

Microsoft Fabric の用語

[アーティクル] • 2023/06/07

Synapse Data Warehouse、Synapse Data Engineering、Synapse Data Science、Synapse Real-Time Analytics、Data Factory、Power BI に固有の用語など、Microsoft Fabric で使用される用語の定義について説明します。

① 重要

Microsoft Fabric は **プレビュー段階**です。

一般的な用語

- **容量**：容量は、特定の時点で使用できるリソースの専用セットです。容量では、アクティビティを実行したり、出力を生成したりするリソースの機能が定義されます。アイテムが異なると、一定の時間に消費される容量が異なります。Fabric では、Fabric SKU と試用版を通じて容量が提供されます。詳細については、「[容量とは](#)」を参照してください。
- **経験**：特定の機能を対象とする機能のコレクション。ファブリック エクスペリエンスには、Synapse Data Warehouse、Synapse Data Engineering、Synapse Data Science、Synapse Real-Time Analytics、Data Factory、Power BI が含まれます。
- **項目**：エクスペリエンス内の一連の機能を含む項目。ユーザーは、作成、編集、および削除できます。項目の種類ごとに異なる機能が提供されます。たとえば、Data Engineeringエクスペリエンスには、lakehouse、notebook、Spark ジョブ定義項目が含まれます。
- **テナント**：テナントは、organization用の Fabric の単一インスタンスであり、Azure Active Directory に合わせて配置されます。
- **ワークスペース**：ワークスペースは、コラボレーション用に設計された 1 つの環境でさまざまな機能を組み合わせる項目のコレクションです。これは、実行される作業の容量を活用するコンテナとして機能し、その中のアイテムにアクセスできるユーザー

ーを制御します。たとえば、ワークスペースでは、ユーザーはレポート、ノートブック、データセットなどを作成します。詳細については、[ワークスペースに関する記事を参照してください](#)。

Synapse Data Engineering

- **レイクハウス:** lakehouse は、ビッグ データ処理のために Apache Spark エンジンと SQL エンジンによって使用されるデータ レイク上のデータベースを表すファイル、フォルダー、およびテーブルのコレクションです。Lakehouse には、オープンソースの Delta 形式のテーブルを使用する場合の ACID トランザクションの機能が強化されています。Lakehouse アイテムは、[Microsoft OneLake](#) の一意のワークスペース フォルダー内でホストされます。フォルダーとサブフォルダーに整理されたさまざまな形式 (構造化および非構造化) のファイルが含まれています。詳細については、「[レイクハウスとは](#)」を参照してください。
- **ノートブック:** Fabric ノートブックは、豊富な機能を備えた多言語対話型プログラミング ツールです。これには、コードとマークダウンの作成、Spark ジョブの実行と監視、結果の表示と視覚化、チームとの共同作業が含まれます。これは、データ エンジニアとデータサイエンティストがデータを探索して処理し、コードとローコードの両方のエクスペリエンスを使用して機械学習実験を構築するのに役立ちます。オーケストレーション用のパイプライン アクティビティに簡単に変換できます。
- **Spark アプリケーション:** Apache Spark アプリケーションは、Spark の API 言語 (Scala、Python、Spark SQL、または Java) または Microsoft が追加した言語 (.NET と C# または F#) のいずれかを使用してユーザーが作成したプログラムです。アプリケーションを実行すると、並列で実行される 1 つ以上の Spark ジョブに分割され、データをより高速に処理できます。詳細については、「[Spark アプリケーションの監視](#)」を参照してください。
- **Apache Spark ジョブ:** Spark ジョブは、アプリケーション内の他のジョブと並列で実行される Spark アプリケーションの一部です。ジョブは複数のタスクで構成されます。詳細については、「[Spark ジョブの監視](#)」を参照してください。
- **Apache Spark ジョブ定義:** Spark ジョブ定義は、Spark アプリケーションの実行方法を示す、ユーザーによって設定されるパラメーターのセットです。これにより、バッチ ジョブまたはストリーミング ジョブを Spark クラスターに送信できます。詳細については、「[Apache Spark ジョブ定義とは](#)」を参照してください。

- **V オーダー**: Parquet ファイル形式への書き込み最適化により、高速読み取りが可能になり、コスト効率とパフォーマンスが向上します。既定では、すべての Fabric エンジンが v オーダーの Parquet ファイルを書き込みます。

Data Factory

- **コネクタ**: Data Factory には、さまざまな種類のデータ ストアに接続できる豊富なコネクタが用意されています。接続したら、データを変換できます。詳細については、「[コネクタ](#)」を参照してください。
- **データ パイプライン**: Data Factory では、データの移動と変換を調整するためにデータ パイプラインが使用されます。これらのパイプラインは、Fabric のデプロイ パイプラインとは異なります。詳細については、「Data Factory の [パイプラインの概要](#)」を参照してください。
- **データフロー Gen2**: データフローは、数百のデータ ソースからデータを取り込み、データを変換するためのローコード インターフェイスを提供します。Fabric のデータフローは、Dataflow Gen2 と呼ばれます。データフロー Gen1 は Power BI に存在します。Dataflow Gen2 には、Azure Data Factory または Power BI のデータフローに比べて追加の機能が用意されています。Gen1 から Gen2 にアップグレードすることはできません。詳細については、「Data Factory の概要」の「[データフロー](#)」を参照してください。

Synapse Data Science

- **データ ラングラー**: Data Wrangler は、探索的データ分析を実行するためのイマーシブ エクスペリエンスをユーザーに提供するノートブック ベースのツールです。この機能は、グリッドに似たデータ表示と動的な概要統計と一連の一般的なデータ クレンジング操作を組み合わせたものであり、選択したアイコンがいくつか用意されています。各操作では、再利用可能なスクリプトとしてノートブックに保存できるコードが生成されます。
- **実験**: 機械学習実験は、関連するすべての機械学習実行の organization と制御の主要な単位です。詳細については、「[Microsoft Fabric での機械学習の実験](#)」を参照してください。

- **モデル**：機械学習モデルは、特定の種類のパターンを認識するようにトレーニングされたファイルです。一連のデータに対してモデルをトレーニングし、そのデータ セットを推論して学習するために使用するアルゴリズムを提供します。詳細については、「[機械学習モデル](#)」を参照してください。
- **実行**：実行は、モデル コードの 1 回の実行に対応します。[MLflow](#) では、追跡は実験と実行に基づいています。

Synapse データ ウェアハウス

- **SQL エンドポイント**：各 Lakehouse には SQL エンドポイントがあり、ユーザーは TSQL を使用して TDS 経由でデルタ テーブルデータに対してクエリを実行できます。詳細については、「[SQL エンドポイント](#)」を参照してください。
- **Synapse Data Warehouse**：Synapse Data Warehouse機能は従来のデータ ウェアハウスであり、エンタープライズ データ ウェアハウスに期待される完全なトランザクション T-SQL 機能をサポートします。詳細については、「[Synapse Data Warehouse](#)」を参照してください。

Synapse Real-Time Analytics

- **KQL データベース**：KQL データベースは、それに対して KQL クエリを実行する形式でデータを保持するデータベースの表現です。詳細については、「[KQL データベースのクエリ](#)」を参照してください。
- **KQL クエリセット**：KQL Queryset は、Data Explorer データベースのデータに対するクエリの実行、結果の表示、およびクエリ結果の操作に使用される項目です。クエリセットには、データベースとテーブル、クエリ、結果が含まれます。KQL Queryset を使うと、クエリを保存して後で使ったり、クエリをエクスポートして他のユーザーと共有したりできます。詳細については、「[KQL Queryset でのデータのクエリ](#)」を参照してください。
- **イベント ストリーム**：Microsoft Fabric イベント ストリーム機能は、コードなしのエクスペリエンスでリアルタイム イベントをキャプチャし、変換し、宛先にルーティングするための Fabric プラットフォームの一元的な場所を提供します。イベント ストリ

ームは、さまざまなストリーミング データ ソース、インジェスト先、および変換が必要な場合のイベント プロセッサで構成されます。詳細については、「[Microsoft Fabric イベント ストリーム](#)」を参照してください。

OneLake

- **ショートカット**：ショートカットは、他のファイル ストアの場所を指す OneLake 内の埋め込み参照です。直接コピーしなくても既存のデータに接続する方法が提供されます。詳細については、「[OneLake ショートカット](#)」を参照してください。

次のステップ

- [Microsoft Fabric ホーム ページからアイテムに移動する](#)
- [OneLake データ ハブ内のデータ項目を検出する](#)
- [Microsoft Fabric のエンド ツー エンドのチュートリアル](#)