



2. システム設計 & 構築

Azure Machine Learning – 構築・運用編

Keita Onabuta

FastTrack for Azure
Senior Customer Engineer for AI/ML



Agenda

- 計算リソース
- データソース
- ネットワーク構成
- 認証認可
- データ保護
- システム監視
- 監査ポリシー
- コスト管理

計算リソース

基礎知識

- Azure ML - Computes

検討事項

- マネージド計算環境の比較ポイント
- 推論環境の比較ポイント
- Public IP or Private IP の選択

ガイドライン・実装手順

- 推論環境の選択ガイドライン



Computes

Computes は学習スクリプトを実行したり、学習済み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 など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)	✓	✓

マネージド計算環境の比較ポイント

Azure Machine Learning のマネージドな計算環境であるコンピューティングインスタンス (Compute Instance) と コンピューティングクラスター (Compute Clusters) の機能比較をします。

	Compute Instance 	Compute Clusters 
用途	開発/テスト環境	本番環境、大規模なデータ処理、学習、バッチ推論用
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
起動停止	起動・停止をスケジュール設定可能 (Preview)	ワークロードに応じて計算ノードが自動で起動・停止
低優先度	非対応	対応

推論環境の比較ポイント

Azure Machine Learning が推論環境としてサポートしている Azure Container Instance と Azure Kubernetes Service の比較をします。

	Azure Container Instance 	Azure Kubernetes Service 	Azure ML Pipeline 
用途	開発/テスト環境	本番環境	開発/テスト/本番環境
方式	リアルタイム推論	リアルタイム推論	バッチ推論
スケーラビリティ	単一インスタンス	スケールアウト可能	スケールアウト、スケールアップ可能
自動シャットダウン	なし	スケールイン可能	スケールイン可能
GPU 対応	なし	あり	あり

※ 現在プレビュー中の Managed Inference について記載していません。

Public IP or Private IP の選択

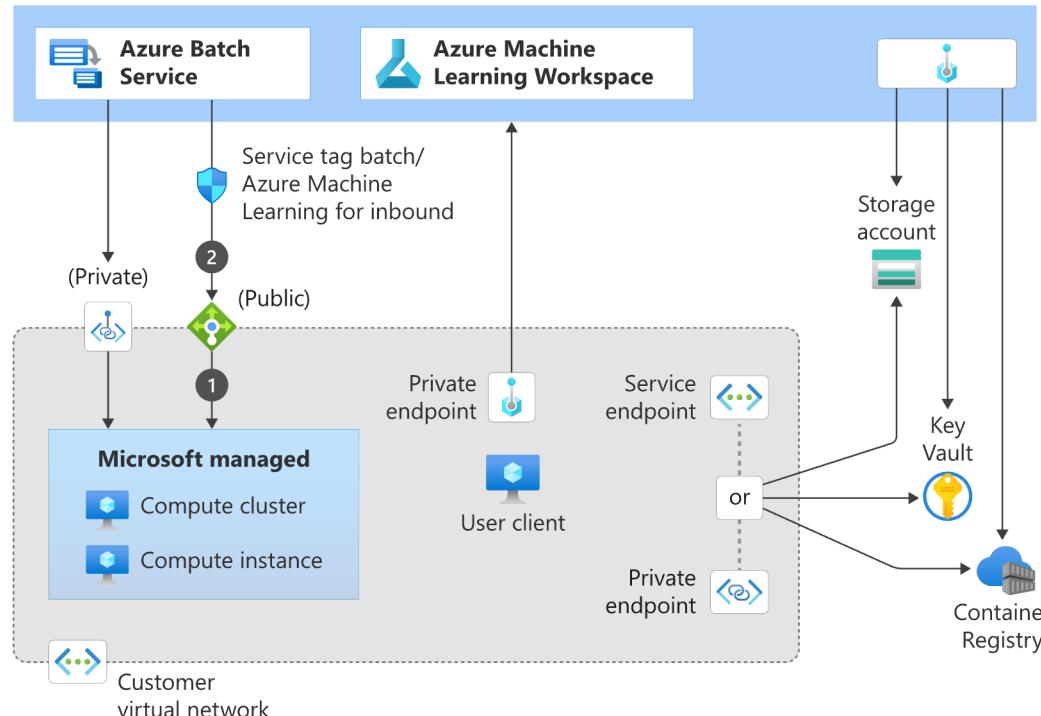
Compute (Compute Instance, Compute Clusters) の IP アドレスは、デフォルトでは Public IP になります。Private IP の設定は Public Preview で提供されています。(2022年4月現在)

- **Public IP の Compute**

- Compute Cluster は Azure Batch Service をベースに構築されています。
受信アクセス (Inbound access) としてサービスタグを利用して
「AzureMachineLearning」と「Batch NodeManagement」を許可する
必要があります。

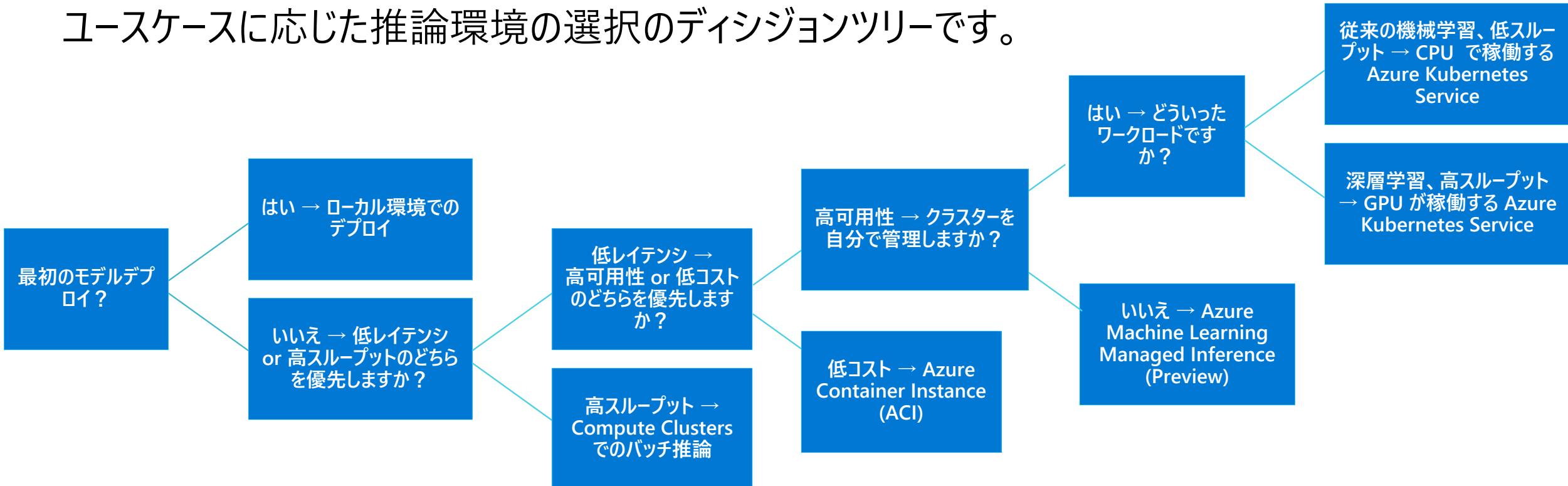
- **Private IP の Compute**

- インターネットからの受信アクセス (Inbound access) の設定は必要がありません。
- 送信アクセス (Outbound access) のために Azure Firewall の Egress Firewall などを設定する必要があります。
- Preview 機能のため正式なサポートは受けられることは予め承知ください。



推論環境の選択ガイドライン

ユースケースに応じた推論環境の選択のディシジョンツリーです。



データソース

基礎知識

- Azure ML – Datastore & Datasets

検討事項

- データソースの選択

ガイドライン・実装手順

- データセット形式の選択ガイドライン
- パフォーマンス最適化のポイント



Datastores & Datasets

- 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。

- Datastores

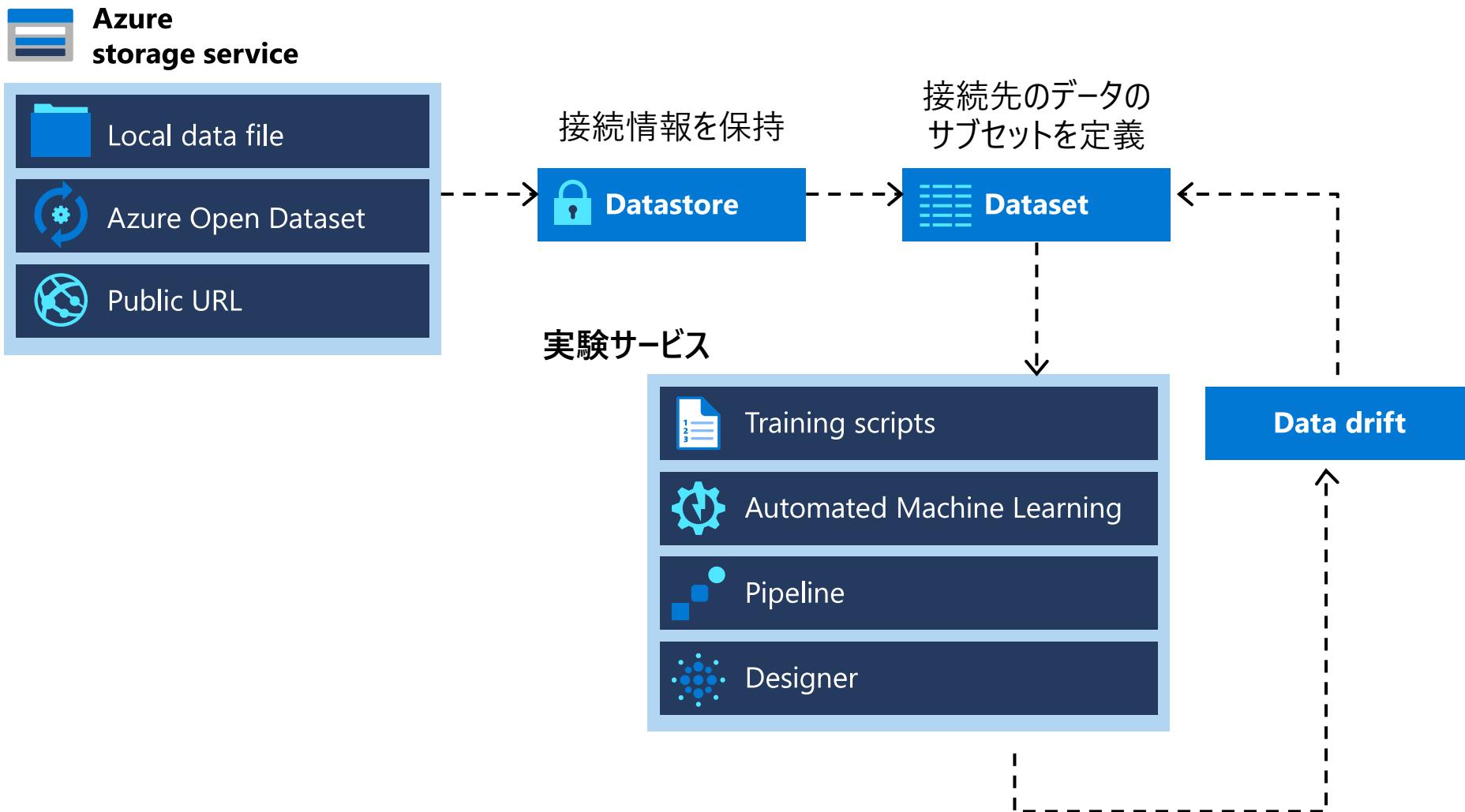
- Datastores は認証情報や元のデータソースの完全性を危険にさらすことなく接続に必要な情報を保持する。サブスクリプション ID や認証トークンといった接続に必要な情報は Workspace に紐づけられた Key Vault に保存しているため、スクリプトに機密情報を直接実装することなくストレージに対してセキュアにアクセスできる。

- 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。



Azure Machine Learning



データソースの選択

Azure ML Datastores がサポートするデータソースから利用するものを選択します。

データソースの種類	資格情報ベースの認証	ID ベースの認証
Azure Blob Storage	Account key SAS token	対応
Azure File Share	Account key SAS token	
Azure Data Lake Storage Gen 1	Service principal	対応
Azure Data Lake Storage Gen 2	Service principal	対応
Azure SQL Database	SQL authentication Service principal	対応
Azure PostgreSQL	SQL authentication	
Azure Database for MySQL	SQL authentication	

データセットの形式の選択ガイドライン

データセット (Datasets) は “表形式” と “ファイル形式” の 2 種類をサポートしています。それぞれの使い分け方について説明します。

ファイル形式データセット

- ・ 画像、音声、動画ファイルなどの非構造化データを扱う場合には必須です。
- ・ 数値系のデータを扱う場合でも、まずはファイル形式の利用を検討します。

表形式データセット

- ・ 実験サービスの AutoML や Designer、データセットのプロファイルやモニタリングを利用する際には必須です。
- ・ Pandas に変換できる機能などがあるが、データ型の指定ができず、変換も難しいため、厳密にデータ型を定義したい場合はファイル形式で利用します。

パフォーマンス最適化のポイント

大規模データを扱う場合やデータの出し入れが頻繁の場合はパフォーマンスが問題になりやすいです。ここではパフォーマンスを向上させるポイントを挙げます。

□ Parquet の利用

- ・ 列指向でデータを保持するため高効率なデータ圧縮を実現。機械学習で必要な列が一部の場合でも高速。

□ リージョン選択

- ・ 計算リソースとデータソースはなるべく同じ・近いリージョンにデプロイする。

□ Mount or Download

- ・ ファイル形式のデータセットは計算リソースにデータをマウントもしくはダウンロードすることができる。

□ Spark の活用

- ・ 大規模データの前処理が必要な場合は、Synapse Analytics Spark Pool との連携機能を用いて、Synapse 側でデータの前処理を実行する。

□ Azure Data Lake Storage Gen2 の利用

- ・ 階層型ストレージの Azure Data Lake Storage Gen2 を利用することでファイル編集速度が向上するケースが多いです。

ネットワーク構成

基礎知識

- Azure Private Link 概要
- Azure Bastion 概要

検討事項

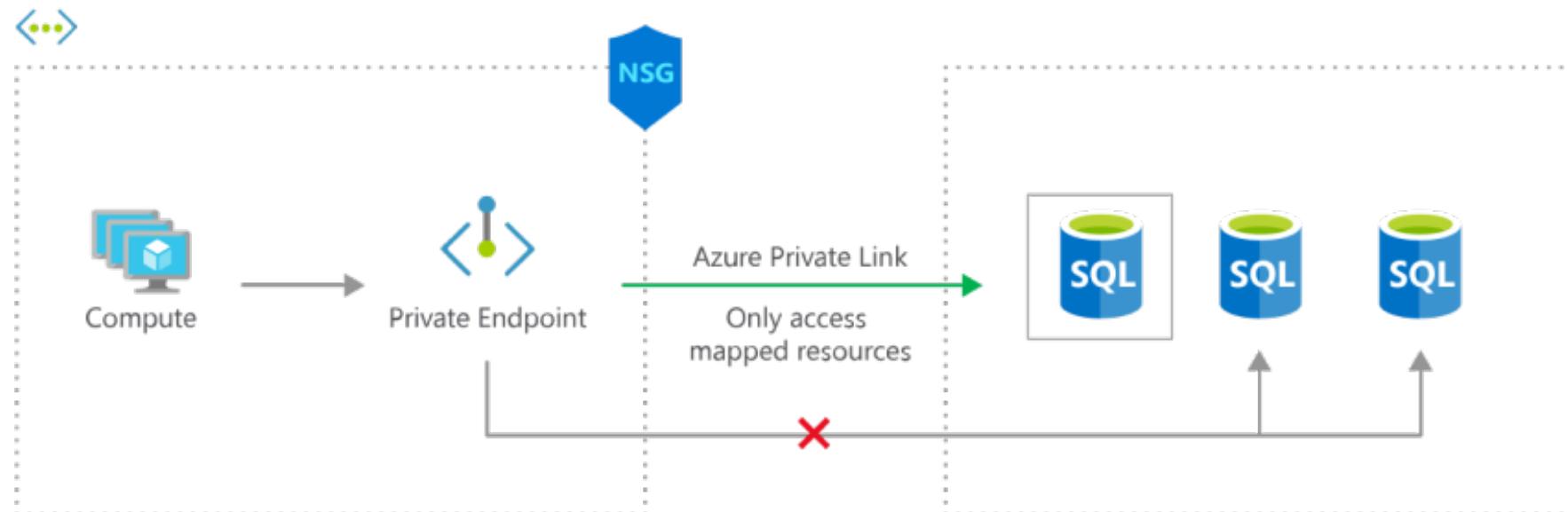
- Private Link 利用の検討
- ネットワークアーキテクチャの設計

ガイドライン・実装手順

- Private Link 利用時の構成ガイドライン

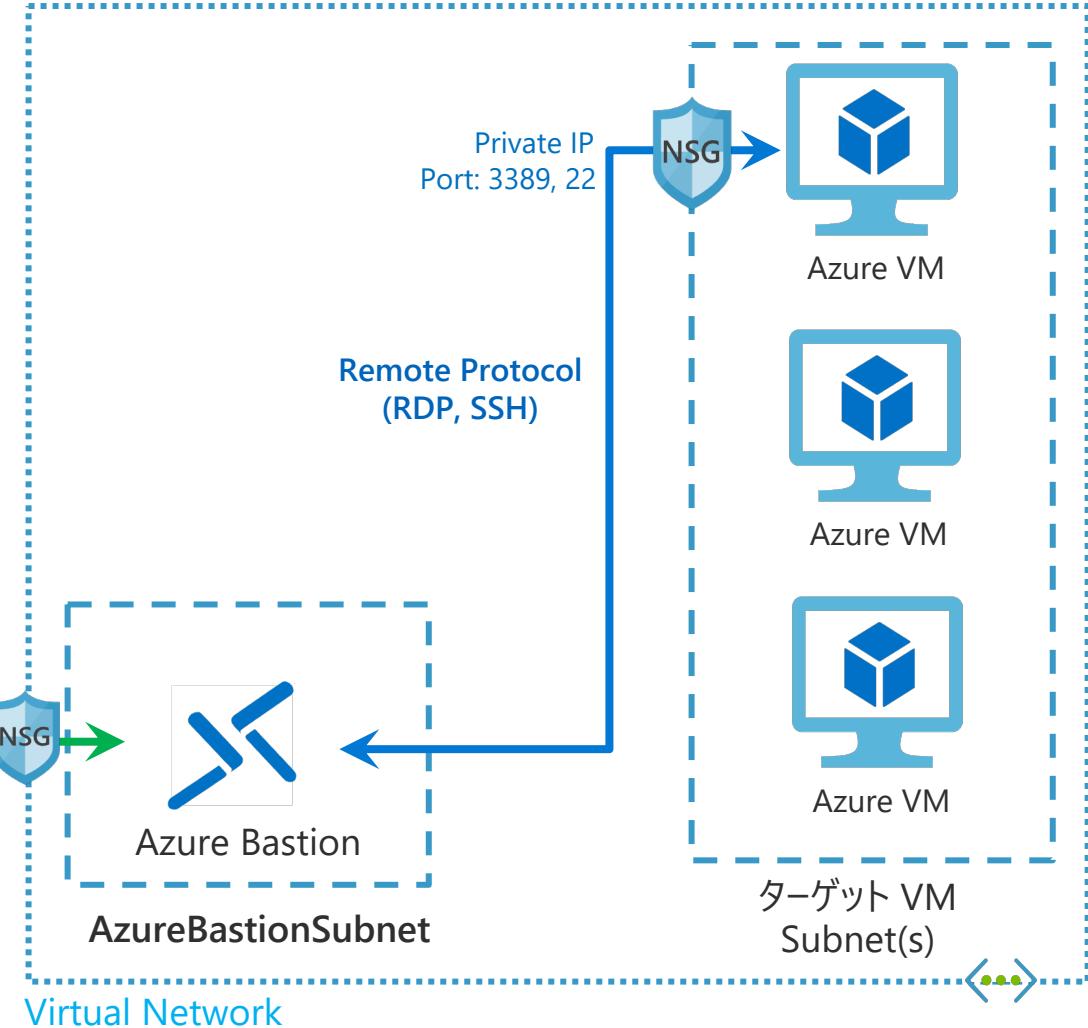
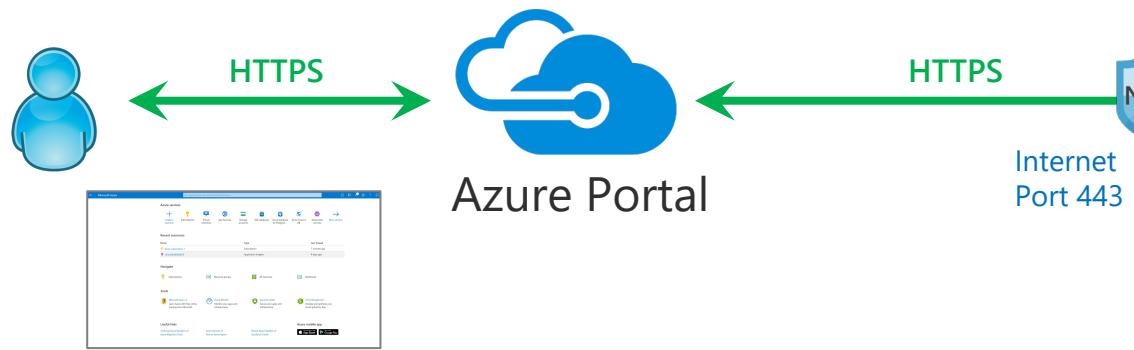
Azure Private Link 概要

Private Link を利用して、Private Endpoint と Azure PaaS サービスをマッピングします。Private Endpoint 経由での通信のみアクセスが許可されるため、インターネット環境などの外部からのアクセスを遮断できます。



Azure Bastion 概要

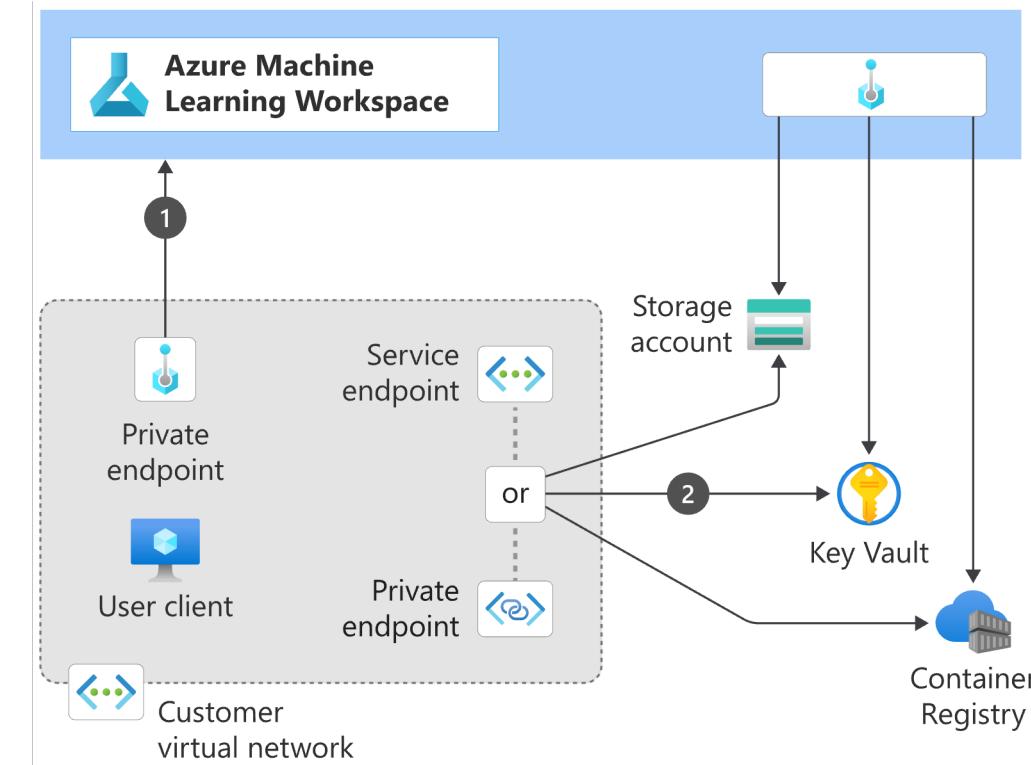
- Azure Bastion は、より安全でシームレスな Remote Desktop Protocol (RDP) および Secure Shell Protocol (SSH) による仮想マシン (VM) へのアクセスを提供するフル マネージド サービス
- パブリック IP アドレスを用いることなく、仮想ネットワーク内に接続し、仮想マシンに対してリモートアクセスを提供



Private Link 利用の検討

Azure ML の閉域化は Private Link を利用する必要があります。

Azure ML Workspace への Private Link の設定 (①) に加えて、Storage Account、Key Vault、Container Registry への Private Link (もしくは Service Endpoint) の設定 (②) も必要です。



基本アーキテクチャ

ネットワークアーキテクチャの設計

代表的なネットワーク構成とそれぞれの特徴を示します。

	構成	ワークスペース	計算環境	ストレージ	特徴
1	最小限度	パブリック	パブリック	パブリック	Azure Portal からデプロイした場合のデフォルト構成。仮想ネットワークを全く利用していない。
2	Azure AD と仮想ネットワークによるセキュアな環境	パブリック	仮想ネットワークへの配置	Firewall 設定 (特定の仮想ネットワーク & クライアント PC の IP アドレスからのアクセスのみ許可)	Azure ML ワークスペースはインターネットからアクセス可能だが Azure AD による認証が必要であるが。ユーザがクライアントからワークスペース経由でストレージにアクセスする際はストレージの Firewall が有効になっている。
3	閉域環境	Private Link	仮想ネットワークへの配置	Private Link	Azure ML workspace は Azure AD による認証 + 特定の仮想ネットワークからのアクセスのみ可能になる。その他ストレージなどの関連サービスも同様に特定の仮想ネットワークに対応させる。

Private Link 利用時の構成ガイドライン

Private Link 利用時に特に注意すべきポイントです。詳細はドキュメントや参考資料を参照ください。

□ 関連 Azure サービスの閉域化

- 通常 Private Link の利用を推奨します。
 - 通信コストは発生するがサービスエンドポイントより制約が少ない。
 - Private Link はやや通信コストが発生する。
 - オンプレミスからの接続に対応しやすい。

□ DNS による名前解決

- 特にオンプレミスから直接 Azure に接続する場合、Azure DNS へ直接アクセスできないため、Private DNS Zone による名前解決ができない。そのため DNS Proxy を配置する必要がある。
 - PaaS としては Azure Firewall で DNS Proxy が利用可能
 - Azure Bastion でホストされる VM など Azure 内部からのアクセスであれば自動で名前解決される。

	Private Endpoint	VNet Service Endpoint
Machine Learning Workspace	○	×
Storage	○	○
Key Vault	○	○
Container Registry	○ ※1	△ ※2
Application Insights	×	×

※1 ACR 側設定により処理が分岐

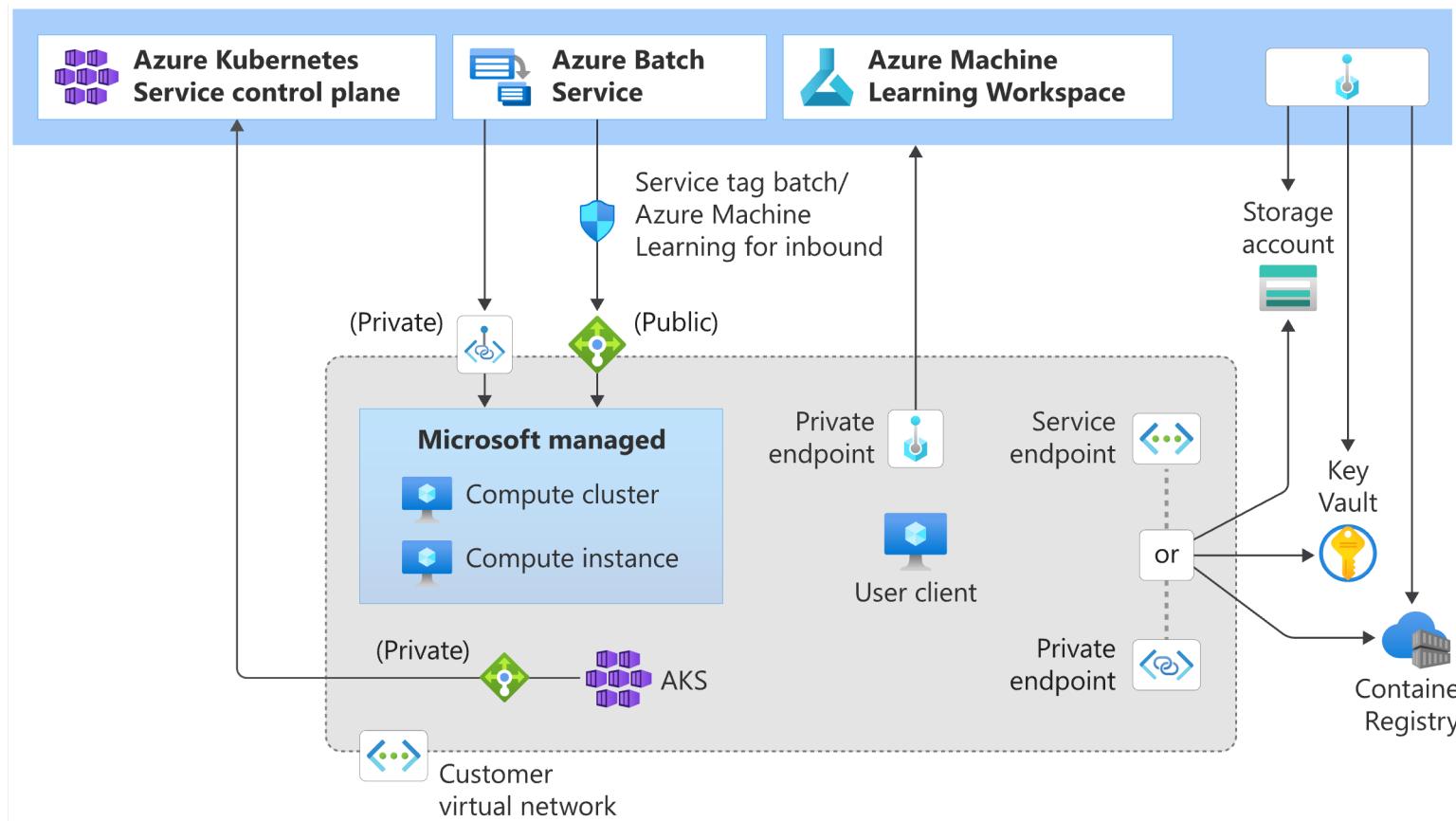
※2 ノーサポート

Private Link 利用時の構成ガイドライン

- マネージドな計算環境はデフォルトは Public IP を利用
 - Compute リソース (Compute Clusters, Compute Instance) にはロードバランサー LB が付属しており Public IP が割り当てられ、Azure Batch (Batch Nodemanagement) と Azure ML Workspace からのインバウンド通信の許可が必要である。
- 仮想ネットワークの着信/送信トラフィックの構成
 - 多くのユースケースで、機械学習で必要な Python パッケージなどのインストールで Microsoft 以外のサイトへのインターネットへの送信が必要である。
 - 直接インターネットへ通信させず、Azure Firewall や自社オンプレミス環境のファイアウォールを用いたfiltratingや Private Package を用いたパッケージ管理を行う場合もある。

Private Link 利用時の構成ガイドライン

Private Link 構成時のアーキテクチャのイメージです。



認証認可

基礎知識

- Azure Active Directory 概要
- Azure RBAC 概要

検討事項

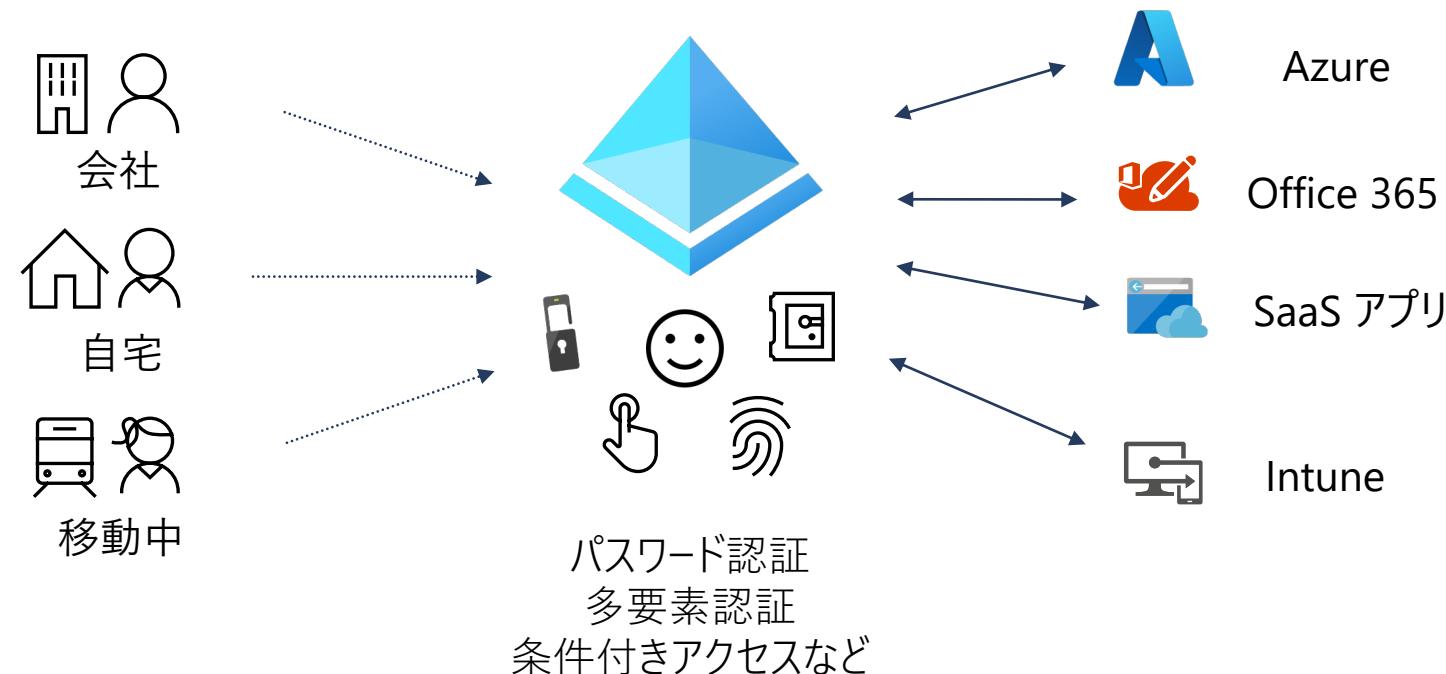
- 認証方法の検討
- ロールの設計

ガイドライン・実装手順

- ワークロードに応じたロールの設定

Azure Active Directory 概要

Azure Active Directory (aka Azure AD) は、マイクロソフトが提供する、マルチテナント対応のクラウドベースの ID およびアクセス管理サービスです。Microsoft 365、Azure portal やその他のさまざまな SaaS アプリケーションなどの外部リソースにアクセスするのに役立ちます。パスワード認証、多要素認証などさまざまな機能が提供されます。



Azure RBAC (ロールベースのアクセス制御) 概要

クラウド リソースに対するアクセスの管理は、クラウドが使用している組織にとって重要な機能です。ロールベースのアクセス制御 (RBAC) は、Azure のリソースにアクセスできるユーザー、そのユーザーがそれらのリソースに対して実行できること、そのユーザーがアクセスできる領域を管理するのに役立ちます。

Azure RBAC は Azure Resource Manager 上に構築された承認システムであり、Azure 内のリソースに対するアクセスをきめ細かく管理できます。

認証方法の検討

Azure ML のコンポーネントや関連する Azure サービスへの認証方法を検討します。

Azure ML Workspace

種類	認証形式	ユースケース
対話型	Azure AD のユーザーアカウント	実験・反復開発中
サービスプリンシパル	Azure AD のサービスプリンシパル	CI/CD パイプライン
Azure CLI セッション	Azure CLI 認証	実験・反復開発中、CI/CD パイプライン
マネージド ID	仮想マシンのマネージド ID	実験・反復開発中、CI/CD パイプライン

※ Azure CLI 認証では Azure AD、サービスプリンシパル、マネージド ID それぞれの認証形式をサポートしている。

認証方法の検討 (cont'd)

計算リソースと関連 Azure サービス

Azure Machine Learning のマネージドな計算リソースや関連する Azure サービスへの接続で用いる認証方法について検討します。



Compute
Instance



Compute
Clusters



Storage



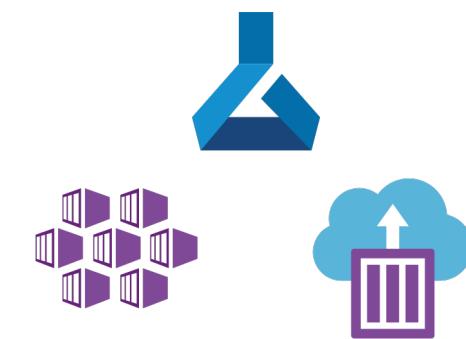
Storage



Data
Lake



SQL



計算リソース

データソース

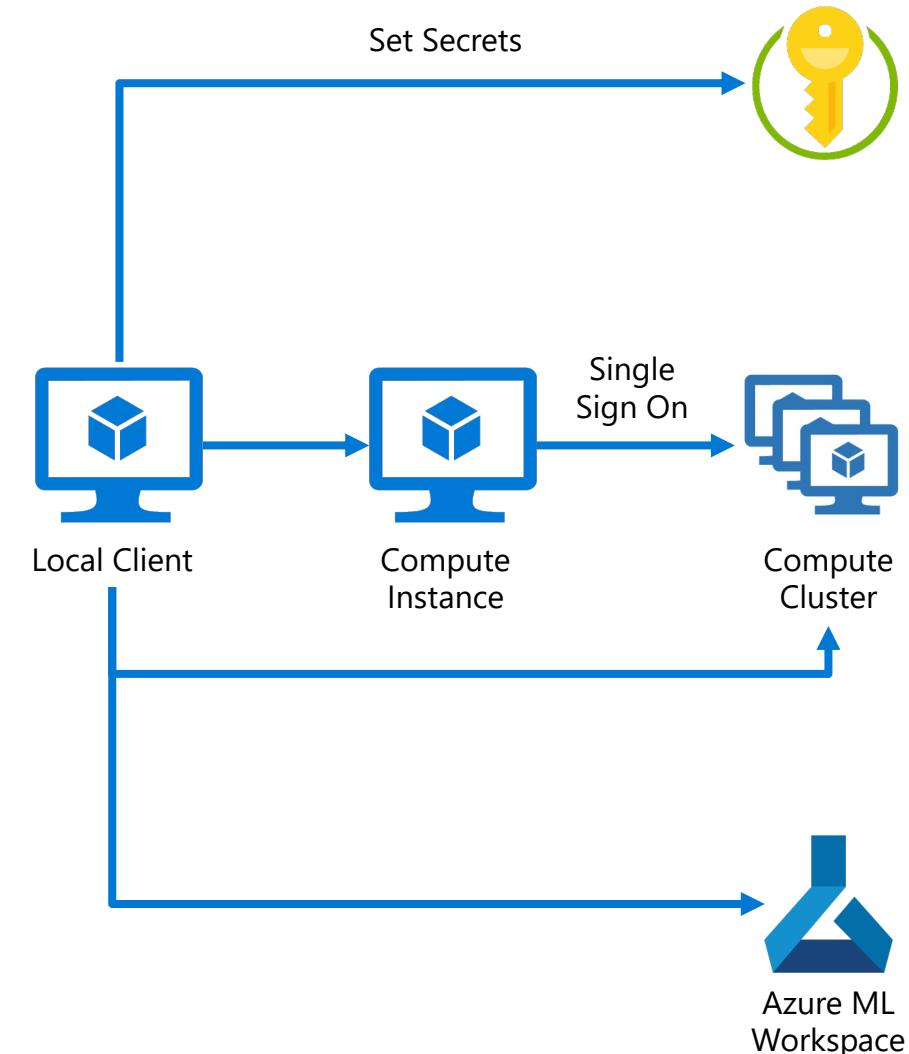
コンテナレジストリー

推論環境

認証方法の検討 (cont'd)

計算リソース

- Compute Instance、Compute Clusters は Azure AD 認証
 - Compute Instance から Compute Cluster は シングルサインオン
- 下記のサービスも Azure AD 認証
 - Azure ML ワークスペース
 - Azure Key Vault
 - Azure ML Studio



認証方法の検討 (cont'd)

データソース

Compute Cluster

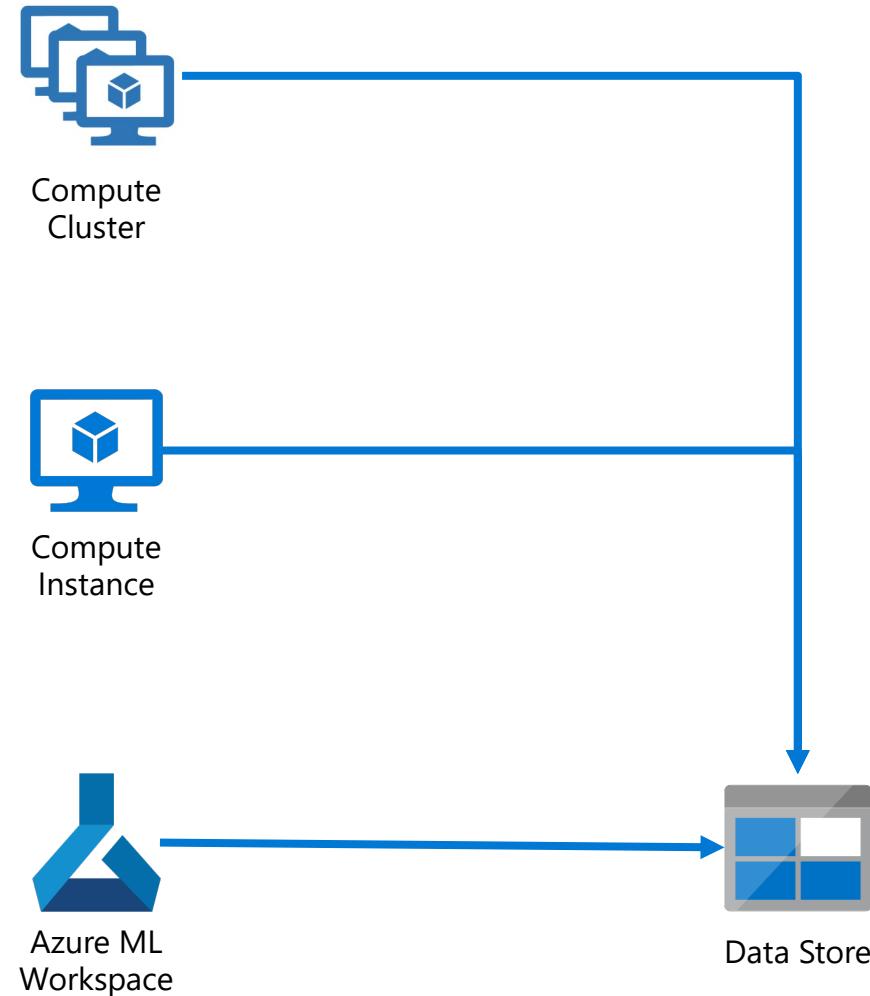
- 資格情報ベースの認証
- ID ベースの認証
 - マネージド ID を用いたデータソースへの認証
 - Azure AD パススルー認証 (Preview, Azure CLI 2.0 のみ)

Compute Instance

- 資格情報ベースの認証
- ID ベースの認証
 - Azure AD 認証
 - マネージド ID を用いたデータソースへの認証 (Private Preview,)

Azure ML ワークスペース

- 資格情報ベースの認証
- ID ベースの認証
 - ワークスペースのマネージド ID による認証
 - Azure AD パススルー認証



認証方法の検討 (cont'd)

コンテナレジストリー

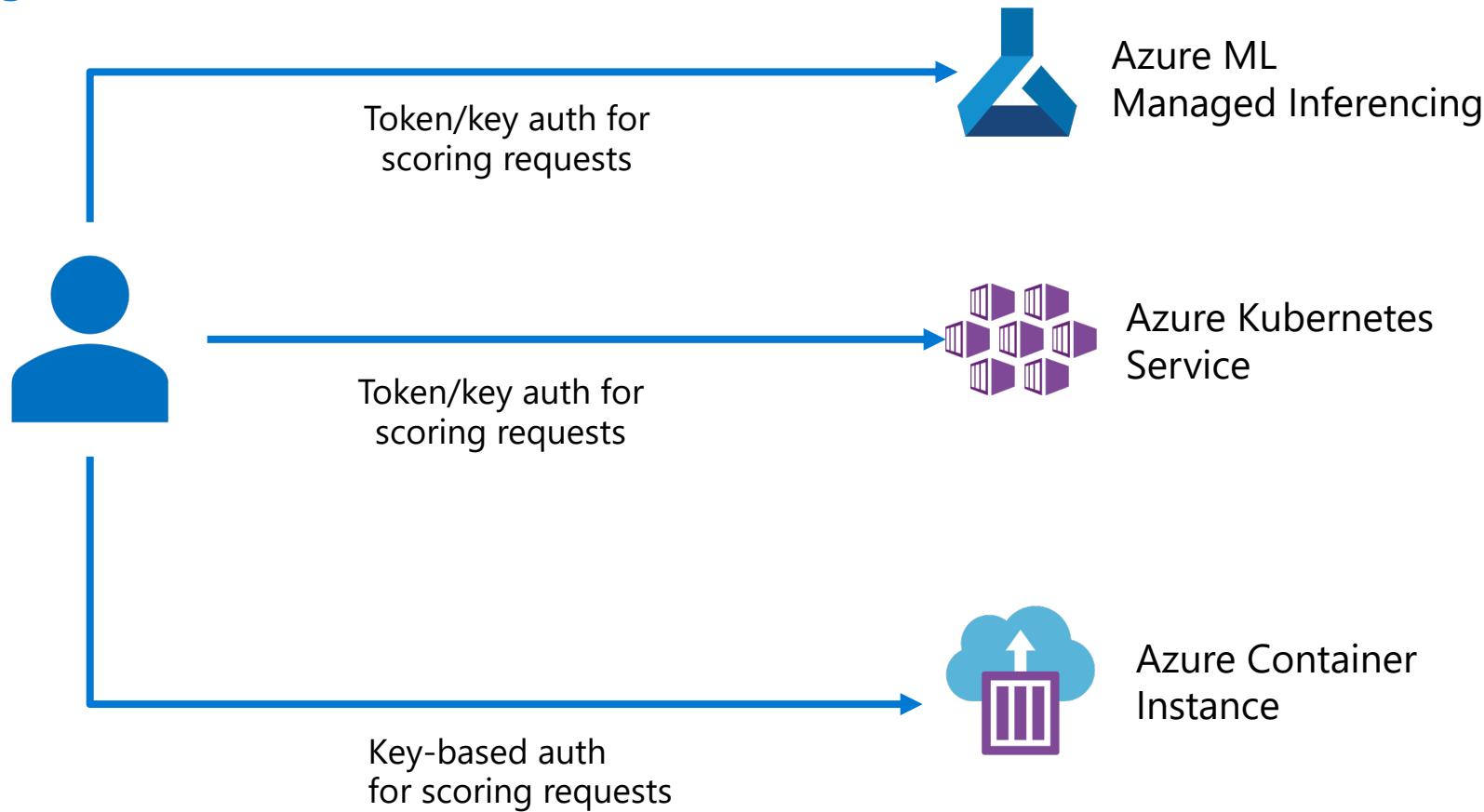
Compute Clusters

- ・ ワークスペースのマネージド ID
- ・ Compute Clusters のマネージド ID
- ・ Azure Container Registry の管理キー



認証方法の検討 (cont'd)

推論環境



ロールの設計

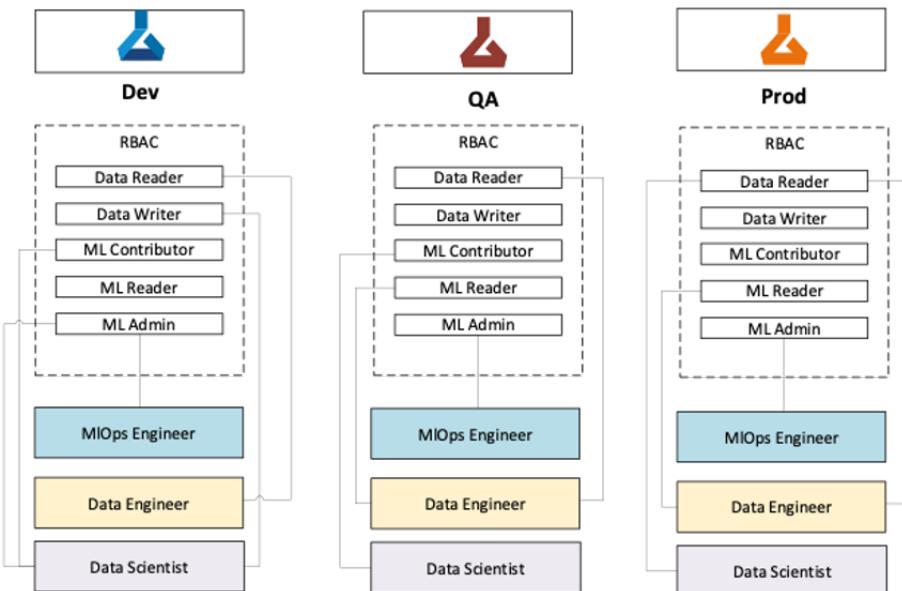
Azure RBAC を用いて Azure ML Workspace の内部リソースへのアクセスを制御する “ロール” をセキュリティプリンシパル (ユーザ、グループ、サービスプリンシパル、マネージド ID) に割り当てます。

下記の 4 つ組み込みロールがありますが、十分でない場合はカスタムロールを作成します。

Role	アクセス レベル
AzureML データ サイエンティスト	コンピューティング リソースの作成または削除とワークスペース自体の変更を除く、Azure Machine Learning ワークスペース内のすべてのアクションを実行できます。
Reader	ワークスペースでの読み取り専用のアクション。閲覧者はワークスペースで資産 (データストア の資格情報を含む) を一覧および表示できます。閲覧者がこれらの資産を作成または更新することはできません。
Contributor	ワークスペース内の資産を表示、作成、編集、削除 (該当する場合) します。たとえば、共同作成者は実験を作成したり、コンピューティング クラスターを作成またはアタッチしたり、実行を送信したり、Web サービスをデプロイしたりできます。
所有者	ワークスペース内の資産を表示、作成、編集、削除 (該当する場合) する機能など、ワークスペースへのフル アクセス。また、ロールの割り当てを変更することができます。

ワークフローに応じたロールの設定

信頼性・セキュリティを高めるために、ワークフローや組織体制に応じて Workspace を分割し、それに対して適切な権限設定を行う。



	Dev	QA	Prod
Data Engineer			
Data Scientist			
MLOps Engineer			

Most permitted permissions

Medium restricted permissions

Highly restricted permissions

データ保護

ガイドライン・実装手順

- カスタマーマネージドキーの利用

カスタマーマネージドキーの利用

Azure Machine Learning では、デフォルトで設定されている Microsoft マネージドな Key だけでなく、お客様管理の Key (カスタマーマネージドキー) の両方を使用した転送中/保存中の暗号化をサポートしています。

カスタマーマネージドキーを有効にした場合は、元々 Microsoft Subscription で管理していた Cosmos DB などの Azure サービスがユーザーの Subscription にデプロイされるため、追加のコストが発生します。

Ignite2020Demo ▾
Private dashboard

+ New dashboard ▾ Refresh Full screen | Edit Share Download Clone Assign to...
Auto refresh : Every hour

Resources
ignite2020demo

ignite2020demoacr	Container registry	eastus2euap
ignite2020demoai	Application Insights	South Central US
ignite2020demokv	Key vault	eastus2euap
ignite2020demosa	Storage account	eastus2euap
ignite2020demovnet	Virtual network	eastus2euap

Resources
ignite2020demows_53b28649-aa2a-4783-9ac1-5186e0fc9ba8

cosmosdb895056896	Azure Cosmos DB acco...	eastus2euap
ignite2020de787886a4	Search service	eastus2euap
sa1814799360	Storage account	eastus2euap
vnet	Virtual network	eastus2euap

Cosmos DB などがユーザの Subscription 内で立ち上がる (課金対象)



システム監視

基礎知識

- Azure Monitor 概要

検討事項

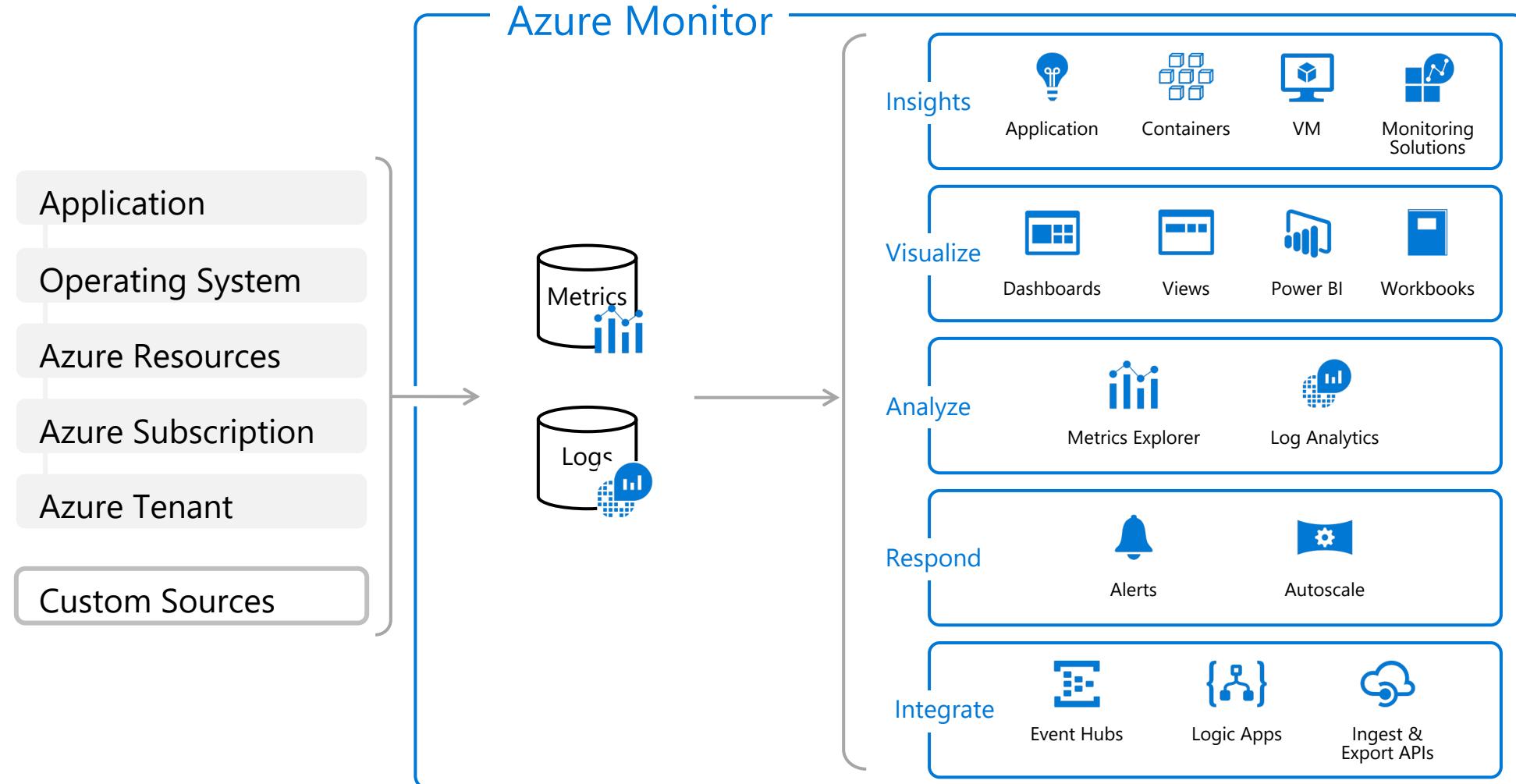
- 監視データの取得と分析方法の検討

ガイドライン・実装手順

- 監視ダッシュボードのテンプレート

Azure Monitor 概要

ログの収集、分析とその後の対応をカバーするサービス



監視データの取得と分析方法の検討

チームで利用する分析環境やミッションクリティカルな推論環境として Azure Machine Learning を利用している場合などには監視の仕組みを本格的に検討する必要があります。Azure Machine Learning では Azure Monitor の機能を利用して一元的に監視データの生成、分析、またはそれに基づくアラート発報ができます。

監視データ	概要	格納先	データ保持期間	一般的な分析
アクティビティログ	ワークスペース、計算リソースの作成・更新など	Azure Monitor に自動で収集・格納されるが、他の場所にルーティングすることもできる。	Azure Monitor の場合 : 90日間保持される。	Azure Monitor の可視化
プラットフォームメトリック	実験の実行、モデルの統計情報、クオータ情報など	Azure Monitor に自動で収集・格納されるが、他の場所にルーティングすることもできる。 93日間保持される。	Azure Monitor の場合 : 90日間保持される。	Azure Monitor の可視化
リソースログ	アセットやジョブの作成・削除・読み取りのイベント情報など	“診断設定”から明示的に収集するログの種類とルーティング先を指定する。	Log Analytics の場合 : 最小 30 日間、最大 730 日間保持される。	<ul style="list-style-type: none">Log Analytics でのクエリAzure Monitor の可視化

監視データの取得と分析方法の検討 (cont'd)

アクティビティログ

操作	説明
Machine Learning ワークスペースを作成または更新します	ワークスペースが作成または更新されました
CheckComputeNameAvailability	コンピューティング名が既に使用されているかどうかを確認します
コンピューティング リソースを作成または更新します	コンピューティング リソースが作成または更新されました
コンピューティング リソースを削除します	コンピューティング リソースが削除されました
シークレットのリスト	Machine Learning ワークスペースの操作のシークレットのリスト

メトリック – モデル

メトリック	ユニット	説明
モデル登録成功	Count	このワークスペースで成功したモデル登録の数
モデル登録失敗	Count	このワークスペースで失敗したモデル登録の数
モデル デプロイ開始	Count	このワークスペースで開始されたモデル デプロイの数
モデル デプロイが成功しました	Count	このワークスペースで成功したモデル デプロイの数
モデル デプロイ失敗	Count	このワークスペースで失敗したモデル デプロイの数

監視データの取得と分析方法の検討 (cont'd)

メトリック – リソース

メトリック	ユニット	説明
CpuUtilization	Count	CPU ノードの使用率 (%)。使用率は 1 分間隔で報告されます。
GpuUtilization	Count	GPU ノードの使用率 (%)。使用率は 1 分間隔で報告されます。
GpuMemoryUtilization	Count	GPU ノードのメモリ使用率 (%)。使用率は 1 分間隔で報告されます。
GpuEnergyJoules	Count	GPU ノードでのコンセントのエネルギーの間隔 (ジュール単位)。エネルギーは 1 分間隔で報告されます。

メトリック – クオータ

メトリック	ユニット	説明
ノード総数	Count	ノードの合計数。この合計には、アクティブ ノード、アイドル状態のノード、使用できないノード、割り込まれたノード、終了中のノードなどが含まれます
アクティブなノード	Count	アクティブなノードの数。ジョブをアクティブに実行しているノード。
アイドル状態のノード	Count	アイドル状態のノードの数。アイドル状態のノードは、ジョブを実行していないノードですが、使用可能な場合は新しいジョブを受け入れることができます。
使用できないノード	Count	使用できないノードの数。使用できないノードは、いくつかの問題が解決されていないため、機能していません。これらのノードは Azure によってリサイクルされます。
割り込まれたノード	Count	割り込まれたノードの数。これらのノードは低優先度のノードであり、使用可能なノード プールから外されます。
終了中のノード	Count	終了中のノードの数。終了中のノードは、ジョブの処理を完了したばかりで、アイドル状態になるノードです。
コアの合計	Count	コアの合計数
アクティブなコア	Count	アクティブなコアの数
アイドル状態のコア	Count	アイドル状態のコアの数
使用できないコア	Count	使用できないコアの数
割り込まれたコア	Count	割り込まれたコアの数
終了中のコア	Count	終了中のコアの数
クオータ使用率	Count	クオータ使用率 (%)

監視データの取得と分析方法の検討 (cont'd)

メトリック - 実行

メトリック	ユニット	説明
Cancelled Runs (取り消された実行数)	Count	このワークスペースに対して取り消された実行の数。実行が正常に取り消されたときに、カウントが更新されます。
Cancel Requested Runs (キャンセルが要求された実行数)	Count	このワークスペースに対してキャンセルが要求された実行の数。実行のキャンセル要求が受信されたときに、カウントが更新されます。
完了した実行数	Count	このワークスペースに対して正常に完了した実行の数。実行が完了し、出力が収集されたときに、カウントが更新されます。
失敗した実行	Count	このワークスペースに対して失敗した実行の数。実行に失敗すると、カウントが更新されます。
Finalizing Runs (終了処理中の実行数)	Count	このワークスペースに対して終了処理状態になった実行の数。実行は完了しているものの、出力の収集がまだ進行中の場合に、カウントが更新されます。
Not Responding Runs (応答していない実行数)	Count	このワークスペースに対して応答していない実行の数。実行が応答していない状態になったときに、カウントが更新されます。
Not Started Runs (未開始の実行数)	Count	このワークスペースに対して未開始状態の実行の数。実行を作成するために要求が受信されたものの、実行情報がまだ設定されていない場合に、カウントが更新されます。
Preparing Runs (準備中の実行数)	Count	このワークスペースに対して準備中の実行の数。実行環境の準備中に実行が準備状態になると、カウントが更新されます。
Provisioning Runs (プロビジョニング中の実行数)	Count	このワークスペースに対してプロビジョニング中の実行の数。実行でのコンピューティング先の作成の待機中、またはプロビジョニング中に、カウントが更新されます。
Queued Runs (キューに入れられた実行数)	Count	このワークスペースに対してキューに入れられた実行の数。コンピューティング先で実行がキューに入れられたときに、カウントが更新されます。必要なコンピューティングノードの準備が整うまで待機しているときに発生する場合があります。
開始した実行数	Count	このワークスペースに対して実行されている実行の数。必要なリソースに対して実行が開始されたときに、カウントが更新されます。
Starting Runs (開始中の実行数)	Count	このワークスペースに対して開始された実行の数。実行の作成要求の後、および実行 ID などの実行情報が設定された後に、カウントが更新されます。
エラー	Count	このワークスペースの実行エラーの数。実行時にエラーが発生するたびに、カウントが更新されます。
警告	Count	このワークスペースの実行警告の数。実行時に警告が発生するたびに、カウントが更新されます。

監視データの取得と分析方法の検討 (cont'd)

リソースログ

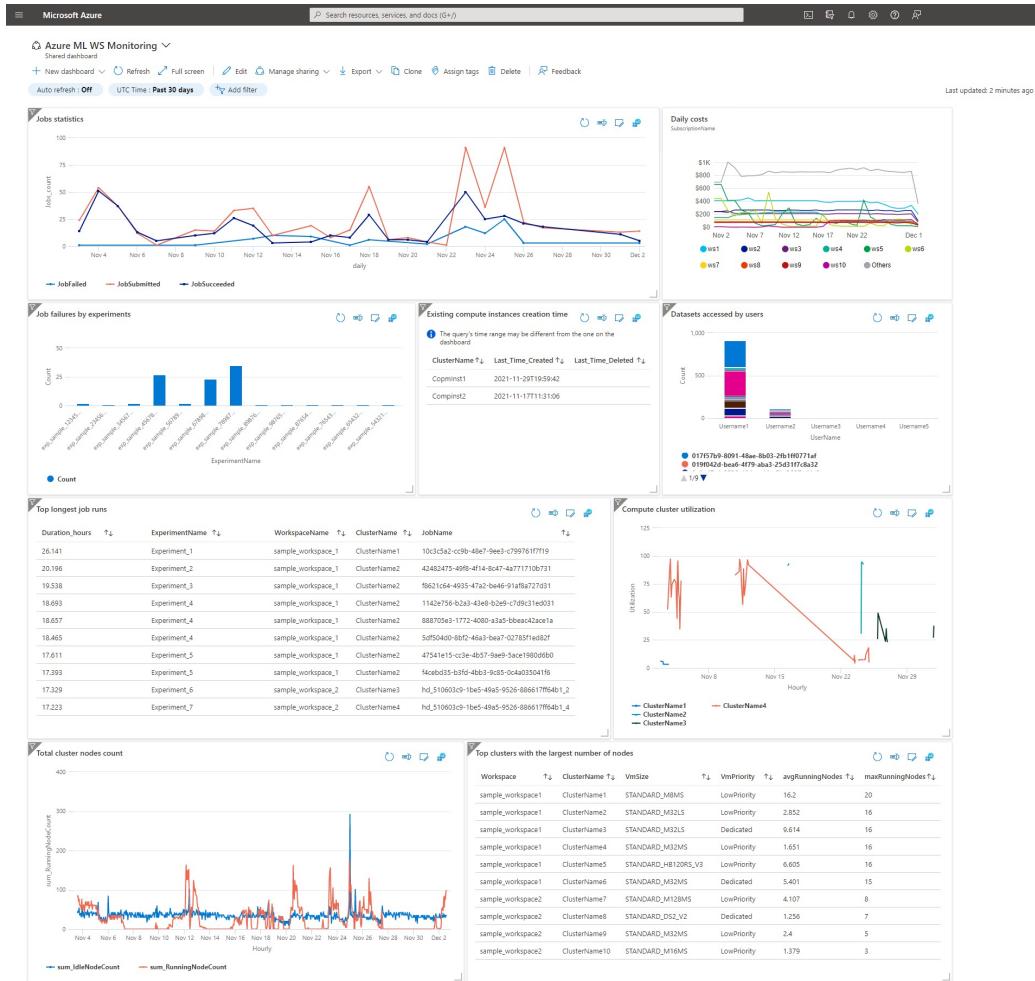
カテゴリ	説明	カテゴリ	説明
AmlComputeClusterEvent	Azure Machine Learning コンピューティング クラスターからのイベント。	DataLabelChangeEvent	データ ラベルまたはそのプロジェクトが作成または削除されたときのイベント。
AmlComputeClusterNodeEvent (非推奨)	Azure Machine Learning コンピューティング クラスター内のノードからのイベント。	DataLabelReadEvent	データ ラベルまたはそのプロジェクトが読み取られたときのイベント。
AmlComputeJobEvent	Azure Machine Learning コンピューティングで実行されているジョブからのイベント。	ComputeInstanceEvent	ML コンピューティング インスタンスがアクセスされたときのイベント (高頻度)。
AmlComputeCpuGpuUtilization	ML サービス コンピューティングの CPU と GPU の使用率ログ。	DataStoreChangeEvent	ML データストアが作成または削除されたときのイベント。
AmlRunStatusChangedEvent	ML の実行状態の変化。	DataStoreReadEvent	ML データストアが読み取られたときのイベント。
ModelsChangeEvent	ML モデルのアクセス、作成、または削除があったときのイベント。	DataSetChangeEvent	ML データストアが作成または削除されたときのイベント。
ModelsReadEvent	ML モデルが読み込まれたときのイベント。	DataSetReadEvent	ML データストアが読み取られたときのイベント。
ModelsActionEvent	ML モデルがアクセスされたときのイベント。	PipelineChangeEvent	ML パイプラインのドラフト、エンドポイント、モジュールが作成または削除されたときのイベント。
DeploymentReadEvent	モデル デプロイが読み込まれたときのイベント。	PipelineReadEvent	ML パイプラインのドラフト、エンドポイント、モジュールが読み取られたときのイベント。
DeploymentEventACI	ACI でモデル デプロイが発生したときのイベント (高頻度)。	RunEvent	ML 実験が作成または削除されたときのイベント。
DeploymentEventAKS	AKS でモデル デプロイが発生したときのイベント (高頻度)。	RunReadEvent	ML 実験が読み取られたときのイベント。
InferencingOperationAKS	コンピューティングの種類が AKS である、推論または関連操作のイベント。		
InferencingOperationACI	コンピューティングの種類が ACI である、推論または関連操作のイベント。		
EnvironmentChangeEvent	ML 環境の構成が作成または削除されたときのイベント。		
EnvironmentReadEvent	ML 環境の構成が読み取られたときのイベント (高頻度)。		

監視ダッシュボードのテンプレート

Log Analytics を用いた Azure Machine Learning のダッシュボードのテンプレートが利用できます。



[Create an Azure ML monitoring Dashboard –Template
\(github.com\)](https://github.com)



Green AI

Azure Machine Learning の計算環境のリソース消費や GPU のエネルギーコストをメトリックとして収集しており、持続可能な機械学習をサポートします。

Microsoft Azure Machine Learning Studio

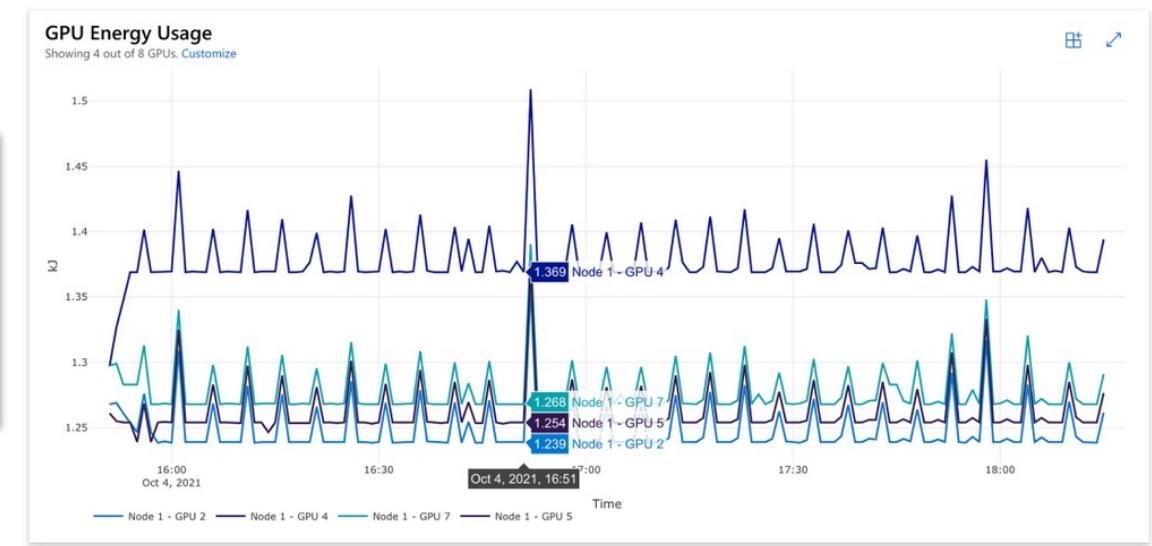
Microsoft > datacachetest > Experiments > Default > busy_boot_hhvij00c

busy_boot_hhvij00c

Refresh Connect to compute Resubmit Cancel Delete Time range Entire run

Details Metrics Images Child runs Outputs + logs Snapshot Explanations (preview) Fairness (preview) Monitoring (preview)

Average CPU Utilization 2.5% Average GPU Utilization 12.1% Average GPU Memory Usage 0.20 GB Total GPU Energy Usage 4994.92 kJ



可視化の例

監査ポリシー

基礎知識

- Azure Policy 概要

ガイドライン・実装手順

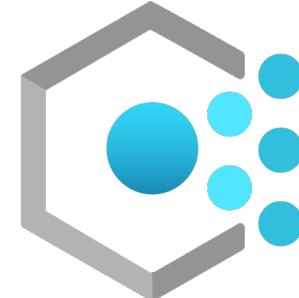
- 組み込みPolicy の使用
- クオータ (Quota) の設定ガイドライン

Azure Policy 概要

Azure Policy は、Azure 内のリソースの設定をビジネスルールに準拠するように管理できる Azure のサービスです。リソースを制御・監査するポリシーを作成し、割り当て、管理します。

1. ポリシー定義の作成

評価方法と対処方法を定義します。



2. ポリシー定義の割り当て

特定のスコープ内で実行されるように割り当てます。

3. 評価結果の確認

評価されると各リソースは “準拠” または “非準拠” としてマークされます。

組み込みポリシーの使用

Azure Policy を利用して Azure Machine Learning の状態を監視して管理することができます。組み込みで提供しているポリシーの種類は下記になります。(2022年4月)

ポリシー	説明
カスタマー マネージド キー	ワークスペースでカスタマー マネージド キーを使用する必要があることを監査または適用します。
プライベート リンク	ワークスペースで仮想ネットワークとの通信にプライベート エンドポイントが使用されているかどうかを監査または強制します。
プライベート エンドポイント	プライベート エンドポイントを作成する必要がある Azure Virtual Network サブネットを構成します。
プライベート DNS ゾーン	プライベート リンクに使用するプライベート DNS ゾーンを構成します。
ユーザー割り当てマネージド ID	ワークスペースでユーザー割り当てのマネージド ID が使用されているかどうかを監査または強制します。
パブリック ネットワーク アクセスの無効化	ワークスペースでパブリック インターネットからのアクセスを無効にするかどうかを監査または強制します。
ローカル認証 (SSH) の無効化	Azure Machine Learning のコンピューティング リソースでローカル認証方法を無効にするべきかどうかを監査または強制します。
ローカル認証 (SSH) の変更または無効化	ローカル認証方法を無効にするためにコンピューティング リソースを構成します。
コンピューティング クラスターおよびインスタンスが仮想ネットワークの背後にいる	コンピューティング リソースが仮想ネットワークの背後にあるかどうかを監査します。

クオータ (Quota) の設定ガイドライン

企業・組織で Azure Machine Learning を利用する際は、計算リソース (Compute) に対する Quota (クオータ) の確認と設定を推奨します。

- Azure Machine Learning においてクオータ (Quota) は Compute Instance と Compute Clusters に対して設定します。
 - 初期の Azure Subscription によってはクオータが 0 になっている場合があります。Azure Machine Learning の利用開始前にクオータの値を確認してください。
- クオータの適用範囲
 - Azure Subscription : VM のインスタンスタイプごとに利用可能な CPU コア数の上限を申請する。
 - Azure ML ワークスペース : 管理者がワークスペース単位でユーザが利用できるコア数を VM のインスタンスタイプごとに制限する。

脆弱性管理

ガイドライン・実装手順

- VM やコンテナの脆弱性管理

VM やコンテナの脆弱性管理

脆弱性の管理は Microsoft とお客様の共同責任です。Microsoft の対策について理解し、ユーザ側での対応方針について考えていく必要があります。

Microsoft マネージド VM イメージ

- Azure ML の Compute Clusters と Compute Instance、Azure Data Science VM のホスト OS VM イメージは、ほとんどのケースでは毎月更新される。
- Compute Instance の更新は Azure ML SDK のリリース周期に合わせて実行される。
- 脆弱性のスキャンを定期実行し、必要に応じて修正を実施する。
- Compute Instance はプロビジョニング時に最新 VM のイメージを取得する仕組みになっているが、現時点（2022年4月）で OS イメージのバージョンアップはサポートされていない。再作成が必要になる。

VM やコンテナの脆弱性管理 (cont'd)

Microsoft マネージド コンテナイメージ

- Azure Machine Learning が管理する[基本 docker イメージ](#)は、新たに検出された脆弱性に対処するためにセキュリティ パッチを頻繁に取得します。Compute Instance の更新は Azure ML SDK のリリース周期に合わせて実行される。
- Azure Machine Learning は、脆弱性に対処するために、サポートされているイメージの更新プログラムを 2 週間ごとにリリースします。Microsoft では、コミットメントとして、サポートされるイメージの最新バージョンで 30 日を超える脆弱性が存在しないことを目指している。
- 基本 docker イメージをベースに追加でパッケージをインストールしたり、ユーザ独自の基本 docker イメージを仕様する場合は、ユーザ側で脆弱性の管理を行っていく必要が出てきます。
- docker イメージは Azure Container Registry に格納されます。Microsoft Defender for Containers を使用して、イメージの脆弱性スキャンを実行することも可能です。

コスト管理

基礎知識

- Azure Machine Learning のコスト

ガイドライン・実装手順

- コスト軽減方法のポイント
- マネージド計算リソースのコスト

Azure Machine Learning におけるコスト

Azure Machine Learning には多数の機能が内蔵されていたり、複数の Azure サービスとの組み合わせ・連携によって構成されるためコストの仕組みは複雑です。



コスト軽減方法のポイント

Azure Machine Learning の機能を有効に活用してコストを抑えることができます。

- Compute Cluster の自動スケールアウト・ダウン機能の活用
- Subscription, Workspace ごとのクオータの設定
- Job の終了ポリシーの設定
- 低優先度 VM の利用
- Compute Instance の自動起動・停止のスケジュール設定
- Azure Arc を含むローカル環境の活用
- Job の並列化
- 中間データのライフサイクルや削除ポリシーの設定
- 同一リージョンへのデプロイによるネットワークコストの削減

[コストの管理と最適化 - Azure Machine Learning | Microsoft Docs](#)

[組織規模での Azure Machine Learning の予算、コスト、クオータを管理する - Cloud Adoption Framework | Microsoft Docs](#)

マネージド計算リソースのコスト

Azure Machine Learning のマネージドな計算環境の Compute Clusters と Compute Instance は非常に似ているが、コストの考え方方が違います。

Computer Clusters

- 起動中のみコストが発生
- コスト対象リソース
 - VM
 - Load Balancer
 - Managed Disk
 - IP Address (Public IP の場合)

Computer Instance

- 停止中でも一部コストが発生
- コスト対象リソース
 - VM
 - Load Balancer
 - Managed Disk
 - IP Address (Public IP の場合)

} 停止中でもコストが発生するリソース



Microsoft AI





© Copyright Microsoft Corporation. All rights reserved.