

サイエンティフィック・システム研究会
オブジェクトストレージ座談会



Object storage on Microsoft Azure

日本マイクロソフト株式会社
パブリックセクター事業本部
文教営業本部

中田 寿穂



な か た ひ さ ほ
中 田 寿 穂

Hisaho.Nakata@microsoft.com

日本マイクロソフト株式会社
パブリックセクター事業本部
クラウドアーキテクト

About me



Academics (~ 1999)

新超伝導体の開発(強相関係の物理、低温物理学、結晶学)



High Performance Computing (1999~2006)

PCクラスタシステムの開発、販売



Cloud Mail, ID Management, SSO (2006.7~)

Google Apps, Office 365 の導入支援



Cloud Services (2015.08~)

Microsoft のクラウド製品(Azure, Office 365)のプリセールス
Azure を使用したサービス開発



Academics (2021.04~)

機械学習 & ロボティクス

Azure Storage サービスの概要

機能	説明	使用する場合
Azure Files	業界標準のサーバー メッセージ ブロック (SMB) プロトコル経由でどこからでもアクセスできるフル マネージドのクラウド ファイル共有を提供します。Azure ファイル共有は、Windows、Linux、macOS のクラウドまたはオンプレミスのデプロイからマウントできます。	既に、ネイティブなファイル システム API を使用して、Azure で実行されている他のアプリケーションとの間でデータを共有しているアプリケーションをクラウドに“リフト アンド シフト”したい場合。 オンプレミスのファイル サーバーまたは NAS デバイスを置き換えるか、またはそれを補完したい場合。 多くの仮想マシンからアクセスする必要がある開発ツールとデバッグ ツールを格納する。
Azure BLOB	非構造化データを大規模にブロック BLOB に格納してアクセスできるようにします。 エンタープライズ ビッグ データ分析ソリューション用の Azure Data Lake Storage Gen2 もサポートされています。	アプリケーションでストリーミングとランダム アクセスのシナリオに対応する。 アプリケーションのデータにどこからでもアクセスできるようにする。 Azure 上にエンタープライズ Data Lake を構築し、ビッグ データ分析を実行する。
Azure ディスク	データを永続的に格納し、アタッチされた仮想ハード ディスクからアクセスできるようにします。	ネイティブなファイル システム API を使用して永続ディスクとの間のデータの読み取りと書き込みを行っているアプリケーションを“リフト アンド シフト”したい場合。 外部 (ディスクがアタッチされている仮想マシンの外) からのアクセスが不要なデータを格納する。
Azure キュー	アプリケーション コンポーネント間の非同期メッセージ キューを可能にします。	アプリケーション コンポーネントを分離し、非同期メッセージングを使用してそれらの間の通信を行いたい場合。 Queue Storage および Service Bus キューをそれぞれいつ使用するかに関するガイダンスについては、 「Storage キューと Service Bus キューの比較」 を参照してください。
Azure テーブル	構造化された NoSQL データをクラウドに格納し、スキーマなしの設計でキー/属性ストアを実現できるようにします。	Web アプリケーションのユーザー データ、アドレス帳、デバイス情報、サービスに必要なその他の種類のメタデータのような柔軟なデータセットを格納したい場合。 テーブル ストレージと Azure Cosmos DB Table API をそれぞれいつ使用するかに関するガイダンスについては、 「Azure Cosmos DB Table API と Azure Table Storage を使用した開発」 を参照してください。

Azure Storage サービスの概要

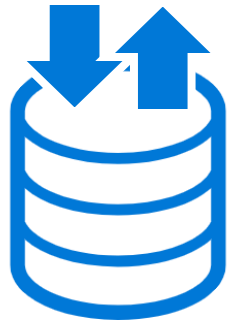
機能	説明	使用する場合
Azure Files	業界標準のサーバー メッセージ ブロック (SMB) プロトコル経由でどこからでもアクセスできるフル マネージドのクラウド ファイル共有を提供します。Azure ファイル共有は、Windows、Linux、macOS のクラウドまたはオンプレミスのデプロイからマウントできます。	既に、ネイティブなファイル システム API を使用して、Azure で実行されている他のアプリケーションとの間でデータを共有しているアプリケーションをクラウドに“リフト アンド シフト”したい場合。 オンプレミスのファイル サーバーまたは NAS デバイスを置き換えるか、またはそれを補完したい場合。 多くの仮想マシンからアクセスする必要のある開発ツールとデバッグ ツールを格納する。
Azure BLOB	非構造化データを大規模にブロック BLOB に格納してアクセスできるようにします。エンタープライズ ビッグ データ分析ソリューション用の Azure Data Lake Storage Gen2 もサポートされています。	アプリケーションでストリーミングとランダム アクセスのシナリオに対応する。 アプリケーションのデータにどこからでもアクセスできるようにする。 Azure 上にエンタープライズ Data Lake を構築し、ビッグ データ分析を実行する。
Azure ディスク	データを永続的に格納し、アタッチされた仮想ハード ディスクからアクセスできるようにします。	ネイティブなファイル システム API を使用して永続ディスクとの間のデータの読み取りと書き込みを行っているアプリケーションを“リフト アンド シフト”したい場合。 外部（ディスクがアタッチされている仮想マシンの外）からのアクセスが不要なデータを格納する。
Azure キュー	アプリケーション コンポーネント間の非同期メッセージ キューを可能にします。	アプリケーション コンポーネントを分離し、非同期メッセージングを使用してそれらの間の通信を行いたい場合。 Queue Storage および Service Bus キューをそれぞれいつ使用するかに関するガイダンスについては、 「Storage キューと Service Bus キューの比較」 を参照してください。
Azure テーブル	構造化された NoSQL データをクラウドに格納し、スキーマなしの設計でキー/属性ストアを実現できるようにします。	Web アプリケーションのユーザー データ、アドレス帳、デバイス情報、サービスに必要なその他の種類のメタデータのような柔軟なデータセットを格納したい場合。 テーブル ストレージと Azure Cosmos DB Table API をそれぞれいつ使用するかに関するガイダンスについては、 「Azure Cosmos DB Table API と Azure Table Storage を使用した開発」 を参照してください。

Azure BLOB ストレージ

Azure Blob Storage は、Microsoft のクラウド用オブジェクト ストレージ ソリューションです。
BLOB ストレージは、テキスト データやバイナリ データなどの大量の非構造化データを格納するために最適化されています。

BLOB ストレージは、次の目的に最適です。

バックアップ
リストア



画像、ドキュメントを
ブラウザに直接配信



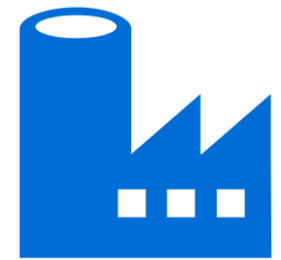
動画、ストリーミング
配信



分散アクセス用の
ファイルの格納



ビッグデータ解析用
データレイク



利用用途に合わせて3種類のBLOB

ブロック BLOB

テキストとバイナリ データが格納されます。ブロック BLOB は、個別に管理できるデータブロックで構成されます。ブロック BLOB には、最大約 4.75 TB のデータが格納されます。プレビューではもっと大きなブロック BLOB を利用できます。最大約 190.7 TB です。

追加 BLOB

ブロック BLOB と同様にブロックで構成されますが、追加操作用に最適化されています。追加 BLOB は、仮想マシンのデータのログ記録などのシナリオに最適です。

ページ BLOB

最大 8 TB のランダム アクセス ファイルを格納できます。ページ BLOB は、仮想ハードドライブ (VHD) ファイルを格納し、Azure 仮想マシン用のディスクとして機能します。

データの冗長化

ローカル冗長ストレージ

1 つのデータセンター内でデータのコピーを複数保持することで、1 年にわたって 99.999999999% (イレブン ナイン) 以上のオブジェクトの耐久性を提供するように設計されています。

ゾーン冗長ストレージ

複数のデータセンターまたは複数のリージョン間でデータのコピーを複数保持することで、1 年にわたって 99.999999999% (トゥエルブ ナイン) 以上のオブジェクトの耐久性を提供するように設計されています。

Geo 冗長ストレージ

あるリージョンでデータのコピーを複数保持し、2 つ目のリージョンに非同期にレプリケートすることで、1 年にわたって 99.99999999999999% (シックスティーン ナイン) 以上のオブジェクトの耐久性を提供するように設計されています。

読み取りアクセスの Geo 冗長ストレージ

GRS で使用される 2 つ目のリージョンからの読み取りアクセスを許可することで、1 年にわたって 99.99999999999999% (シックスティーン ナイン) 以上のオブジェクトの耐久性と、読み取りの 99.99% の可用性を提供するように設計されています。

BLOB Storage へのアクセス(データの移動)



CLI ツール

Powershell、AzCopy(Windows, Linux, macOS 用のCLI)

Azure Storage Data Movement Library

.NET、Java、Node.js、Python、PHP、Ruby

Azure Data Factory

90種類以上のコネクタを提供

NFS (Preview)

NFS 3.0 プロトコルを使用してLinuxシステムからコンテナをマウント

blobfuse

Linuxで利用できる仮想ファイルシステムドライバ

Azure Databox

データ移行用専用のデバイスを貸し出し

Azure Import/Exportサービス

お客様の用意したHDDを使用

Azure BLOB ストレージ

ストレージ料金	Hot: ¥2.24/GB Cool: ¥1.68/GB Archive: ¥0.224/GB
---------	---

容量上限	200ストレージアカウント×500TB ※申請により上限5PBまで引き上げ可能
------	--

オブジェクト数の上限	無制限
------------	-----

1オブジェクト容量上限	4.77TB
-------------	--------

Hot 頻繁にアクセスされるデータの格納に最適化されています。

Cool アクセスされる頻度は低いものの、少なくとも 30 日以上保管されるデータの格納に最適化されています。

Archive ほとんどアクセスされず、少なくとも 180 日以上保管され、待ち時間の要件が柔軟（数時間単位）であるデータの格納に最適化されています。

Azure BLOB ストレージ

ホット アクセス層

ホット アクセス層は、クール層とアーカイブ層に比べてストレージ コストが高くなるものの、アクセス コストが最も低くなります。ホット アクセス層の使用シナリオの例には、次のようなものがあります。

- 活発に使用されたり、頻繁に読み取りおよび書き込みがされると予想されるデータ
- 処理段階にあり、最終的にはクール アクセス層に移行されるデータ

クール アクセス層

クール アクセス層は、ホット ストレージに比べてストレージ コストが低くなり、アクセス コストが高くなります。この層は、少なくとも 30 日以上クール層で保持されるデータを対象としています。クール アクセス層の使用シナリオの例には、次のようなものがあります。

- 短期的なバックアップおよびディザスター リカバリー
- 頻繁には使用されないが、アクセスされたときにはすぐに利用できることが期待されている古いデータ
- コスト効率よく格納する必要があり、さらなるデータが今後処理するため収集されている、大規模なデータ セット

アーカイブ アクセス層

アーカイブ アクセス層はストレージ コストが最も低くなりますが、ホットとクールの各層に比べてデータ取得コストが高くなります。データは、少なくとも 180 日間、アーカイブ層に保持される必要があります。

アーカイブ アクセス層の使用シナリオの例には、次のようなものがあります。

- 長期バックアップ、セカンダリ バックアップ、アーカイブ データセット
- 最終的に使用可能な形式に処理した後も保持する必要がある、元の（未加工の）データ
- 長期間格納しておく必要があり、ほとんどアクセスされることがない、コンプライアンスおよびアーカイブ データ

Azure Data Lake Storage Gen2

Azure Data Lake Storage は、BLOB Storage に基いて構築されており、ビッグデータ分析専用の機能セットが備わっています。

Azure Data Lake Services



HDInsight
Hadoop as a Service



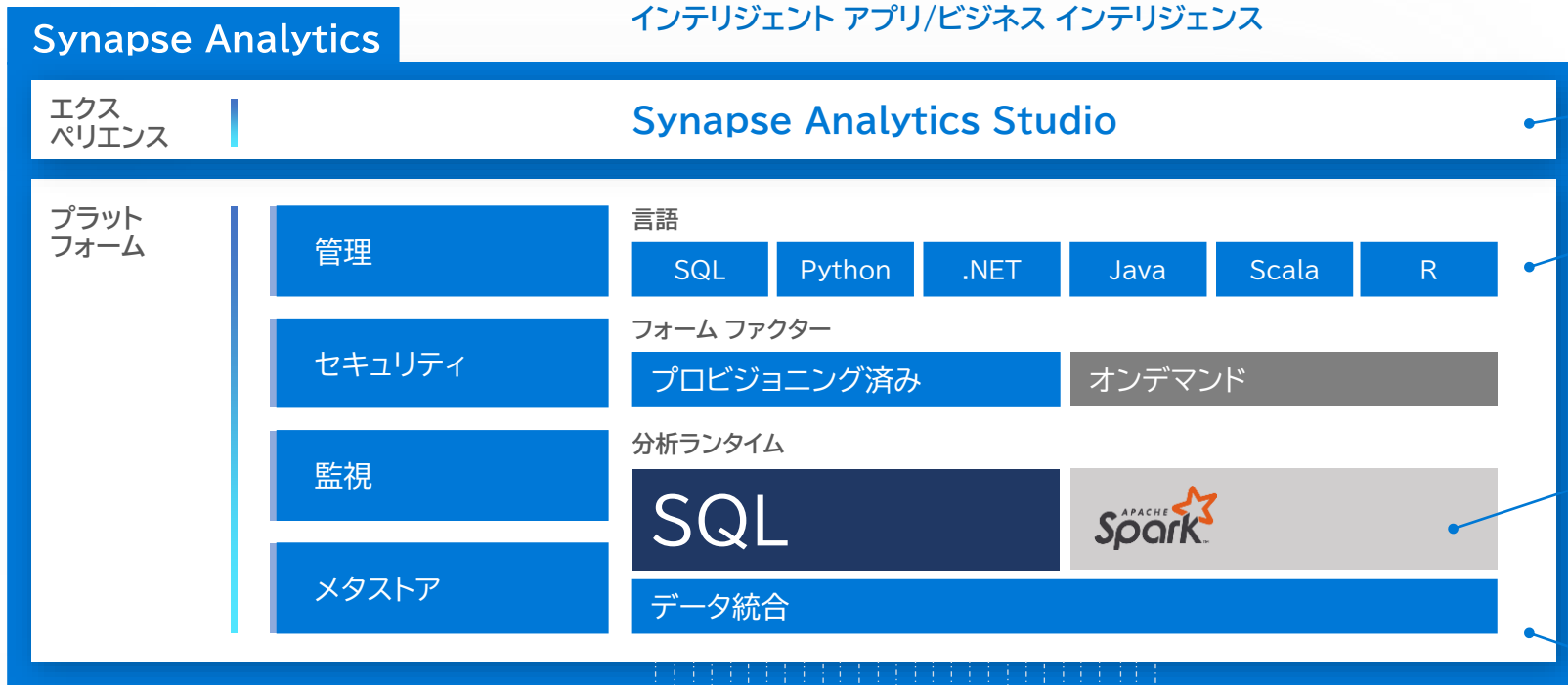
Analytics
Big Data Query
as a Service



Store
容量無制限、Raw Data
アクセスコントロール

Azure Synapse Analytics

人工知能/機械学習/モノのインターネット
インテリジェント アプリ/ビジネス インテリジェンス



あらゆる規模の分析ワークロードに向けて設計

コード不要、コード ファーストの SaaS 開発者エクスペリエンス

複数の言語でさまざまな分析ワークロードに適合

統合分析ランタイムをプロビジョニング済みおよびサーバーレス オンデマンドの形式で利用可能

SQL Analytics がバッチ処理、ストリーミング処理、対話型処理用に T-SQL を提供

Apache Spark による、Python、Scala、R、.NET を使用したビッグデータ処理

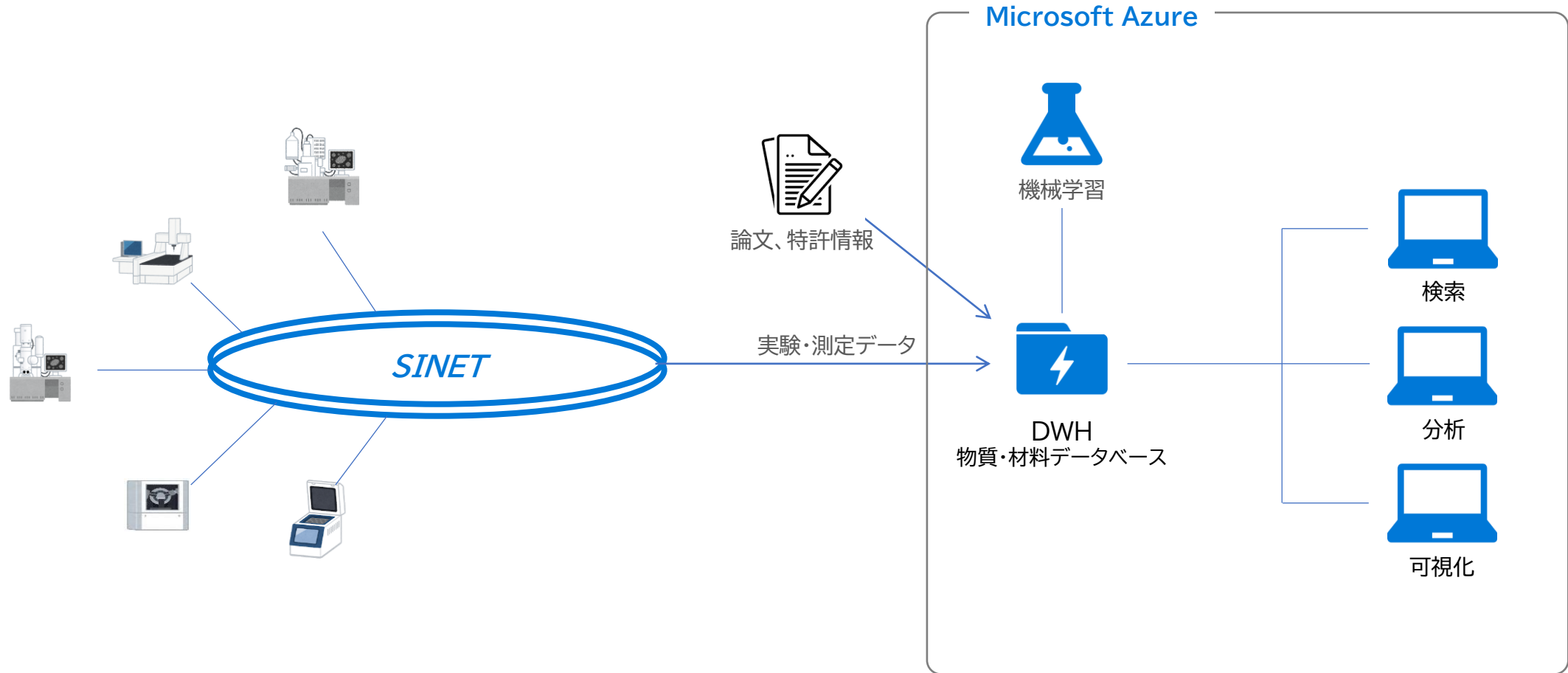
管理、セキュリティ、監視、メタストアのために統合されたプラットフォームサービス

統合されたデータレイクと Common Data Model への対応

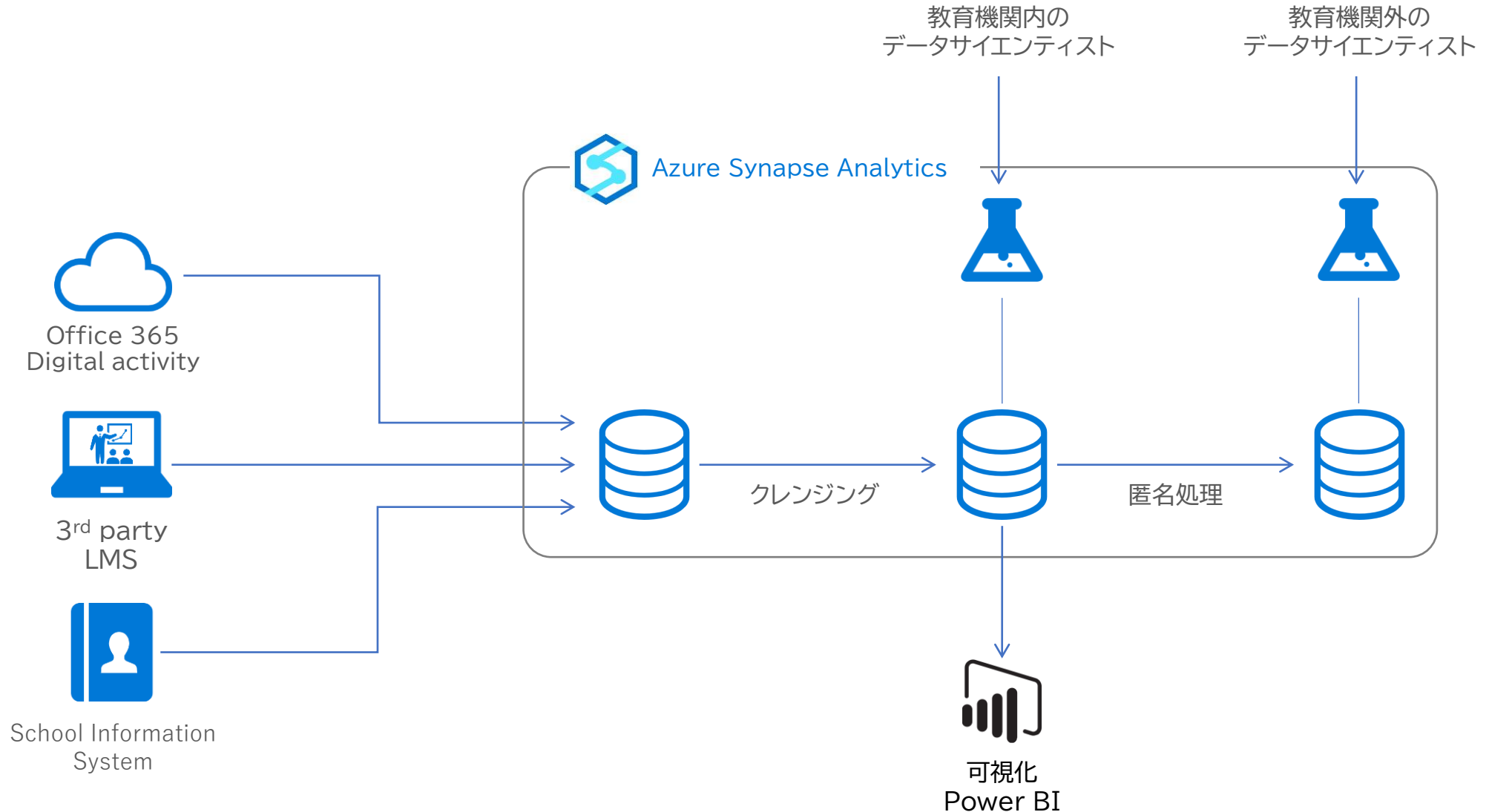
Azure Data Lake Storage

Common Data Model
エンタープライズ セキュリティ
分析に最適化

マテリアルインフォマティクス プラットフォーム



ラーニングアナリティクス プラットフォーム





© 2020 Microsoft Corporation. All rights reserved.

本資料は情報提供のみを目的としており、本資料に記載されている情報は、本資料作成時点でのマイクロソフトの見解を示したものです。状況等の変化により、内容は変更される場合があります。本資料に特別条件等が提示されている場合、かかる条件等は、貴社との有効な契約を通じて決定されます。それまでは、正式に確定するものではありません。従って、本資料の記載内容とは異なる場合があります。また、本資料に記載されている価格はいずれも、別段の表記がない限り、参考価格となります。貴社の最終的な購入価格は、貴社のリセラー様により決定されます。マイクロソフトは、本資料の情報に対して明示的、黙示的または法的な、いかなる保証も行いません。