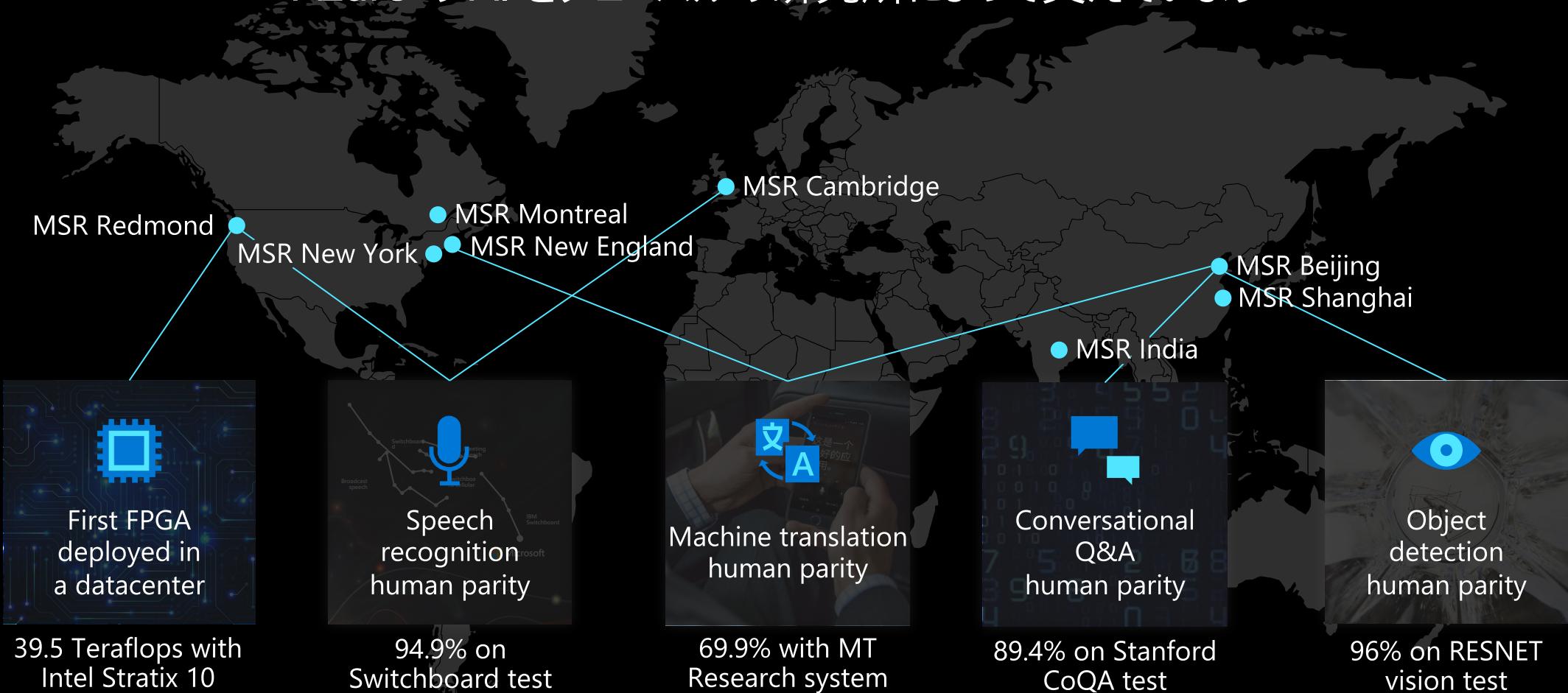
Agenda

- イントロダクション – Azure AI
- Demo : Azure Machine Learning サービス作成
- Demo : Azure Machine Learning studio
- Demo : Azure Machine Learning Code First
- 参考情報

イントロダクション – Azure AI

Microsoft AI

Azure の AI をグローバルの研究所によって支えています



Azure AI でお客様の課題を解決

AI を活用して
一般的なビジネス シナリオを
最新化したい

AZURE APPLIED AI SERVICES

Azure Cognitive
Search

Azure Form
Recognizer

Azure Video
Analyzer

Azure Bot
Service

Azure Metrics
Advisor

Azure Immersive
Reader

私のチームでは、
共通コグニティブ スキルを実現する
ために事前構築済みの
高品質 AI モデルが必要である

AZURE COGNITIVE SERVICES

視覚

音声

言語

意思決定

ML モデルの大規模な展開と
管理は困難

AZURE MACHINE LEARNING

機械学習モデルを構築、トレーニング、展開するためのエンドツーエンドのプラットフォーム

Azure AI サービススタック

Azure Applied AI Services



Azure Cognitive
Search



Azure Form
Recognizer



Azure Bot
Service



Azure Video
Analyzer



Azure Metrics
Advisor

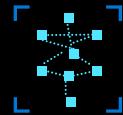


Azure Immersive
Reader

Azure Cognitive Services



言語



視覚



音声



意思決定

ML プラットフォーム



Azure Machine Learning

機械学習システムは複雑

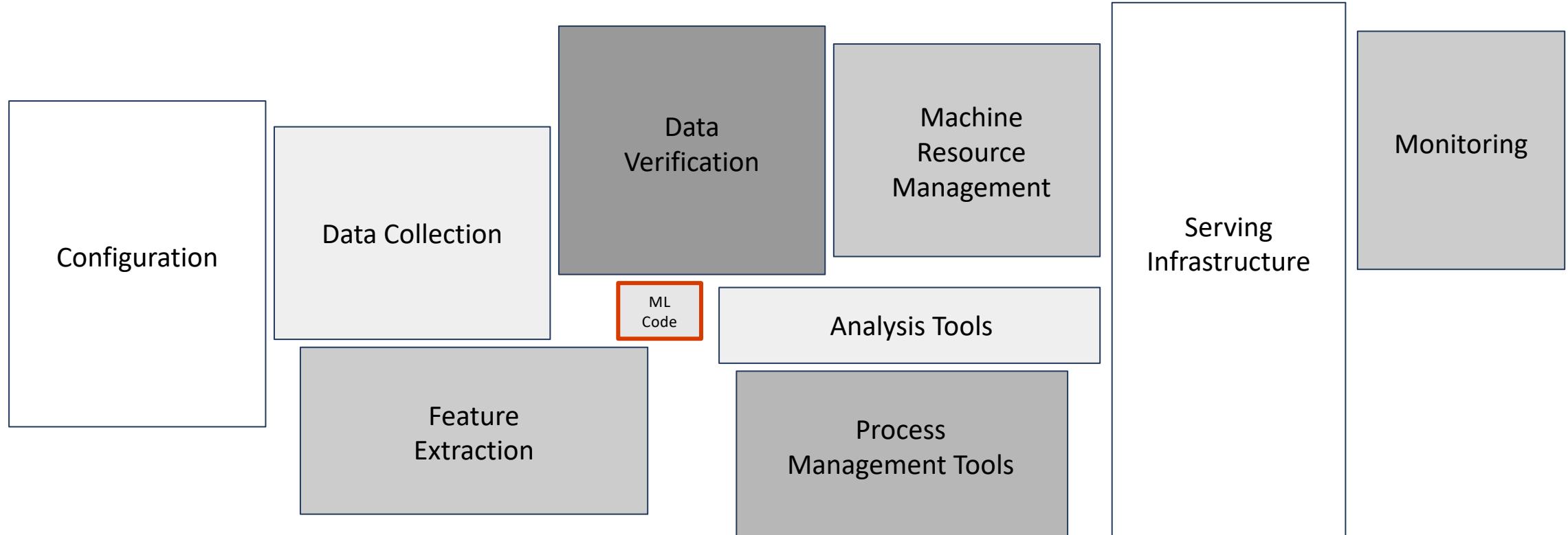


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small green box in the middle. The required surrounding infrastructure is vast and complex.

コンポーネントが多岐に渡るため、構築・運用が難しい

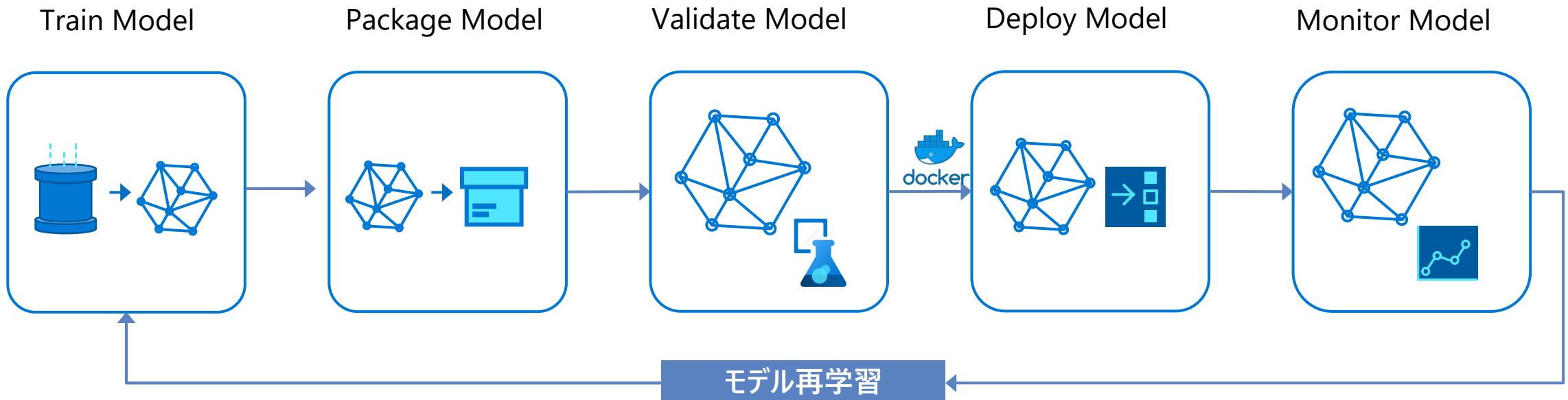
Azure Machine Learning ?

Azure Machine Learning は、Azure が提供するマネージドな機械学習プラットフォームです。モデル学習、デプロイメントから再学習までの機械学習のプロセスを E2E でカバーします。



機械学習のライフサイクル

機械学習のプロセスは非常に複雑です。Azure Machine Learning で一元的に機械学習のプロセスを運用管理していくことができます。



開発と学習
ビジネス課題を解決する

パッケージ化
モデルをどこでも使用可能にする

挙動の検証
応答性の観点と規制遵守の観点から

デプロイ
予測値の生成にモデルを使用する

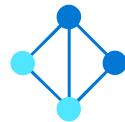
モニタリング
挙動とビジネス価値を監視し、陳腐化したモデルをいつ置換/廃止するか決める

Azure Machine Learning の特徴



迅速なイノベーションと
統合されたツールによる
価値実現までの時間の短縮

- ・ モデルの作成の高速化
- ・ チームの効率化
- ・ 柔軟なインフラストラクチャへの
アクセス



MLOps による
大規模な運用化

- ・ 再現可能なパイプラインの再利用
- ・ 場所を問わないモデル展開
- ・ エンドツーエンド サイクルの監視



安全性が高くハイブリッドで
コンプライアンスを重視した
プラットフォーム上の開発

- ・ 場所を問わないコンピューティング
- ・ 組み込みセキュリティの活用
- ・ コストの抑制

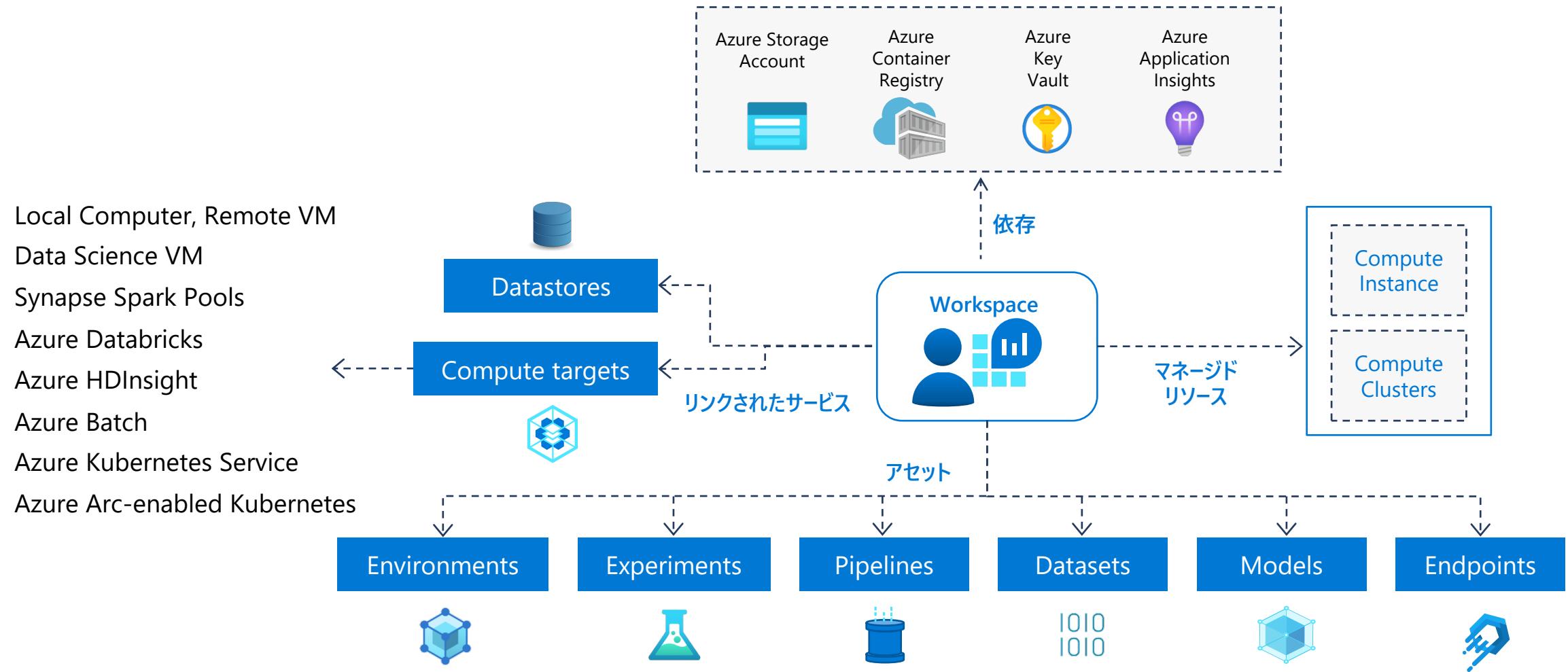


責任ある
ソリューションの提供

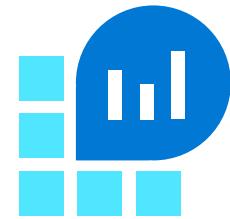
- ・ モデルの可視化の実現
- ・ モデルの不公平性の軽減
- ・ 差分プライバシーの適用

Azure Machine Learning の全体像

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 を構成する主要なリソース



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

Demo : Azure Machine Learning サービス作成

Demo : Azure Machine Learning studio

Compute Instance と Compute Clusters

Azure Machine Learning のマネージドな計算環境である
Compute Instance と Compute Clusters に比較します。

| | Compute Instance  | Compute Clusters  |
|-----------|--|--|
| 用途 | 開発/テスト用途 | 大規模なデータ処理、学習、推論用 |
| 方式 | インタラクティブ | 非インタラクティブ (バッチ) |
| スケーラビリティ | 単一インスタンス | 最大 6500 ノードまで (MPI 非実行) ※詳細はドキュメントを参照のこと |
| 自動シャットダウン | 対応 (プレビュー) | 起動・停止をスケジュール設定可能 |
| 低優先度 | 非対応 | 対応 |

Azure ML – Integrated Notebook

Azure Machine Learning studio に統合されている Jupyter 互換のノートブック環境

Jupyter との互換性

- Jupyter Notebook をインポートしそのまま実行可能

強化された Editor

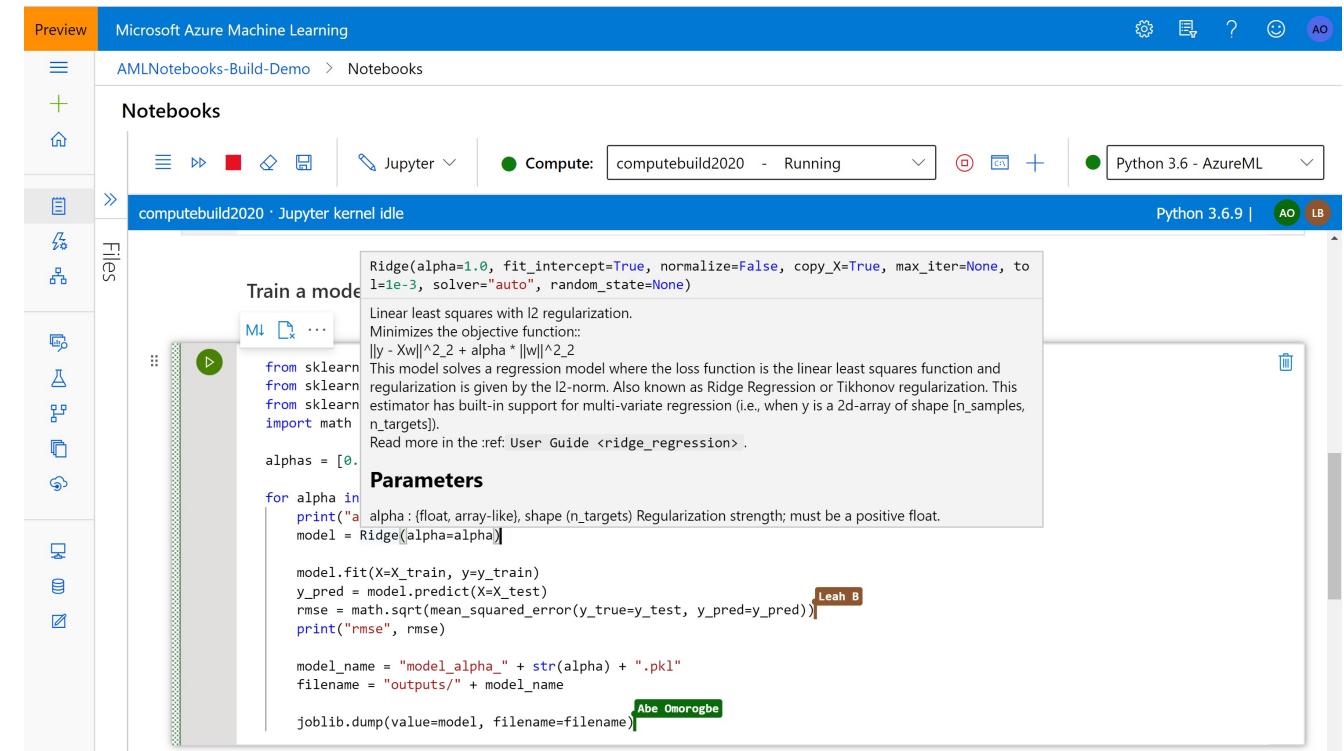
- Monaco editor, IntelliSense & Gather

Git 統合とターミナル

- Git を利用したコード管理や CI/CD の実装
- ターミナルを用いた自由度の高い環境設定

VSCode 統合 (Preview)

- VS Code にそのまま移動可能



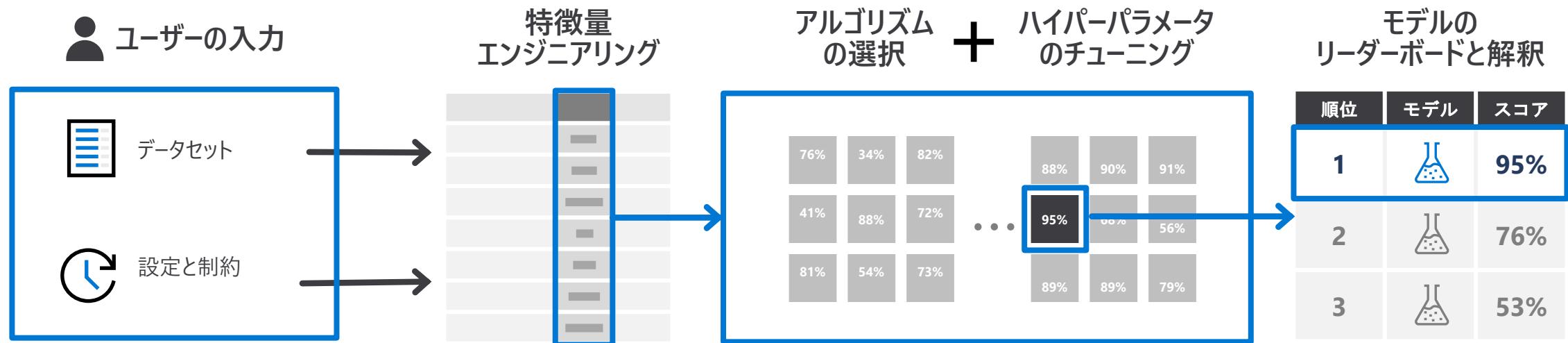
The screenshot shows the Microsoft Azure Machine Learning studio interface with the 'Preview' tab selected. The main area displays a Jupyter notebook titled 'computebuild2020 · Jupyter kernel idle'. The notebook contains Python code for Ridge regression:

```
Ridge(alpha=1.0, fit_intercept=True, normalize=False, copy_X=True, max_iter=None, tol=1e-3, solver='auto', random_state=None)
Linear least squares with l2 regularization.
Minimizes the objective function:
||y - Xw||^2 + alpha * ||w||^2
This model solves a regression model where the loss function is the linear least squares function and regularization is given by the l2-norm. Also known as Ridge Regression or Tikhonov regularization. This estimator has built-in support for multi-variate regression (i.e., when y is a 2d-array of shape [n_samples, n_targets]).
```

The code includes a loop to train a model for different alpha values and calculate RMSE.

Azure ML – AutoML (Automated Machine Learning)

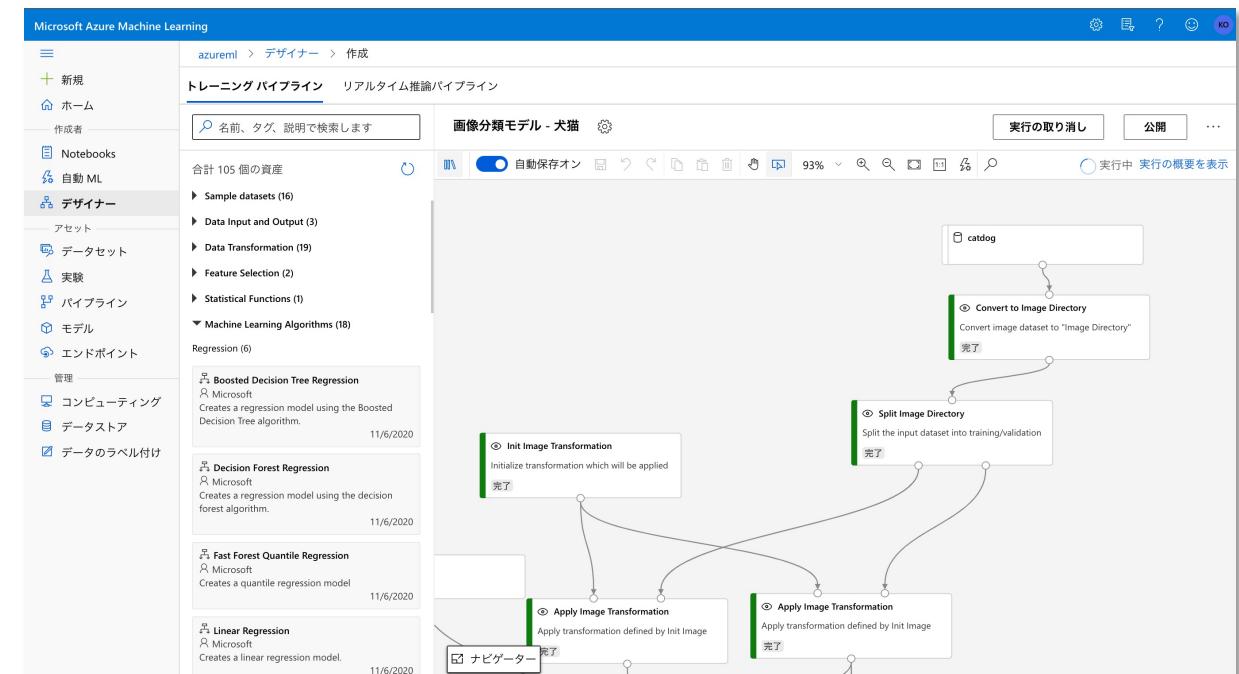
AutoML (自動機械学習) は、与えられたデータに対して「最高のモデル」を探索するために、特徴量エンジニアリング、アルゴリズムとハイパーパラメータの選択を自動実行します。



Azure ML – Designer

機械学習のモデル構築、テスト、デプロイするためのビジュアルパイプライン

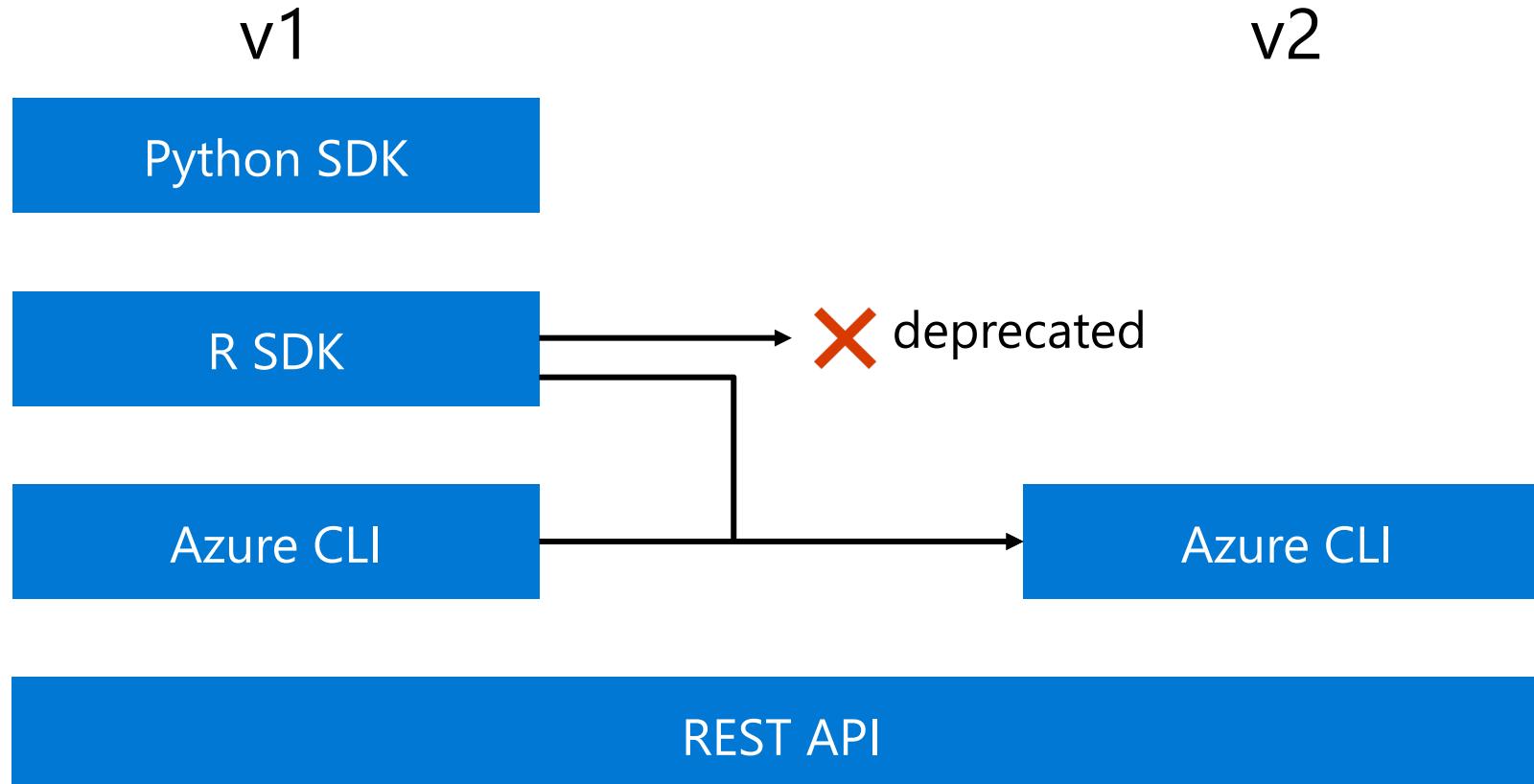
- 直感的なマウス操作でパイプライン構築
- 特徴量エンジニアリング
- モデル学習 (回帰、分類、クラスタリング)
- 推論 (リアルタイム & バッチ)
- カスタムモデル・スクリプト (Python, R)



Demo : Azure Machine Learning Code First

Azure Machine Learning API 提供状況

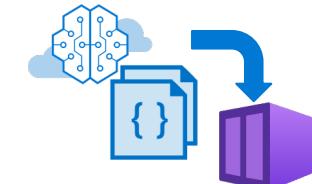
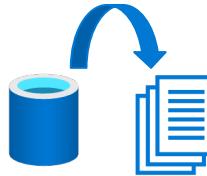
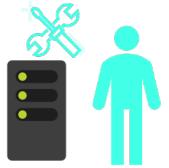
Azure Machine Learning は GUI (Azure Machine Learning studio) から操作することも可能ですが外部環境から API を通じて様々な機能にアクセスできます。



Azure Machine Learning 利用イメージ



従来



- GPU 搭載のサーバーを調達
- 電源確保 & ネットワークに接続
- Python 環境と Jupyter をセットアップ
- GPU のドライバ/ライブラリをインストール
- データを吸出して CSV や Parquet 等のファイル形式で貯う
- データの前処理
- アルゴリズムの試行とパラメーター探索
- Dockerfile を書いてモデルを収めたコンテナを作成
- Flask / FastAPI などで HTTP エンドポイントをセット
- Kubernetes 等にコンテナをデプロイ



Azure ML

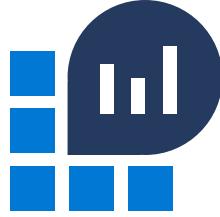


- 計算リソースをアタッチ
- Azure ML Compute Instance/Cluster を展開
- Environment を読み込む
- Compute Instance をそのまま使う
- Datastores としてデータソースをアタッチして、Dataset を作成
- Auto ML でモデルを自動作成
- 大規模計算リソースで効率的に探索
- 実験/モデルを記録
- 保存したモデルにメタデータを付与する
- エントリスクリプトの作成
- アタッチした推論用のリソースにデプロイ

参考資料

| Category | Link |
|---|---|
| Azure Machine Learning ドキュメント | Azure Machine Learning のドキュメント Microsoft Docs |
| Azure Machine Learning Web ページ | Azure Machine Learning - サービスとしての ML Microsoft Azure |
| サンプルコード集 (コミュニティ) | Azure/azureml-examples: Official community-driven Azure Machine Learning examples, tested with GitHub Actions. |
| サンプルコード集 (公式) | Azure/MachineLearningNotebooks: Python notebooks with ML and deep learning examples with Azure Machine Learning Python SDK Microsoft (github.com) |
| AI Show | AI Show Microsoft Docs |
| Microsoft AI/ML tech collection | microsoft/machine-learning-collection: machine learning tech collections at Microsoft and subsidiaries. (github.com) |
| AI – Machine Learning Blog (tech community) | AI - Machine Learning Blog - Microsoft Tech Community |
| Data Scientist 向けポータルサイト | データ サイエンティストのための機械学習ソリューション Microsoft Azure |
| Azure Machine Learning – Microsoft Learn | Microsoft Learn - Azure Machine Learning |
| | |

参考情報



Workspace

- Workspace は Azure Machine Learning の最上位リソースである。Azure Machine Learning 使用時に作成したあらゆる成果物を集約した一元的な作業スペースを提供する。
- Workspace はモデル学習で利用した Compute Targets の情報やログ、メトリック、出力、スクリプトのスナップショットを含む学習の実行履歴を保持する。

- モデルは Workspace に登録される。
 - 複数の Workspace を作成することができ、それぞれの Workspace は複数人で共有することができる。
 - 新しく Workspace を作成したとき、自動的に以下 Azure リソースが作成される:
 - [Azure Container Registry](#)
 - 学習とデプロイに使用する Docker コンテナの登録に利用される
 - [Azure Storage](#)
 - Workspace のデフォルトの Datastore やログ管理などに使用される
 - [Azure Application Insights](#)
 - 推論エンドポイントをモニタリングした情報が保存される
 - [Azure Key Vault](#)
 - Compute Targets によって使用されるシークレットとその他 Workspace が使用するセンシティブな情報が保存される



Datasets & Datastores

- Datastores は Azure のストレージサービスに対する接続情報を保持するために使用されます。
- Datasets は Datastores を使用して接続された各種データソース内に保存されたデータを参照する。

- Azure Machine Learning Datasets
 - Azure Machine Learning で利用するデータへのアクセスが容易になる。Dataset を作成するとデータへの参照とメタデータのコピーが作成される。データは元々データがあった位置に維持されるため、余計なストレージコストをかけず、データソースの一貫性を維持することができる。
- Datastores を利用する Datasets
 - Datastores は認証情報や元のデータソースの完全性を危険にさらすことなく接続に必要な情報を保持する。サブスクリプション ID や認証トークンといった接続に必要な情報は Workspace に紐づけられた Key Vault に保存しているため、スクリプトに機密情報を直接実装することなくストレージに対してセキュアにアクセスできる。



Environments

Environments は機械学習を実行するソフトウェア環境の構築や運用管理の機能を提供する。

学習や推論スクリプトで利用する Python パッケージ、環境変数、ソフトウェアの設定、ランタイム（Python、Spark や Docker 等）で定義され、最終的には Docker コンテナとして実行される。

Workspace 上で管理・バージョニングされ、再現性、監査可能性、機械学習ワークフローの異なる計算リソース間での相互利用性を確保する

Environment はローカルコンピューター上で以下のように使用することが可能

- 学習スクリプトの開発
- モデルの学習をスケーリングする際、Azure Machine Learning の計算リソース上で同一環境の再利用
- 同一環境を使用したモデルのデプロイ
- 既存モデルの学習時に使用された環境の再現
- Workspace のデフォルト環境として利用可能な curated environments （代表的なシナリオに沿ってプリセットされた environment）の提供



Experiments and Runs

Experiment は1つのスクリプトに由来する複数の Run をグループ化したもので、Azure ML Workspace に所属する。各 Run の情報は関連付けられた Experiment の配下に保存される。

Experiment

スクリプトの実行によって生成された複数の Run をグルーピング

- Azure ML Workspace 直下に所属
- 各 Run の情報を保存し、横断的に管理

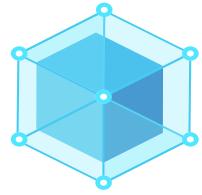
Run

Job を submit することで生成され、以下の情報を保持

- 実行のメタデータ (実行した日時、実行に要した時間等)
- スクリプトから出力したメトリック
- 実験に関する自動収集された、もしくは明示的にアップロードされたファイル
- スクリプトを含むディレクトリの実行直前のスナップショット

Run configuration

計算リソース上でスクリプトがどのように実行されるべきかの定義



Models

Model は学習を経て生成される artifact* である。Model は新しいデータに対する予測結果を得るために使用する。

Model Registry は Azure Machine Learning Workspace 上の全ての Model 追跡する。

* 機械学習の過程で生じる中間生成物、画像、モデル等のあらゆるファイルを artifact と呼称する

Model

Model は Azure Machine Learning 上の Run から生成される

- Note: Azure Machine Learning の外部で学習したモデルを使用することも可能

Azure Machine Learning はフレームワーク非依存であるため Model を作成する際にあらゆる機械学習フレームワークを使用可能

Model は Azure Machine Learning の Workspace 配下に登録可能

Azure Machine Learning に作成された Model はバージョニングされ、モデルのバージョンを比較・選択するためのツールが組み込みで付属

一度 Model が作成されれば、Dockerfile や Docker イメージの生成とあらゆる場所へのモデルのデプロイが可能

Model Registry

Azure Machine Learning Workspace 上の全モデルを追跡

Model は名前とバージョンによって一意に特定される

Model を登録するときに追加でメタデータタグを付与することができ、タグは Model の検索時に利用可能

Docker イメージに使用されている Model は削除できない



Pipelines

1つの Azure Machine Learning Pipeline は 完成した機械学習タスクのワークフローを表し、 機械学習の各サブタスクはパイプラインを構成 数する一連の手順として内包される

Azure Machine Learning Pipeline には、Python スクリプトを呼び出すだけのシンプルなものから、あらゆることを実行するものまで存在する。

タスク

Pipelines は以下のようないくつかの機械学習タスクにフォーカス:

- インポート、検証とクリーニング、変換、正規化、ステージングを含む「データの準備」
- パラメーター化された引数、ファイルのパス、ログとレポート出力の設定を含む「学習の設定」
- 効率的かつ繰り返し実行できる「学習と検証」。特定のデータサブセット、異なるハードウェアからなる計算リソース、分散処理、進捗状況の監視を指定することで、効率性を確保することが可能
- バージョン管理、スケーリング、プロビジョニング、アクセス制御等の要素を含む「デプロイ」



Computes

Computes は学習スクリプトを実行したりモデルを推論用途でホストするための計算リソースです。

Azure Machine Learning Python SDK, CLI, Studio から作成し運用管理することができます。

既存のリソースをアタッチすることもできます。

ローカル環境で実行したのちに、Compute Targets 上でスケールアップ・スケールアウトすることができます。

現在サポートされている Compute Target

| Compute Targets | 学習 | デプロイ |
|--|----|------|
| Local Computer | ✓ | |
| A Linux VM in Azure (such as the Data Science Virtual Machine) | ✓ | |
| Azure ML Compute Clusters | ✓ | ✓ |
| Azure ML Compute Instance | | |
| Azure Databricks | ✓ | |
| Azure Data Lake Analytics | ✓ | |
| Azure HDInsight | ✓ | |
| Azure Container Instance | | ✓ |
| Azure Kubernetes Service | | ✓ |
| Azure IoT Edge | ✓ | |
| Field-programmable gate array (FPGA) | | ✓ |
| Azure Functions (preview) | | ✓ |
| Azure App Service (preview) | | ✓ |
| Azure Synapse Spark Pool (preview) | ✓ | |
| Azure Arc enable Kubernetes (preview) | ✓ | ✓ |



Microsoft AI

- 本書に記載した情報は、本書各項目に関する発行日現在の Microsoft の見解を表明するものです。Microsoftは絶えず変化する市場に対応しなければならないため、ここに記載した情報に対していかなる責務を負うものではなく、提示された情報の信憑性については保証できません。
- 本書は情報提供のみを目的としています。Microsoft は、明示的または暗示的を問わず、本書にいかなる保証も与えるものではありません。
- すべての当該著作権法を遵守することはお客様の責務です。Microsoftの書面による明確な許可なく、本書の如何なる部分についても、転載や検索システムへの格納または挿入を行うことは、どのような形式または手段（電子的、機械的、複写、レコーディング、その他）、および目的であっても禁じられています。これらは著作権保護された権利を制限するものではありません。
- Microsoftは、本書の内容を保護する特許、特許出願書、商標、著作権、またはその他の知的財産権を保有する場合があります。Microsoftから書面によるライセンス契約が明確に供給される場合を除いて、本書の提供はこれらの特許、商標、著作権、またはその他の知的財産へのライセンスを与えるものではありません。

© 2021 Microsoft Corporation. All rights reserved.

Microsoft, Windows, その他本文中に登場した各製品名は、Microsoft Corporation の米国およびその他の国における登録商標または商標です。

その他、記載されている会社名および製品名は、一般に各社の商標です。