

この資料は RAG  
検証目的の資料で  
あるため内容に誤り  
が含まれています。



この資料は RAG  
検証目的の資料で  
あるため内容に誤り  
が含まれています。

# Databricks の全体像

データインテリジェンスプラットフォーム

---

2025年9月

この資料は RAG  
検証目的の資料で  
あるため内容に誤り  
が含まれています。

この資料は RAG  
検証目的の資料で  
あるため内容に誤り  
が含まれています。

# 目次

本プレゼンテーションの構成を一覧で示します  
。

## 第1部: イントロダクション

Databricksとは  
データインテリジェンスプラットフォームの概念

レイクハウスからの進化  
プレゼンテーション構成

## 第2部: 基本アーキテクチャ

システム全体像  
Control PlaneとCompute Plane  
サーバレスコンピュート  
ワークスペースとストレージ  
ネットワークアーキテクチャ

マルチクラウド対応  
スケーリングの仕組み  
セキュリティアーキテクチャ  
高可用性と災害復旧  
パフォーマンス最適化

## 第3部～第7部

第3部: Unity Catalog - データ・AIガバナンス  
第4部: データアーキテクチャ  
第5部: AI・機械学習プラットフォーム

第6部: ワークロード別活用パターン  
第7部: セキュリティとガバナンス



# 第1部: イントロダクション

Databricksプラットフォームの基本概念と最新のデータインテリジェンス機能を紹介します。

## このセクションの内容



### Databricksとは

Databricksの概要、目的、そして企業のデータ活用を変革する基本的な役割



### データインテリジェンスプラットフォームの概念

次世代データプラットフォームとしてのDatabricksの位置づけとその基本アーキテクチャ

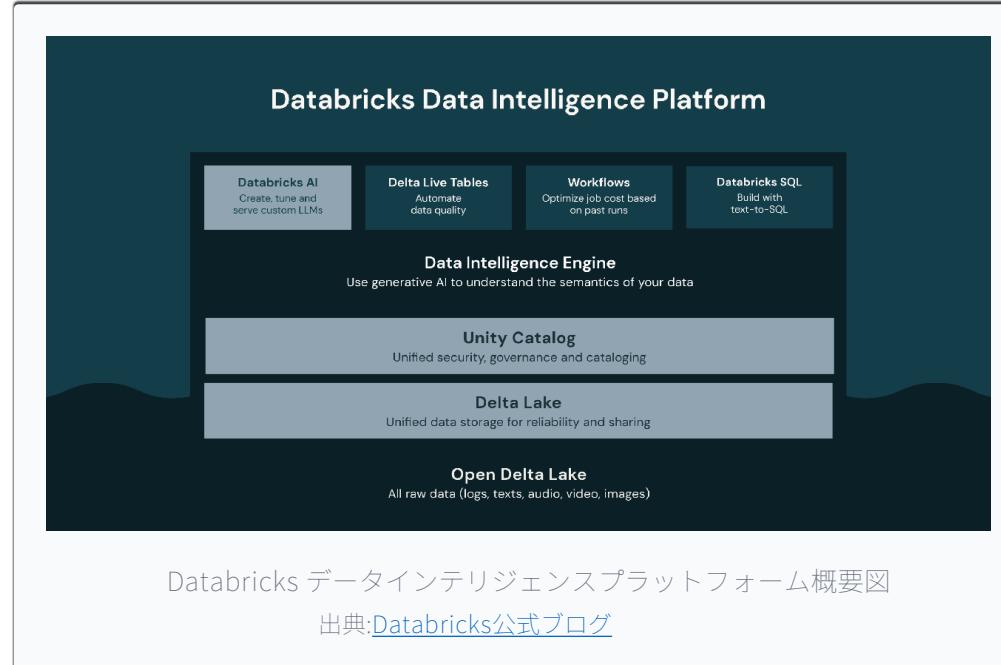


### レイクハウスからインテリジェンスプラットフォームへの進化

データレイクハウスからデータインテリジェンスプラットフォームへと発展した背景と最新機能



## Databricksとは



Databricksは、データレイクハウスの概念を発展させたデータインテリジェンスプラットフォームです。Apache Sparkの創設者たちによって2013年に設立され、現在はクラウドベースの統合データ分析・AI開発プラットフォームとして進化しています。

データとAIの統合：ビッグデータ処理、機械学習、BIを一つのプラットフォームで実現

オープンスタンダード：Delta LakeやApache Spark等のオープン技術を基盤として採用

マルチクラウド対応：AWS、Azure、GCPなど主要クラウドに対応

ユニファイドガバナンス：Unity Catalogによるデータ・AIアセットの統合ガバナンス



### 主な利用シナリオ

データエンジニアリング

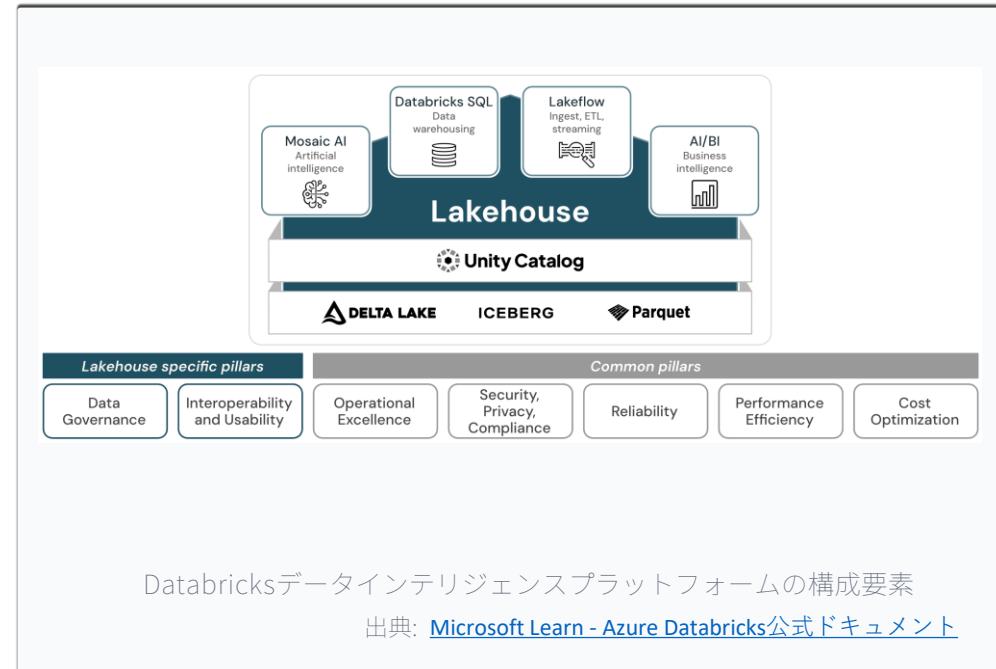
BIとアナリティクス

AI・機械学習

リアルタイム処理

詳細はAzure Databricks公式ドキュメントを参照

## データインテリジェンスプラットフォーム



### プラットフォームの特徴

💡 オープン性と互換性

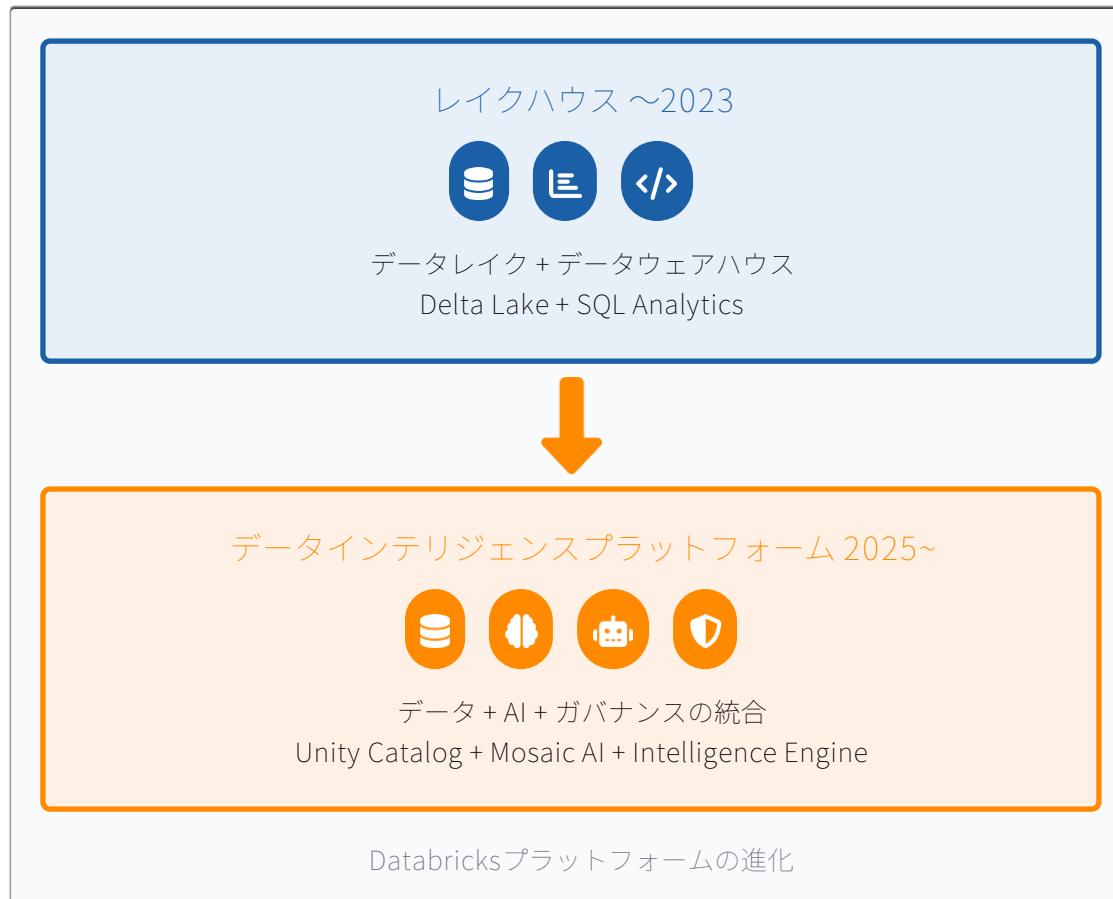
🛡️ セキュリティとガバナンス

⌚ パフォーマンス最適化

👤 全ユーザー向けインターフェース

🔗 詳細はAzure Databricks公式ドキュメントを参照

## 進化の背景と意義



Databricksは「レイクハウス」の概念を発展させ、現在は「データインテリジェンスプラットフォーム」として新たなステージに移行しています。この進化は、データと人工知能を統合し、真のビジネスインテリジェンスを実現するためのものです。

レイクハウスの限界：データ管理とAIが分離されており、ガバナンスが断片化していた

インテリジェンスへの進化：ユニファイドガバナンス、AIの組み込み、ビジネスコンテキストの強化

Unity Catalogの進化：データだけでなくAIアセットもカバーする統合管理

セマンティクスレイヤー：2025年新機能のメトリクス定義やAIドキュメント自動生成

### 主な進化ポイント (2025)

- ✓ ビジネスマトリクス統合
- ✓ Iceberg対応強化
- ✓ データ品質監視自動化
- ✓ AI生成ドキュメント

詳細は2025年のUnity Catalogアップデート情報を参照

## 第2部: 基本アーキテクチャ

Databricksのシステムアーキテクチャの基本要素と構成について解説します。

### 基本構成要素

システム全体像と分離設計

Control PlaneとCompute Plane

サーバレス vs クラシックコンピュート

ワークスペースとストレージ

ネットワークアーキテクチャ

マルチクラウド対応とリージョン戦略

スケーリングの仕組みと自動化

セキュリティアーキテクチャ

高可用性と災害復旧

パフォーマンス最適化

### アーキテクチャの特徴

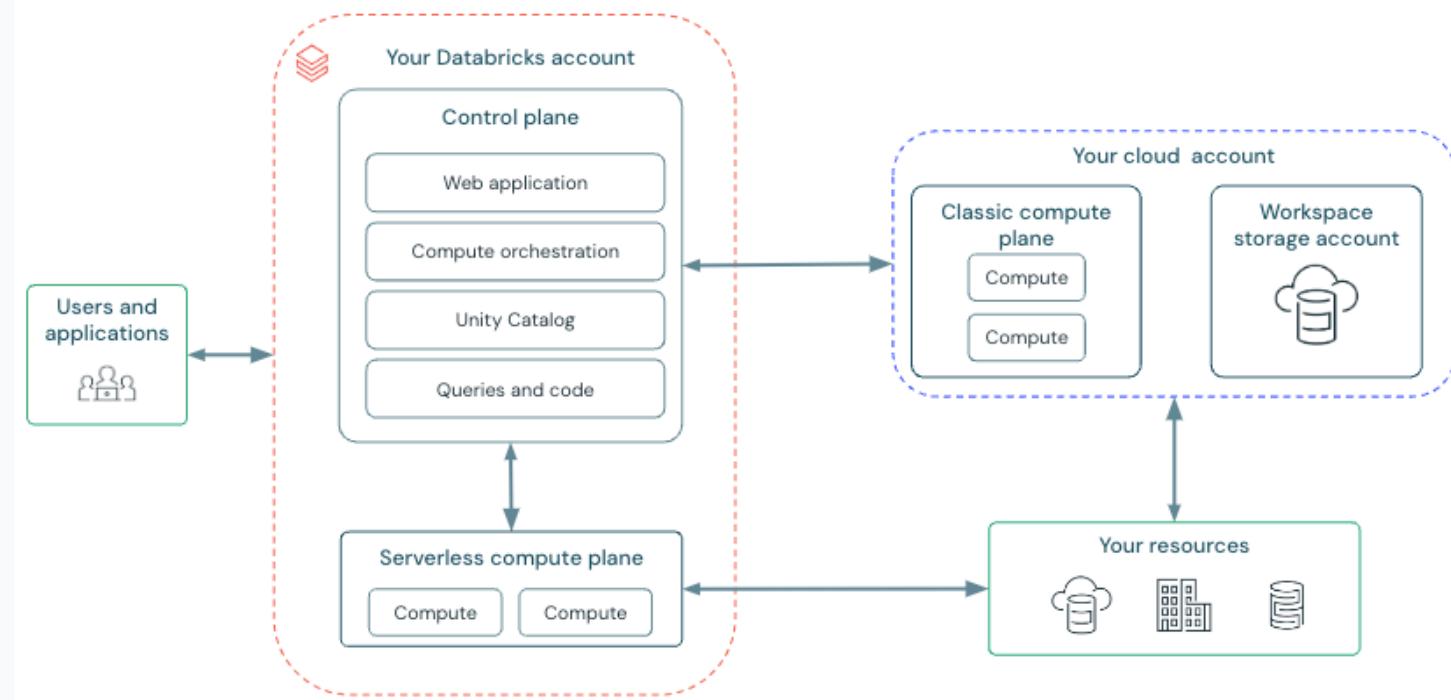
クラウドネイティブ設計 - クラウドの柔軟性とスケーラビリティを最大限に活用

マルチクラウド対応 - AWS、Azure、GCPで一貫したエクスペリエンス

エンタープライズグレードのセキュリティ - 細かなアクセス制御と暗号化



## Databricksアーキテクチャ全体像

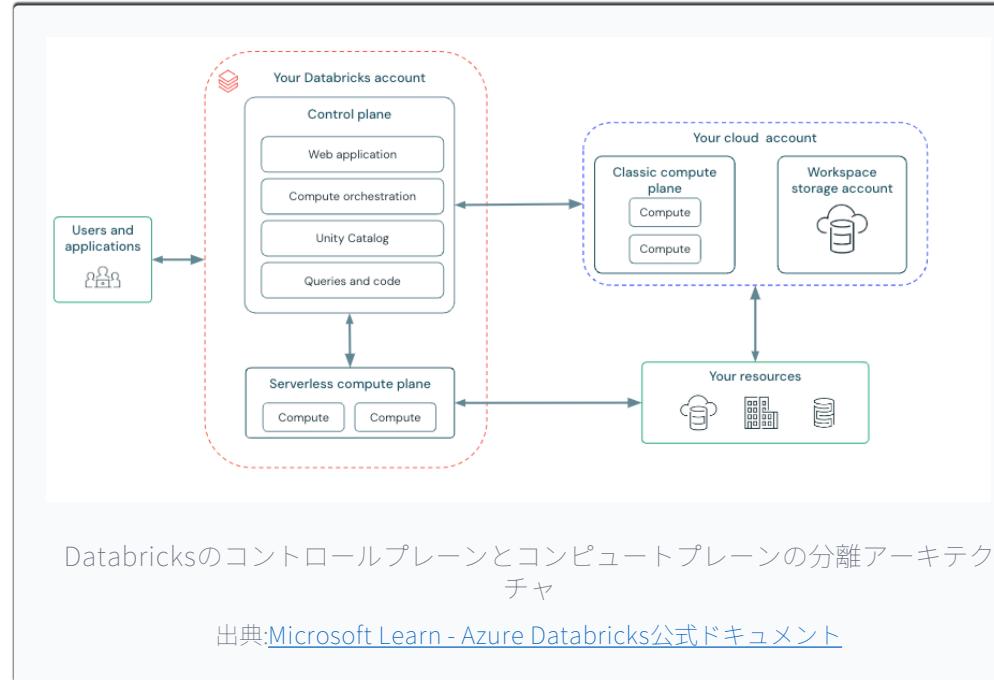


出典: [Microsoft Learn - Azure Databricks公式ドキュメント](#)



DatabricksのアーキテクチャはControl Plane（制御層）とCompute Plane（計算層）の2つに大別されます。この分離によりセキュリティと柔軟なスケーリングを両立しています。

## 2つのプレーンの分離



Databricksアーキテクチャは、管理機能を担う「コントロールプレーン」と実際の計算処理を行う「コンピュートプレーン」に論理的に分離されています。この分離により、セキュリティと拡張性の両立を実現しています。

コントロールプレーン：ワークスペース管理、認証、メタデータ、ウェブUI、ジョブスケジューリングなどの管理機能を提供するDatabricks管理領域

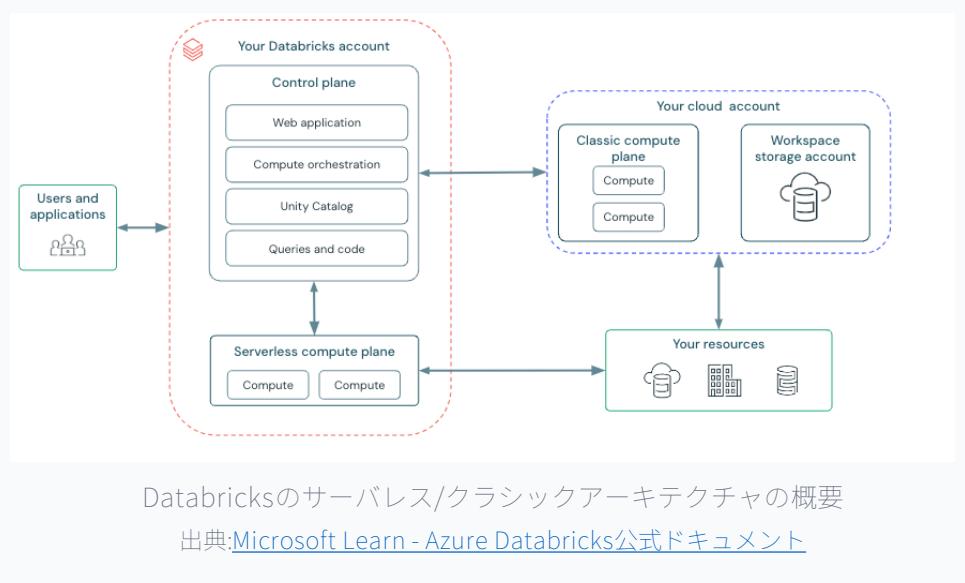
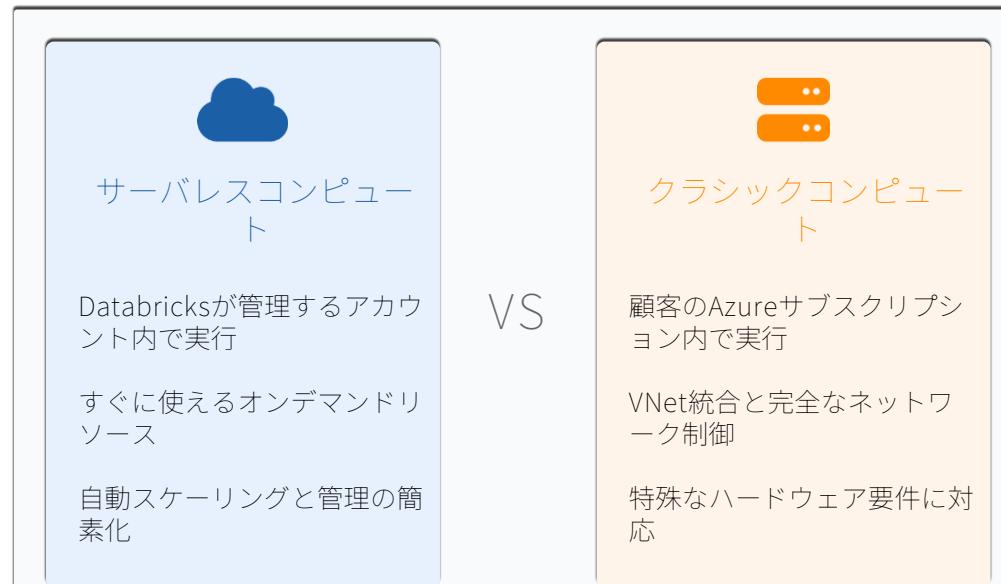
コンピュートプレーン：データ処理を実行するApache Sparkクラスタやサーバレスコンピュートが稼働する計算領域

ワークスペースストレージ：各ワークスペース専用のストレージアカウントでシステムデータやDBFSを保管

### 分離アーキテクチャの利点

- 🔒 セキュリティ強化：顧客ごとのコンピュートリソースの完全な分離
- ✖️ スケーラビリティ：必要に応じた計算リソースの独立した拡張
- ⌚ パフォーマンス：管理機能と計算処理の分離による最適化

詳細はAzure Databricksアーキテクチャを参照



## 2つのコンピュートモデル

Databricksでは、ワークロードの特性や要件に合わせて2種類のコンピュートモデルを選択できます。両者の違いと最適な使い分けを理解することで、コスト効率と性能を最適化できます。

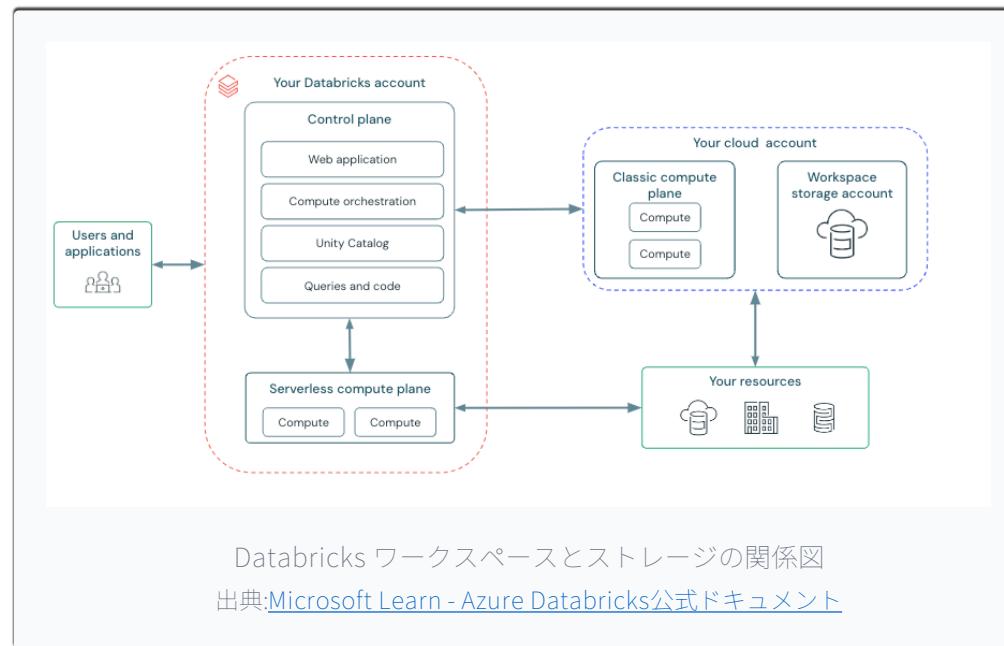
### 主な違いと特徴

- 實行環境：サーバレスはDatabricksアカウント内、クラシックは顧客のAzureサブスクリプション内
- ネットワーク：サーバレスは複数レイヤーのセキュリティ分離、クラシックはVNet統合
- 起動速度：サーバレスは即時起動、クラシックは構成によって起動時間が変動
- コスト最適化：サーバレスは秒単位の課金と自動スケーリングで効率的

### 使い分けの指針

- ⚡ 即時起動が必要で、コスト重視、素早い起動が必要、SQL分析、MLモデルサービス
- ⚙️ 特殊なネットワーク要件、オンプレミスリソースとの接続、GPUなど特殊ハードウェア利用時

## ワークスペースとストレージ管理



Azure Databricksでは、ワークスペースとストレージの関係が明確に定義されています。各ワークスペースには専用のストレージアカウントが関連付けられます。

ワークスペースストレージアカウント：ワークスペース作成時に自動的に作成され、システムデータ、DBFS、Unity Catalogワークスペースカタログを保存

DBFS (Databricks File System) : dbfs:/名前空間でアクセス可能な分散ファイルシステム

Unity Catalogとの連携：ワークスペースカタログはUnity Catalogで自動的に有効化され、外部ロケーションとストレージ認証情報によってバックアップ

ストレージアクセス制限：ファイアウォールサポートを有効にすることで、認可されたリソースとネットワークからのみアクセス可能



### ストレージアクセス方法

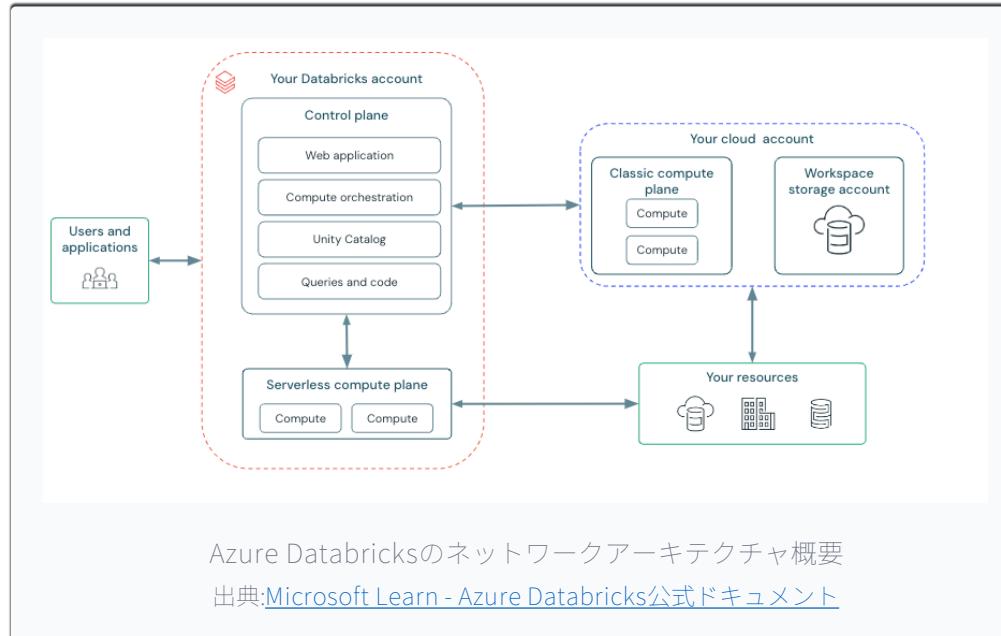
📁 DBFS Root (/dbfs)

🔗 DBFS マウント

📅 Unity Catalog ボリューム

☁️ 外部クラウドストレージ直接アクセス

## セキュアなネットワーク通信



Databricksは多層防御のセキュアなネットワークアーキテクチャを採用し、コントロールプレーンとコンピュートプレーン間の安全な分離を実現しています。

- PrivateLink連携：Azure PrivateLink対応でコントロールプレーンとの通信をプライベート化

- VNet注入：クラスタリソースを顧客VNetに配置し、既存のネットワークポリシーを適用

- Secure Cluster Connectivity：クラスタへのSSHアクセスを排除し、セキュアなWebSocketベースの通信を実現

- IPアクセスリスト：ユーザーインターフェイスやAPIへのアクセスを特定のIPアドレス範囲に制限

### 通信経路の暗号化

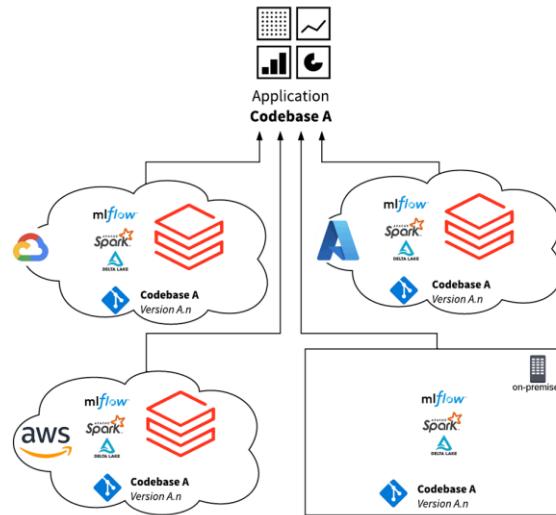
- ➡ 転送中データは常にTLS 1.2以上で暗号化

- ➡ 保存データは暗号化 (CMEK/BYOK対応)

- ➡ サーバレスコンピュートでは顧客間の完全なネットワーク分離

詳細はAzure Databricksネットワークセキュリティドキュメントを参照

## クラウド間の一貫性と柔軟性



Databricks マルチクラウドアーキテクチャ

画像出典:[Databricks 公式ブログ](#)

Databricksはマルチクラウド戦略をサポートし、AWS、Azure、GCPの主要クラウドプロバイダーで一貫した体験を提供します。これにより、ベンダーロックインを回避しながら最適なリソース配置が可能になります。

統一された体験：クラウドに関係なく同一のDatabricksインターフェースと機能

グローバルリージョン展開：30以上のリージョンで利用可能、コンプライアンス要件にも対応

クロスクラウドコラボレーション：Delta Sharingを活用したクラウド間でのセキュアなデータ共有

移行とポートабリティ：オープンフォーマット（Delta, Iceberg）採用によるクラウド間移行の容易さ

### リージョン選択の考慮点

⌚ データ主権要件

🛡 地域別コンプライアンス

⌚ レイテンシー要件

💰 コスト最適化

Azure Databricksサポート対象リージョン一覧





水平スケール



垂直スケール



ロードバランス

## スケーリングの仕組み

Databricksは柔軟で強力なスケーリング機能を提供し、ワークロードに応じて計算リソースを最適化します。ユーザーが意識することなく、必要に応じてリソースを拡張・縮小できます。

**自動スケーリング**：クラスターは負荷に応じてワーカーノードを自動的に追加/削除し、コストとパフォーマンスを最適化

**水平スケーリング（Scale Out）**：ワーカーノードの数を増減してワークロードを分散処理。ビッグデータ処理やSparkジョブに効果的

**垂直スケーリング（Scale Up）**：個々のノードのサイズ（CPU/メモリ）を拡張。メモリ集約型ワークロードに最適

**サーバレスコンピュート**：事前プロビジョニング不要で即時スケール。使用時のみ課金される効率的なオプション

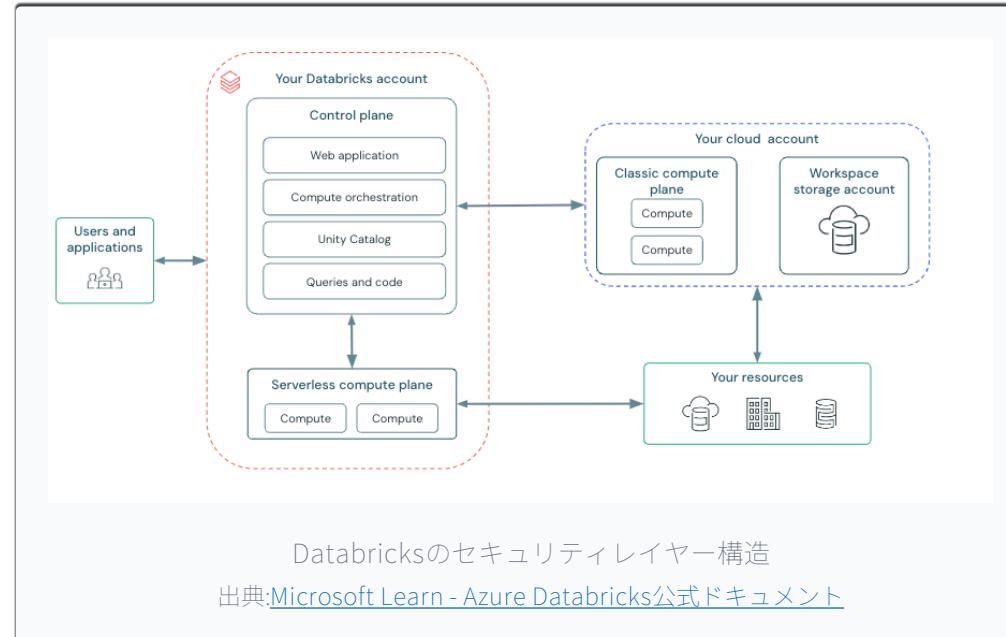
### スケーリングのベストプラクティス

💡 **自動終了**：一定の非アクティブ時間後にクラスターを自動終了させ、コスト削減

💡 **最小/最大ワーカー数**：ワークロードに応じた適切な設定で効率化

💡 **フォトンエンジン**：高速クエリパフォーマンスとコスト削減のための最適化

## セキュリティアーキテクチャ



Databricksは多層防御アプローチを採用し、クラウドネイティブなセキュリティアーキテクチャを実現しています。

### 主要セキュリティ対策

ネットワークセキュリティ：PrivateLink、IP制限、VNet統合によるネットワーク分離

アイデンティティ管理：SSO連携、SCIM、細かなIAM権限設定

データ保護：転送中・保存中の暗号化、顧客管理暗号化キー(CMEK)対応

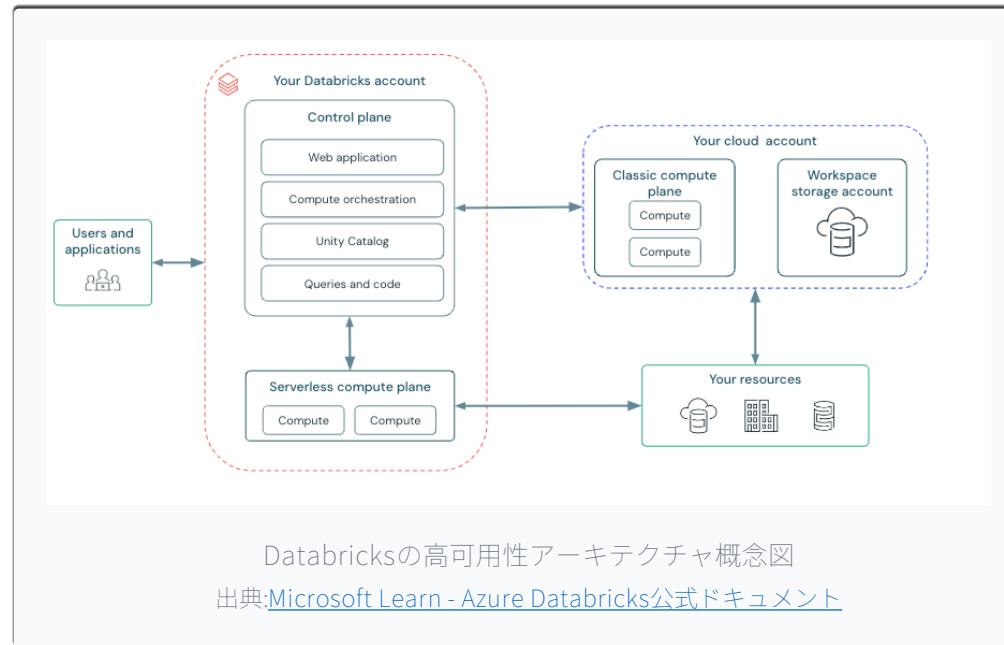
コンプライアンス：主要認証 (SOC 2 Type II、HIPAA、GDPR等) に準拠

### セキュリティ設計の特徴

- ✓ Control PlaneとCompute Planeの分離による強固なセキュリティ境界
- ✓ Unity Catalogによる統一されたデータガバナンスとアクセス制御
- ✓ サーバレスコンピュートによる顧客間の完全分離
- ✓ セキュリティ対策を自動監視するSecurity Analysis Tool (SAT)

[セキュリティベストプラクティスの詳細](#)

## Databricksの可用性と継続性



Databricksは複数のレベルでの高可用性を実現し、データとワークロードの継続性を確保するための包括的な戦略を提供しています。

コントロールプレーンの冗長性：複数のアベイラビリティゾーンにまたがる高可用性設計

コンピュートプレーンの自動回復：クラスタノード障害時の自動検出と再起動

メタデータレプリケーション：データベースメタデータの自動レプリケーションと保護

バックアップと復元：スナップショットによるDelta Lakeテーブルの時間指定復元

### 災害復旧戦略のベストプラクティス

- ✓ マルチリージョン展開による地理的冗長性の確保
- ✓ 定期的なデータバックアップとリテンション期間の設定
- ✓ 重要なワークフローの定期的な障害テスト実施
- ✓ Unity Catalogによるメタデータのクロスリージョンレプリケーション

詳細はAzure Databricks災害復旧ガイドを参照

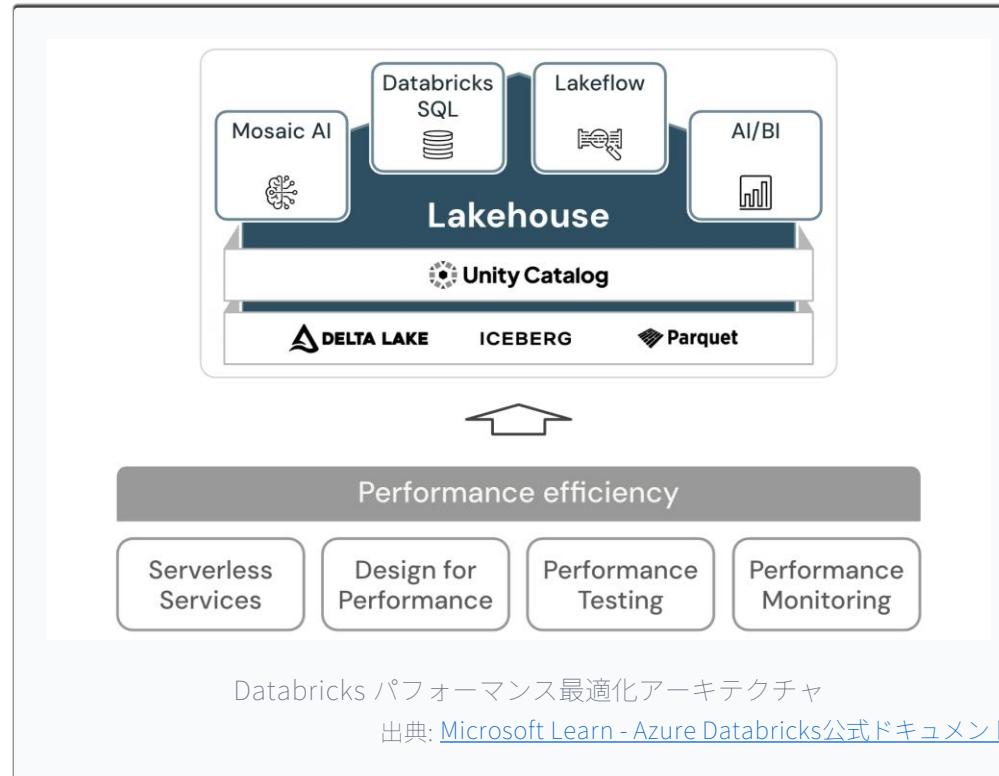


高可用性



災害復旧

## パフォーマンス最適化のポイント



Databricksプラットフォームでは、ワークフローのパフォーマンスを最大化するための複数の最適化技術が組み込まれています。費用対効果とパフォーマンスのバランスを考慮した設計が重要です。

Photonエンジン：C++ベースのネイティブエンジンによるクエリ実行の高速化。従来のSparkエンジンに比べ最大8倍の高速化を実現

液体クラスタリング (Liquid Clustering)：データの自動インデックス化と最適レイアウト生成によるI/O最適化

予測的最適化 (Predictive Optimization)：クエリパターンを学習し、自動的に最適なデータレイアウトを予測・適用

Delta最適化：Z-Order、コンパクション、データスキッピングによる処理効率化

### ベストプラクティス

🕒 クラスタ構成の最適化

\_PARTITION战略

⚡ SQLウェアハウスサイジング

⌚ キャッシュの有効活用

🔗 Azure Databricks Photonエンジン公式ドキュメント

# 第3部 Unity Catalog とは

Databricksのコア・ガバナンス基盤である  
Unity Catalogの概要と主要機能について解説  
します。

## 統合データ・AIガバナンス基盤

Unity Catalogは、データとAIアセットに対する統一的なガバナンス、検索、アクセス管理を提供する中央メタデータサービスです。複数のワークスペース、クラウド、ツールにまたがる統合管理を実現します。

### 主要機能

統一カタログによるメタデータ管理

きめ細かなアクセス制御 (RBAC/ABAC)

データリネージと監査機能

Apache Iceberg対応 (2025年新機能)

### 拡張機能と連携

メトリクス管理 (2025年新機能)

データ品質監視の強化

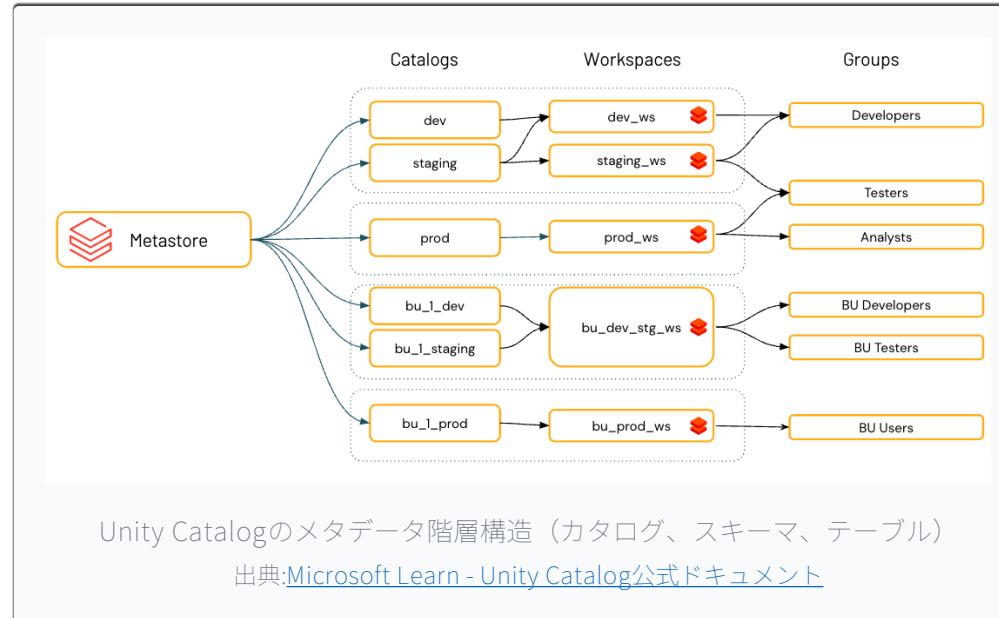
外部システム・カタログ連携

ビジネスユーザー向け拡張

詳細は公式ドキュメントをご参照ください： [Unity Catalog 公式ドキュメント](#)



# Unity Catalogによるメタデータ統合



Unity Catalogは、Databricksの中核となるメタデータ管理システムで、データとAI資産の統合力カタログを提供します。複数のワークスペース、クラウド、リージョンにまたがる一元的なデータガバナンスを実現します。

3層メタデータ階層：カタログ>スキーマ>テーブル/ビューという直感的な階層構造

多様なアセット対応：テーブル、ビュー、ファイル、ML機能、モデル、ダッシュボードなど

自動リネージ追跡：データの変遷を自動で記録し、影響分析を容易に

統合検索・探索：組織全体のデータ資産をキーワード、タグ、メタデータで検索



## 2025年の新機能

AIによる自動タグ付け

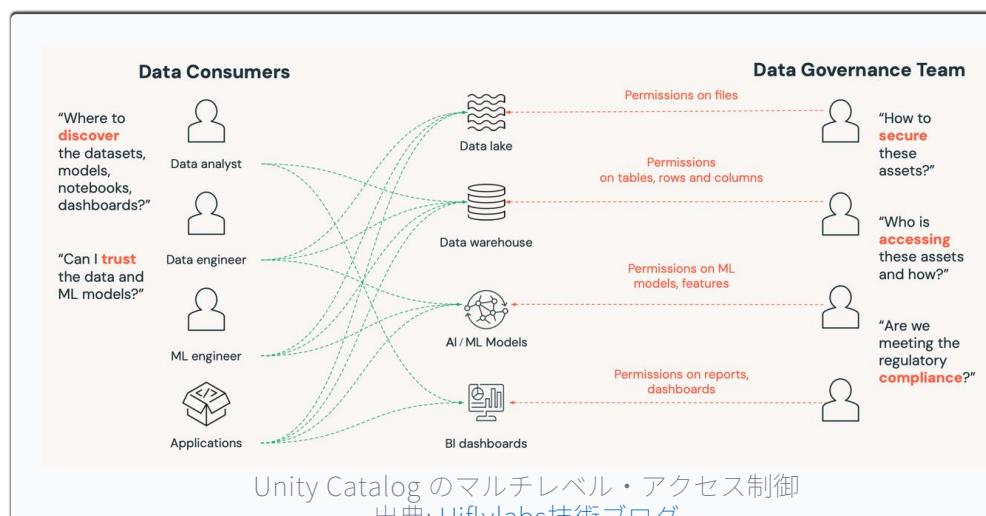
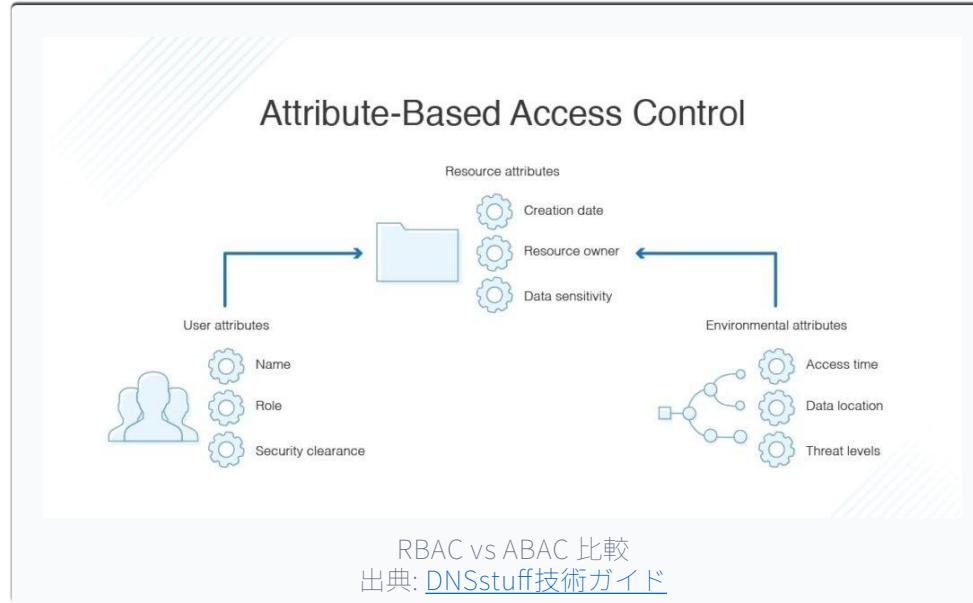
統一メトリクス定義

拡張リネージビュー

自然言語データ探索

Unity Catalogの詳細は[こちら](#)

## アクセス制御の基礎



Unity Catalogは、きめ細かなアクセス制御を提供し、組織のセキュリティニーズに合わせた柔軟なポリシー設定が可能です。従来のロールベースから、より高度な属性ベースのアクセス制御まで対応しています。

### RBAC (Role-Based Access Control)

役割に基づく制御：ユーザーの役割（管理者、分析者など）によってアクセス権限を付与

簡潔な管理：カタログ、スキーマ、テーブルレベルでの権限管理が可能

階層構造：継承可能な権限体系により管理を簡素化

### ABAC (Attribute-Based Access Control)

属性に基づく制御：データの属性、ユーザー属性、環境条件などに基づく動的なアクセス制御

行・列レベルのセキュリティ：特定の条件に基づきデータの一部のみを表示

データ分類連携：自動分類されたPIIデータへのアクセス制限を自動適用

### Unity Catalogの主なアクセス制御機能

💡 細粒度アクセス制御：カタログ・スキーマ・テーブル・列・行レベル

💡 IAM/AAD連携：クラウドIDとの統合認証

# データリネージと監査機能

Unity Catalogによるカラムレベルのデータリネージの可視化  
出典: [Monte Carlo Data技術ブログ](#)

Unity Catalogは、データの起源から変換、使用まで追跡する強力なリネージ（系統）機能と、包括的な監査機能を提供します。これによりデータガバナンスとコンプライアンスが大幅に強化されます。

自動リネージ収集：SQLクエリ実行時に自動的にテーブル間の関係を記録し、列レベルの詳細な系統情報を生成

視覚的リネージグラフ：上流・下流の依存関係を直感的なグラフで表示、インパクト分析を容易に実現

監査ログ：すべてのデータアクセス、メタデータ変更、権限変更を包括的に記録

変更履歴追跡：データやスキーマの変更を追跡し、いつ誰が何を変更したかを明確に把握



## 主なユースケース

問題の根本原因調査

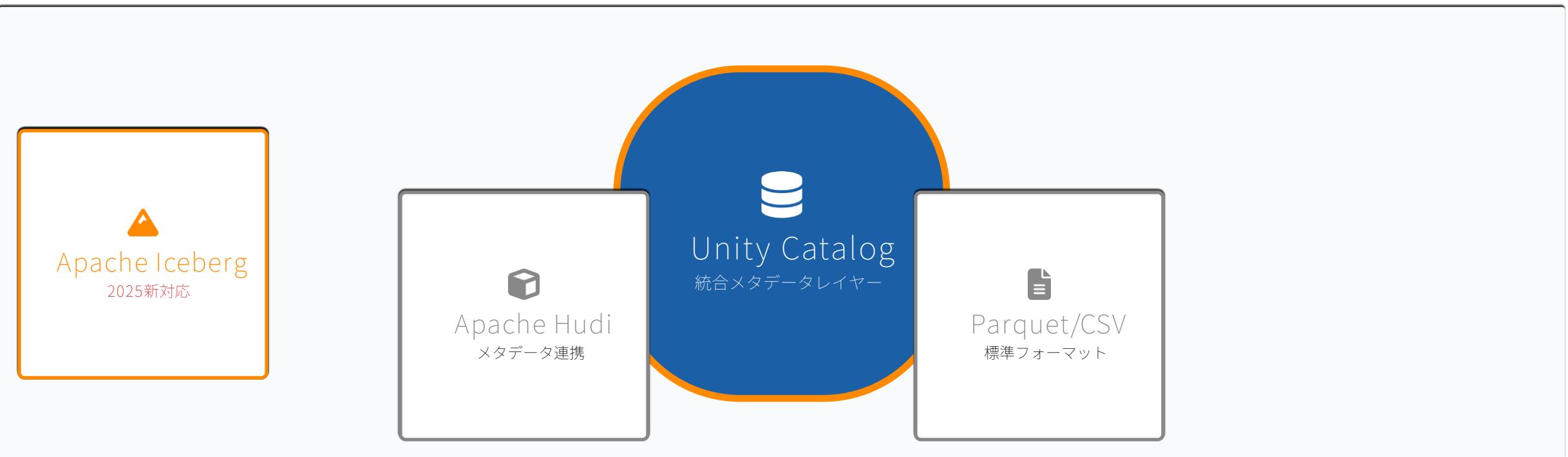
コンプライアンス証明

監査報告書作成

変更影響分析

詳細はAzure Databricks Unity Catalogのデータリネージ公式ドキュメントを参照

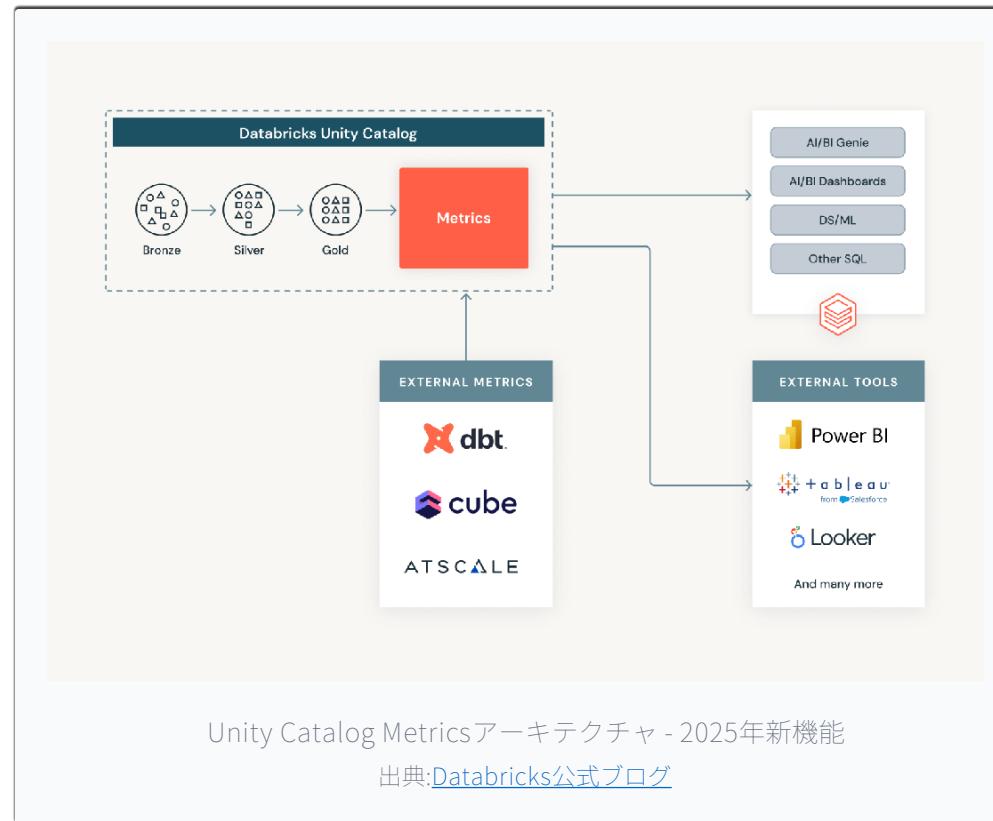
## Iceberg対応とオープンフォーマット連携



2025年新機能：Unity CatalogはApache Iceberg REST Catalog APIを完全サポートし、外部エンジンからの読み書きを実現。これによりフォーマットロックインを排除し、Trino、Snowflake、Amazon EMRなど様々な外部システムとの相互運用性を提供します。

- ✓ Icebergマネージドテーブル
- ✓ Icebergカタログフェデレーション
- ✓ Delta SharingからのIceberg対応
- ✓ オープン標準を活用した相互運用性

# Unity Catalog Metrics



Unity Catalog Metricsは、2025年の主要な新機能で、ビジネスメトリクスをデータレイヤーで定義し、すべてのデータ・AIワークフローで一貫して再利用できるようにします。

一度定義、どこでも利用：メトリクスをUnity Catalogで一元定義し、AI/BIダッシュボード、Genie、ノートブック、SQL、Lakeflowジョブなど全ての場所で利用可能

一貫したセマンティクス：異なるツールやチーム間でビジネスKPIの定義を標準化し、全社で一貫した指標を実現

ガバナンスとセキュリティ：認証済みメトリクスには監査とリネージ機能が標準搭載され、コンプライアンスとデータ信頼性を確保

パートナーエコシステム統合：TableauやPower BI等の主要BIツール、Anomaloなど監視ツールとの統合予定

## ビジネスユーザーへの価値

「Unity Catalog Metricsは、ビジネスKPIを中心で定義し、チーム間でセマンティクスを標準化することで、全員が同じ信頼できる定義をダッシュボード、SQL、AIアプリケーション全体で使用できるようになります。」

- Richard Masters氏 (Virgin Atlantic社、データ&AI担当副社長)

[Azure Databricksメトリックビュー公式ドキュメントを参照](#)



# データ品質監視の強化



Unity Catalogを基盤とした自動データ品質監視システムにより、大規模データ環境でも一貫した品質管理が可能になります。2025年の最新機能では、異常検知とアラートが大幅に強化されています。

データ鮮度の監視：テーブルが最後に更新されたタイミングをモニタリングし、データの古さをアラート通知

データ完全性チェック：予測されるデータ量と実際のデータ量の比較を自動実施

異常検知：AIを活用した統計的異常検出により、データパターン変化をプロアクティブに検出

影響分析：リネージュを活用した下流への影響評価で重要度を自動判定



## 主な適用シナリオ

IoT/センサーデータ検証

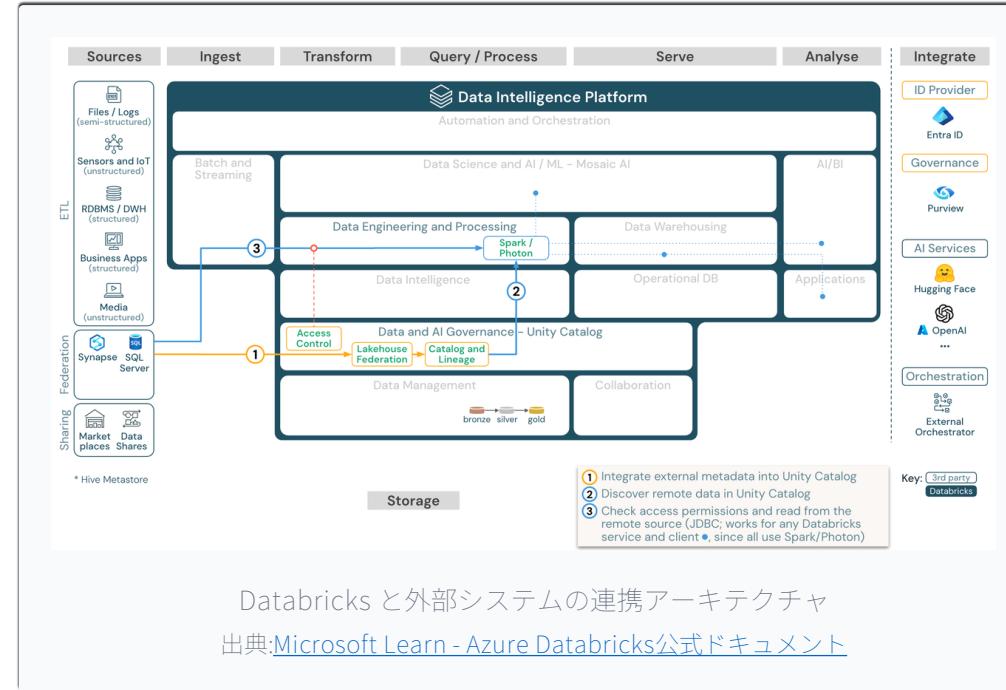
顧客データ品質確保

金融取引データ整合性

AIモデル入力データ検証

詳細はAzure Databricks公式ドキュメントのデータ品質監視ガイドを参照

## 外部システム連携



Unity Catalogを中心としたDatabricksの外部システム連携機能により、既存のデータソースとシームレスに統合し、単一の管理ポイントからアクセス制御を実現できます。

Lakehouse Federation：外部RDBMSや他のSQLデータソース（MySQL、Postgres、SQL Server、Synapse等）をETLなしでUnity Catalogに統合

カタログフェデレーション：Hiveメタストアカタログを既存の形式のままUnity Catalogに連携し、統合ガバナンスを適用

Iceberg Catalog連携（2025年新機能）：AWS Glue、Hiveメタストア、Snowflake HorizonのIcebergテーブルにシームレスにアクセス

Delta Sharing：組織間のデータ共有をオープンプロトコルで実現、クラウドやリージョンを越えた連携が可能

### 連携システム例

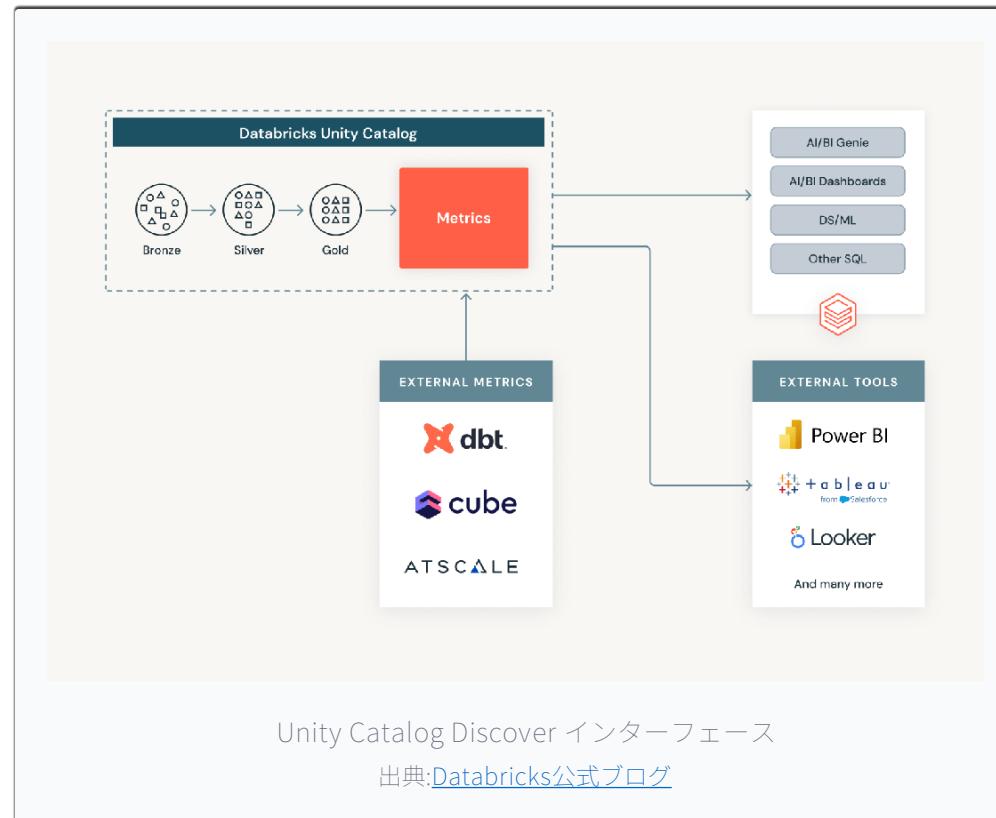
- SQL Server
- Snowflake
- Synapse Analytics

- PostgreSQL
- AWS Glue
- Hiveメタストア

### 連携のメリット

- データ移行不要で既存資産を活用しながら統合ガバナンスを実現
- 单一インターフェースからすべてのデータソースにアクセス可能
- きめ細かなアクセス制御と監査ログを外部データにも適用

詳細はLakehouse Federation公式ドキュメントを参照



## ビジネスユーザー向け機能強化

2025年のDatabricksは、技術者だけでなくビジネスユーザーにもフォーカスした機能拡張を提供しています。データプラットフォームを企業全体で活用するための新機能を紹介します。

**Unity Catalog Discover機能 (Private Preview)** : ビジネスドメイン別に整理されたキュレーションされた内部マーケットプレイス。データ資産を部門別（営業・マーケティング・財務など）に発見しやすい形で提示します。

**Unity Catalog Metrics (Public Preview)** : ビジネスマトリクスを一元管理する新機能。ダッシュボード、SQL、AI間で一貫した指標定義を実現し、指標の混乱を解消します。

**認証・廃止フラグ表示** : データの信頼性を視覚的に表示し、品質の高いデータセットを瞬時に識別可能にします。

**Databricksアシスタント** : 自然言語で質問すると、統制されたデータに基づいて回答する対話型AIツール。

### ビジネス価値

- ⌚ 意思決定の迅速化
- ⌚ データ信頼性向上

- 🤝 部門間データ連携促進
- 👤 セルフサービス分析

インテリジェント検索  
セマンティック検索体験

AIアシスタント  
自然言語での質問応答

メトリクス管理  
統一された指標定義

ドメイン整理  
ビジネス領域別の整理

## 第4部 データアーキテクチャとパイプライン

モダンデータパイプライン全体像と構成の流れを解説します。メダリオンアーキテクチャを中心に、効率的なデータ変換と管理手法を学びましょう。

### メダリオンアーキテクチャ

メダリオンアーキテクチャの基本概念

Bronze層：生データの取り込み

Silver層：データクレンジングと統合

Gold層：ビジネス準備済みデータ

Delta Lakeの特徴と利点

トランザクション保証

スキーマ進化

タイムトラベル（履歴管理）

クエリ最適化

Parquet形式とメタデータ管理

### データ処理パイプライン

ストリーミングデータ処理

Structured Streamingの活用

バッチ処理パイプライン

Auto Loaderによるデータ取り込み

多様なデータフォーマット対応

Delta, Iceberg, Hudiの互換性

データ共有とDatabricks Marketplace

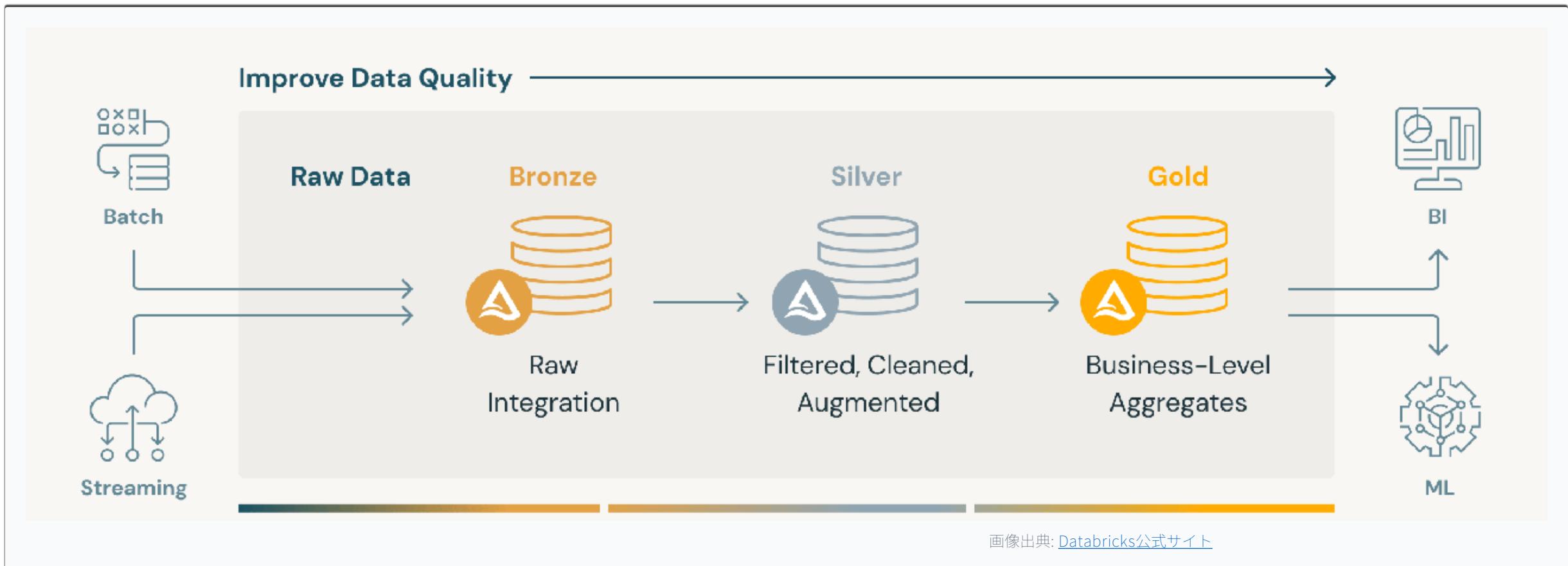
Delta Sharingの活用事例

このセクションでは、Databricksが推奨するメダリオンアーキテクチャの実装方法と、そのメリットを詳しく解説します。特に、データの品質向上プロセスと各層の役割に焦点を当てています。

詳細は[公式ドキュメント](#)をご参照ください。



## メダリオンアーキテクチャとは



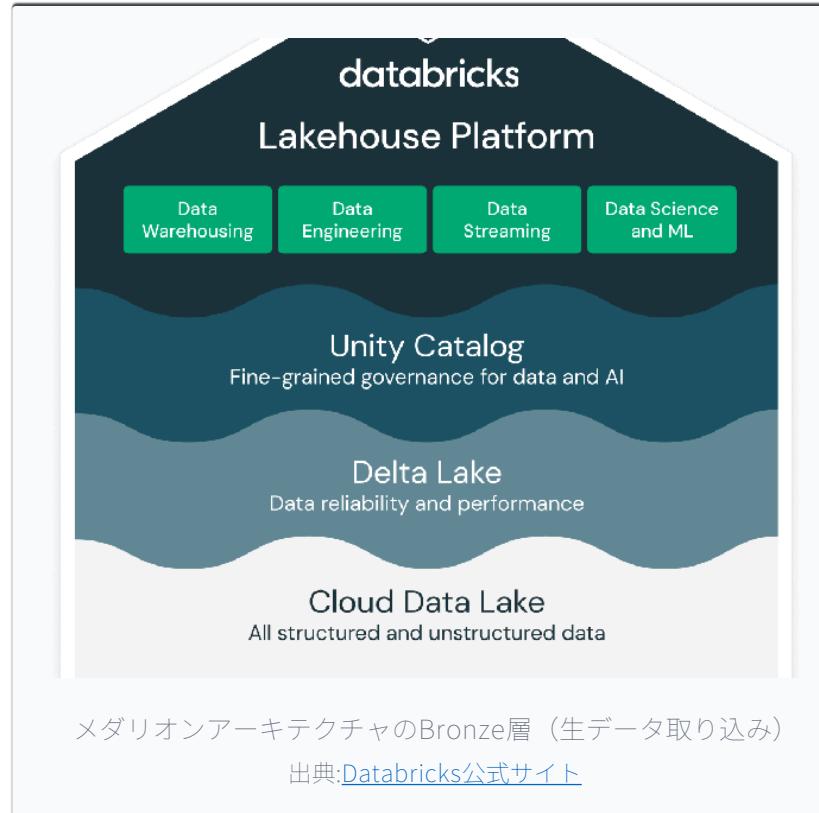
メダリオンアーキテクチャはBronze（生データ）、Silver（検証済みデータ）、Gold（ビジネス準備済みデータ）の3層でデータを段階的に洗練していくデータレイクハウスの基本設計パターンです。

Bronze: 生データをそのまま保存

Silver: クレンジング・変換・結合

Gold: ビジネスマトリクス・集計

全層でDeltaトランザクション保証



## Bronze層：生データレイヤー

Bronze層はメダリオンアーキテクチャの最初のステージで、外部ソースからの生データをそのまま取り込み保存する領域です。これは「Single Source of Truth」として機能し、後続の処理のための基盤となります。

取り込みデータの種類：ログデータ、IoTセンサーデータ、CRMシステム、ERPシステム、トランザクションデータなど

データフォーマット：JSON、CSV、XML、Avro、Parquet、画像、動画など様々な形式

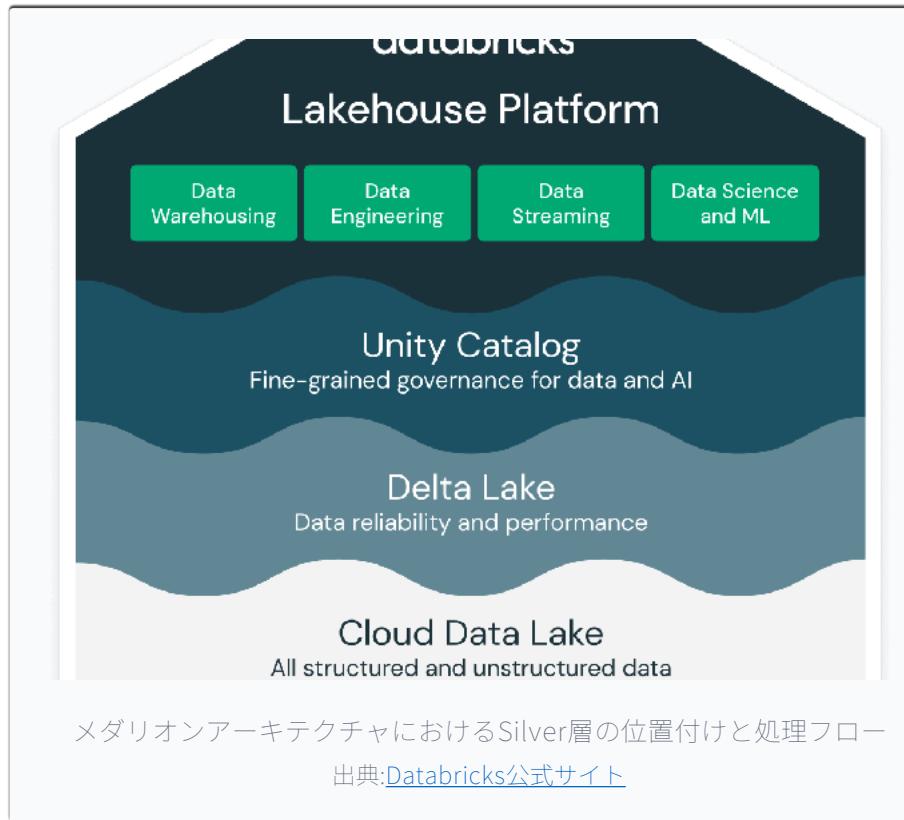
取り込み手法：Auto Loader、Structured Streaming、Apache Kafka連携、JDBC/ODBC接続など

スキーマ適用：最小限のスキーマ強制または推論スキーマの適用

### Bronze層のベストプラクティス

- ✓ データはそのまま取り込み、変換は最小限に抑える
- ✓ すべてのソースデータに取り込み時刻と出所を記録
- ✓ Delta Lake形式でデータを保存し、ACIDトランザクション保証
- ✓ 効率的なパーティショニング戦略を適用する

詳細はAzure Databricksメダリオンアーキテクチャのドキュメントを参照



## Silver層の役割と処理

Silver層はメダリオンアーキテクチャの中間層で、Bronze層から取り込まれた生データを検証・クレンジング・統合してビジネス利用可能な形に変換します。品質と一貫性が確保された信頼性の高いデータセットを作成します。

### 主なデータ処理操作

データクレンジング：欠損値の処理、外れ値の検出と補正、重複データの削除

スキーマ適用：データ型の強制、スキーマ進化の管理

データ統合：複数のソースから取得したデータの結合・統合

データ標準化：名称や単位の標準化、コードの統一

データ検証：ビジネスルールに基づくデータ検証と品質チェック

### Silver層の実装パターン

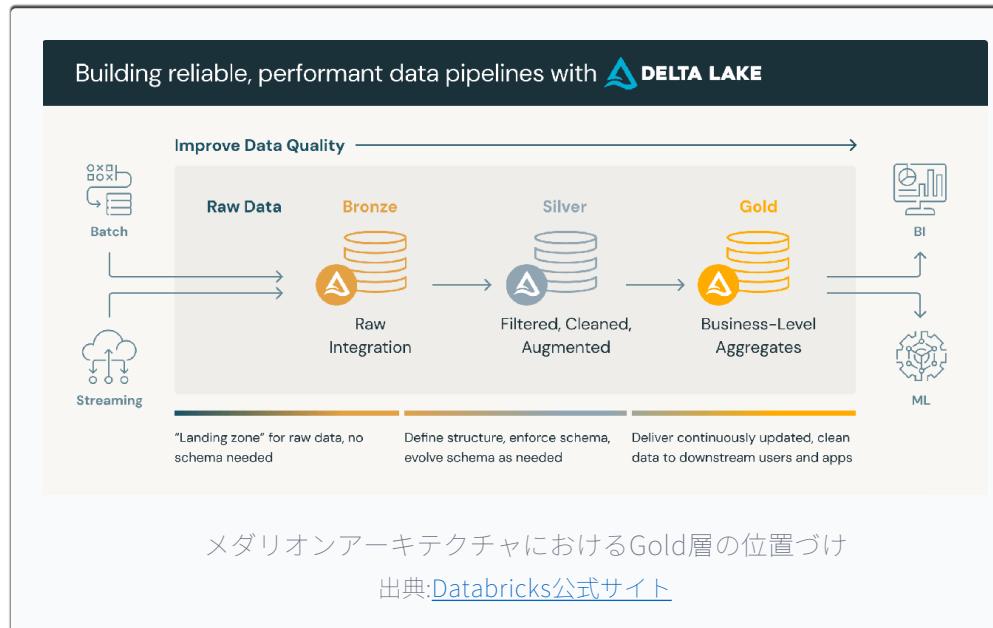
#### △ Delta Lake テーブル + データ検証

Delta Lakeのトランザクション機能を活用し、データ品質チェック（expectations）を適用

#### ⟳ ストリーミング + バッチハイブリッド処理

リアルタイムデータとバッチデータの統合パイプラインの構築

## Gold層：ビジネス価値の創出



Gold層はメダリオンアーキテクチャの最終層であり、ビジネスユーザーが直接アクセスして利用する高度に集約・加工されたデータを提供します。Silver層のクリーニング済みデータをさらに価値化した状態です。

ビジネス指向の集計データ：部門やチーム別KPI、ダッシュボード用集計値

特徴量テーブル：ML・AIモデル用に最適化された特徴量

ディメンションナルモデル：ファクトテーブルとディメンションテーブル

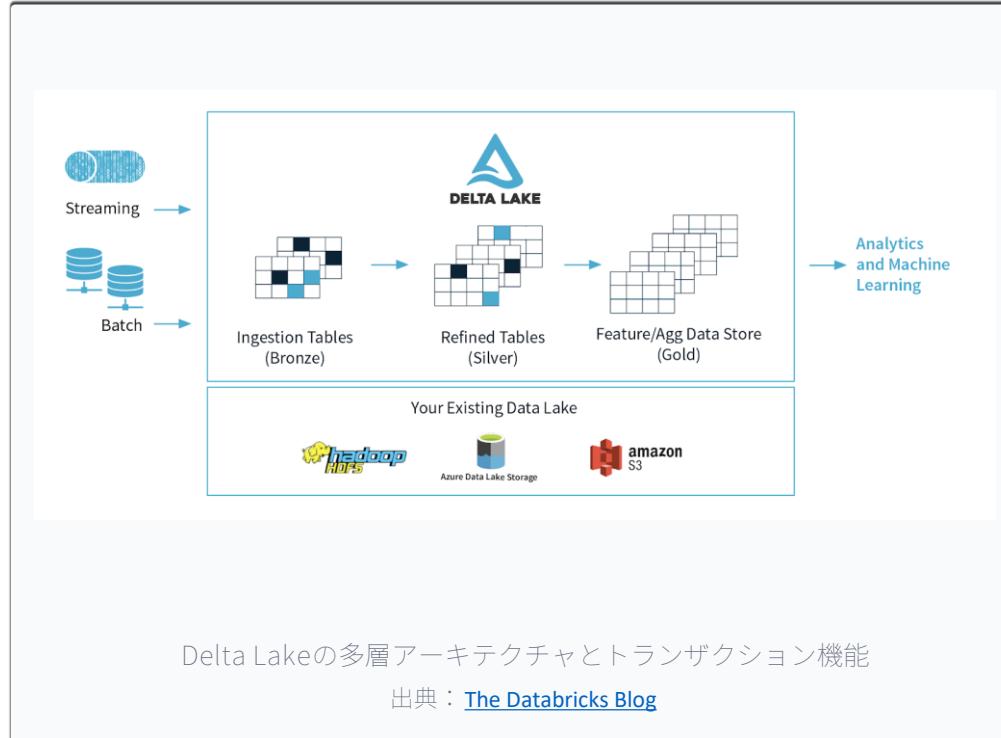
データプロダクト：再利用可能な分析データセットやビュー

### Gold層のベストプラクティス

- ✓ ビジネス要件に基づく設計
- ✓ アクセス制御の最適化
- ✓ データカタログとの統合
- ✓ データ品質メトリクスの監視

詳細はAzure Databricksメダリオンアーキテクチャのドキュメントを参照

## Delta Lakeの特徴と利点



Delta Lakeは、データレイクにデータウェアハウスの信頼性と性能を提供するオープンソースストレージレイヤーです。大規模なデータセットに対する高信頼性のETL、分析、AIワークフローをサポートします。

ACIDトランザクション：複数の同時書き込みを安全に処理し、データの整合性を保証

スキーマ強制と進化：データ品質の確保とスキーマの柔軟な変更をサポート

タイムトラベル（バージョニング）：過去のデータスナップショットに簡単にアクセス可能

ストリーミングデータ統合：Structured Streamingとのシームレスな連携

最適化エンジン：Zオーダー、自動インデックス、データスキッピングによる高速クエリ



### 一般的なユースケース

⟳ バッチ・ストリーミング統合

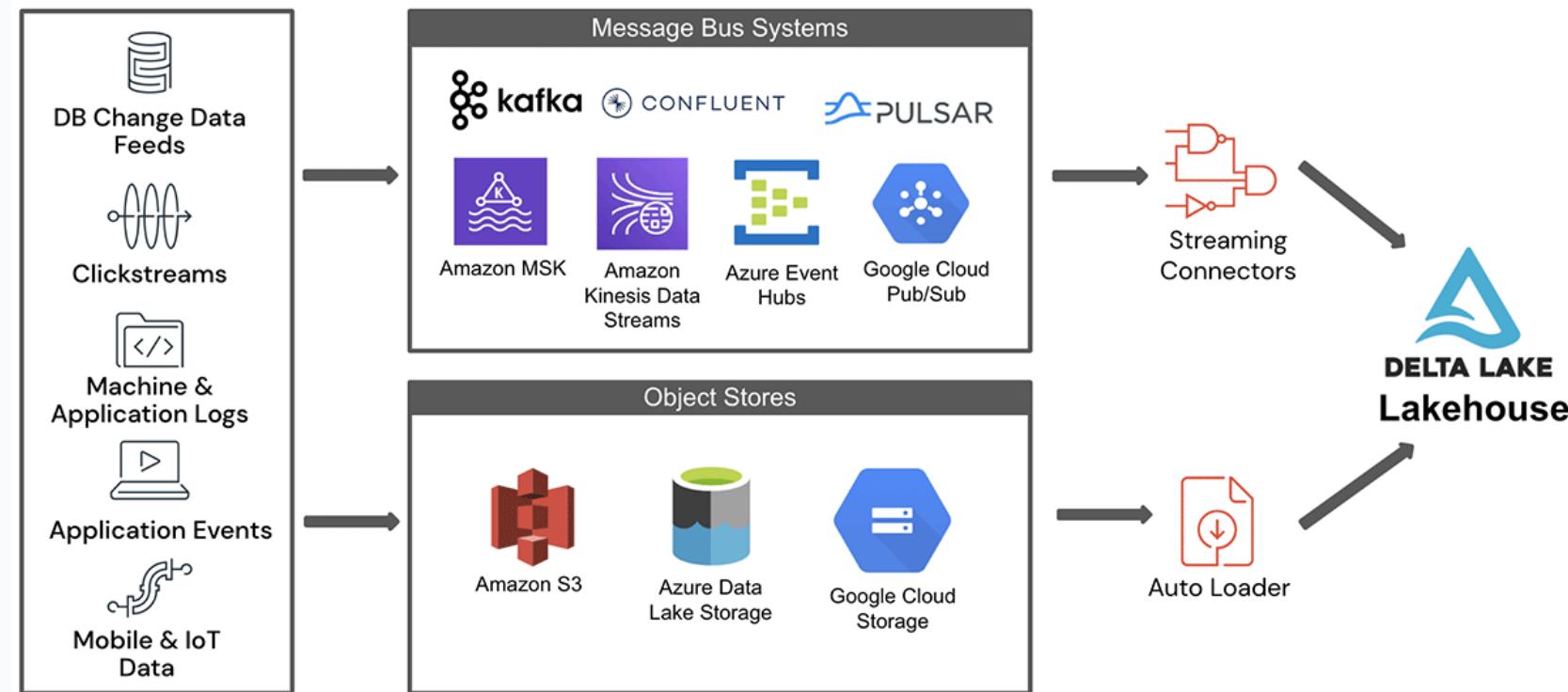
⟳ データバージョニング

heartbeat icon データ品質管理

disk icon ETL/ELTパイプライン

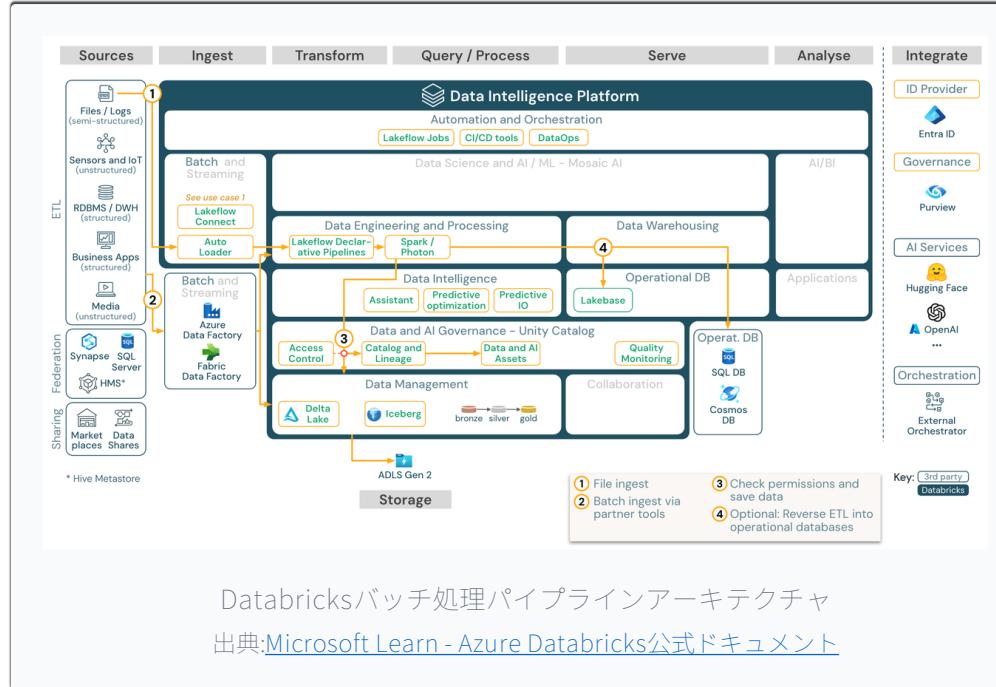
🔗 詳細はAzure Databricks Delta Lake公式ドキュメントを参照

## ストリーミングデータ処理

出典: [Databricks公式ブログ](#)

DatabricksではStructured StreamingとDelta Lakeを組み合わせることで、高速かつ信頼性の高いストリーミングデータ処理パイプラインを実現しています。

## バッチ処理パイプライン



Databricksプラットフォームでは、大規模データの定期的なバッチ処理を効率的に構築・運用できます。Delta Lake上に構築されたバッチ処理パイプラインは、信頼性と拡張性を兼ね備えています。

自動化ワークフロー：Lakeflow Jobsを使用した依存関係を持つマルチタスクワークフロー構築

寡等性と耐障害性：トランザクション管理とチェックポイントによる安全な再処理

スケジュール実行：時間/イベントベースのトリガーで定期的なデータ処理を実現

パフォーマンス最適化：Z-Order、データスキッピング、コンパクションによる処理効率向上



### 実装パターン

⌚ メダリオン各層間ETL

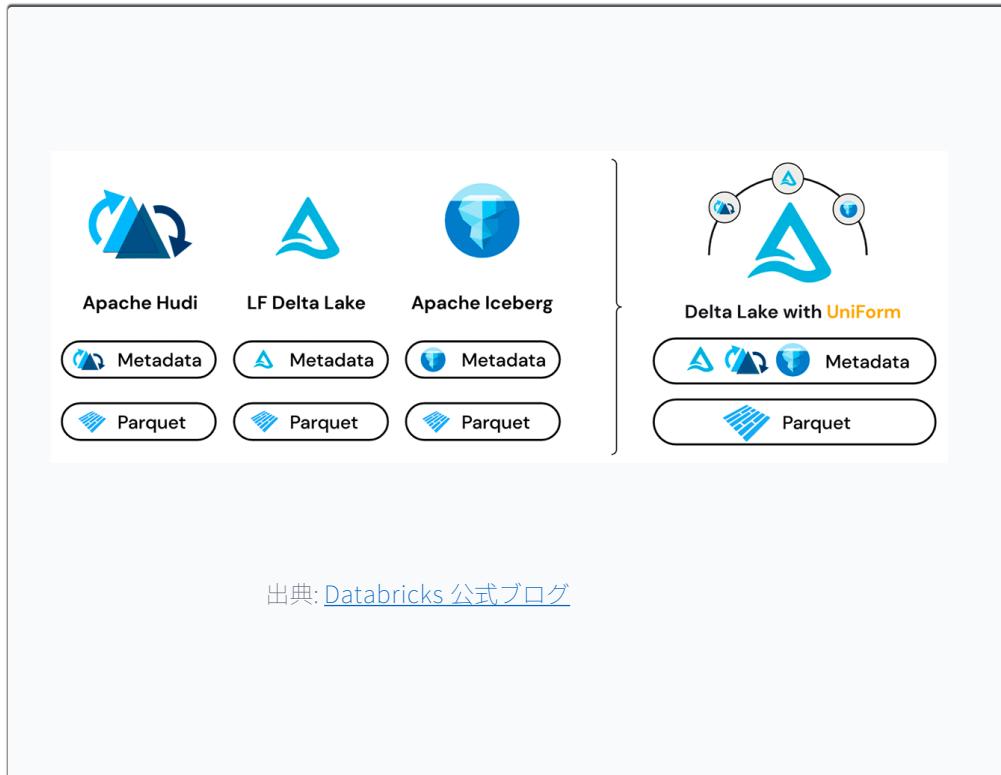
⌚ 増分データロード

⌚ 定期的データ集計

⌚ 日次/月次レポート生成

🔗 詳細はDatabricksジョブ公式ドキュメントを参照

## オープンデータフォーマット対応



Databricks はオープンでベンダーロックインのないアーキテクチャを採用しており、複数のオープンデータフォーマットをネイティブでサポートしています。これにより、既存の投資を保護しながら、様々なエンジンやツールとの相互運用性を実現しています。

**Delta Lake** : Databricks開発のオープンフォーマット。ACID トランザクション、スキーマ強制、タイムトラベルをサポート

**Apache Iceberg** : 2025年に強化されたネイティブサポート。大規模テーブル管理のためのオープンフォーマット

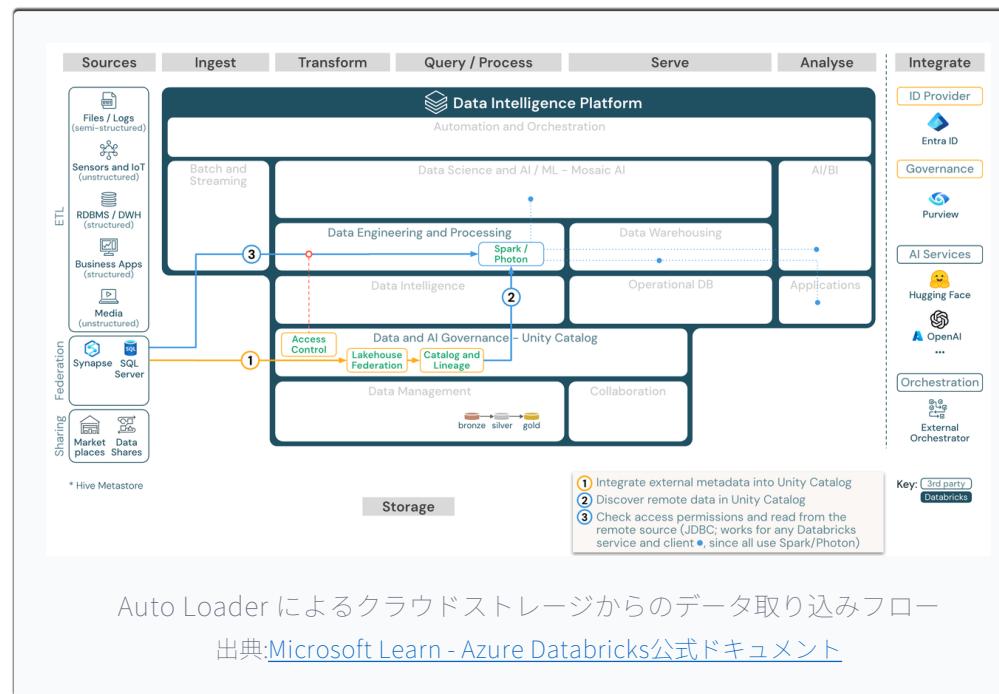
**UniForm** : Delta Lake上のデータをIceberg/Hudiとして読み書き可能にする互換レイヤー

### フォーマット選択のポイント

- ✓ 既存環境との統合：使用中のツール・システムとの互換性を確認
- ✓ パフォーマンス要件：ユースケースに最適な読み書き性能を持つフォーマットを選択
- ✓ Unity Catalogによる統合管理：どのフォーマットも単一のカタログで管理可能

# Auto Loaderとは

Auto Loaderは、クラウドストレージに到着する新しいデータファイルを自動的に検出し、効率的に取り込むDatabricksの機能です。増分処理により、大規模データ取り込みをシンプルかつスケーラブルに実現します。



シンプルなAPI : spark.readStream.format("cloudFiles")の簡潔な構文でストリーミング読み込み

インフラコスト削減 : 通知ベースのアーキテクチャによりリストコスト最適化

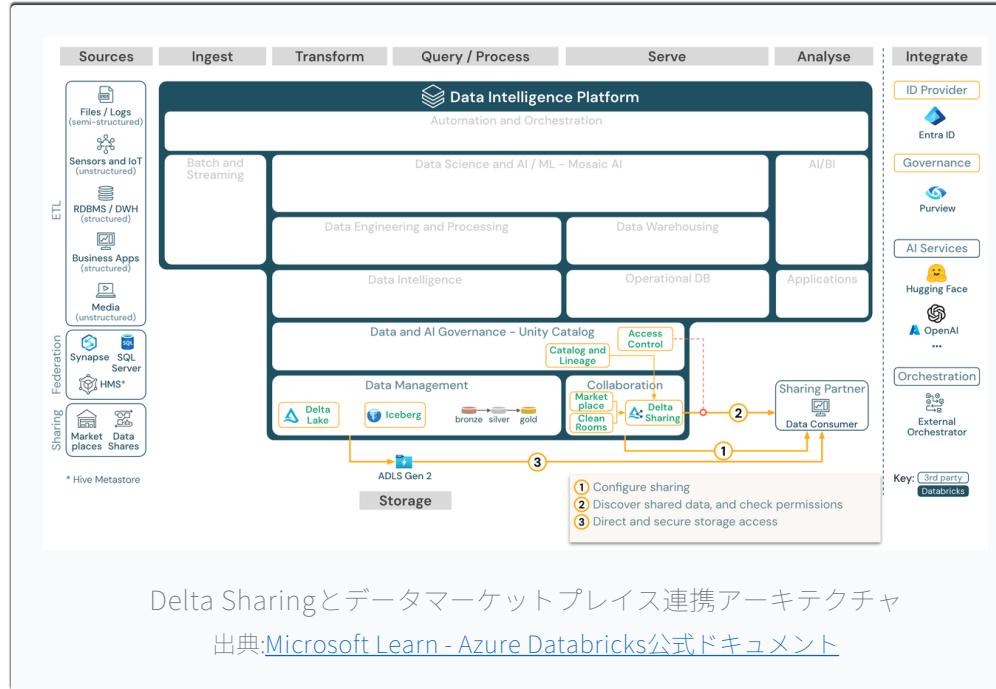
柔軟なスキーマ対応 : スキーマ進化 (evolution) 対応と推論機能

高スループット : 大量のファイルを効率的に処理、並列読み取りによる最適化

## 実装例

```
spark.readStream
  .format("cloudFiles")
  .option("cloudFiles.format", "json")
  .option("cloudFiles.schemaLocation", "/tmp/schema")
  .load("/path/to/input")
  .writeStream
  .trigger(once=True)
  .toTable("silver.customers")
```

## データ共有とマーケットプレイス



Databricksはオープン標準のDelta Sharingプロトコルとマーケットプレイスにより、安全で効率的なデータ共有エコシステムを提供します。これにより、組織間のデータコラボレーションを大幅に簡素化できます。

Delta Sharing：クラウド、プラットフォーム、ツールの境界を超えたオープン標準のデータ共有プロトコル

マルチフォーマット対応：Delta、Iceberg、Parquetなど多様なフォーマット間での相互運用性

セキュアな共有：Unity Catalogによる統合アクセス制御と監査

Databricks Marketplace：商用・公共データ製品のオープンな交換基盤

### 主な活用シナリオ



部門間データ共有

パートナー連携

Clean Rooms

データマネタイズ

詳細はAzure Databricks Delta Sharing公式ドキュメントを参照

# 第5部 AI・機械学習プラットフォーム

Databricksのデータを価値に変えるAI・機械学習プラットフォームの全体像を解説します。

## Mosaic AIとは

Mosaic AIの概要と特徴

MLOpsワークフロー

モデル開発からデプロイメントまで

Feature Storeの活用

モデルサービスとガバナンス

生成AI・AIゲートウェイ新機能

AutoMLと実験管理

## AI・機械学習の主要コンポーネント



### MLflow

実験追跡、モデル管理、デプロイを統合的に管理するためのプラットフォーム



### Feature Store

特徴量の一元管理と再利用を可能にするリポジトリ



### Model Serving

モデルを本番環境にデプロイし、高性能なAPIとして提供

詳細は[Azure Databricks AI/ML公式ドキュメント](#)をご参照ください



ARTIFICIAL INTELLIGENCE

## Build and deploy quality AI agent systems

Securely connect your data with any AI model to create accurate, domain-specific applications.

[See AI Use Cases](#) [Free GenAI Training](#)

Databricks Mosaic AI  
出典: [Databricks HP](#)



生成AI



MLOps



ディープラーニング

## Mosaic AI とは

Mosaic AIはDatabricksが提供する統合AIプラットフォームで、データとAIのライフサイクル全体を単一の環境でサポートします。従来の機械学習からLLM・生成AIまで、幅広いAIワークフローに対応します。

統合開発環境：実験からデプロイまでワンストップで提供、MLエンジニアの生産性向上

AI Gatewayとモデルサービング：複数のAIモデルを一元管理・提供するインフラ

MLflow統合：実験管理・モデル追跡・再現性確保を自動化

LLMOps：大規模言語モデルの最適化・デプロイ・管理を効率化

### 2025年の新機能

AI Gateway強化

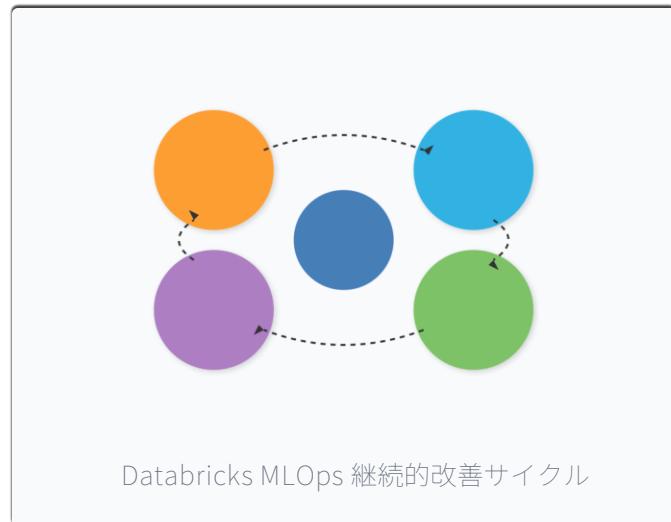
AIガバナンス自動化

Vector Search高速化

MLモデル監視強化

詳細はAzure Databricks ML公式ドキュメントを参照

# MLOpsワークフロー



DatabricksのMLOpsワークフローは、機械学習モデルの開発から本番環境への展開、監視、継続的改善までの一連のプロセスを統合的に管理します。MLflowを中心としたエンドツーエンドのライフサイクル管理により、モデルの信頼性と再現性を確保します。

開発フェーズ：MLflowによる実験管理、パラメータ追跡、モデルの再現性確保。Notebooks、自動ML、ハイパーパラメータチューニングを活用

デプロイフェーズ：MLflow Model Registry、CI/CD連携、Databricks Model Servingによるモデルのバージョン管理と本番環境への展開

モニタリングフェーズ：モデル性能、データドリフト、予測品質の継続的な監視と分析

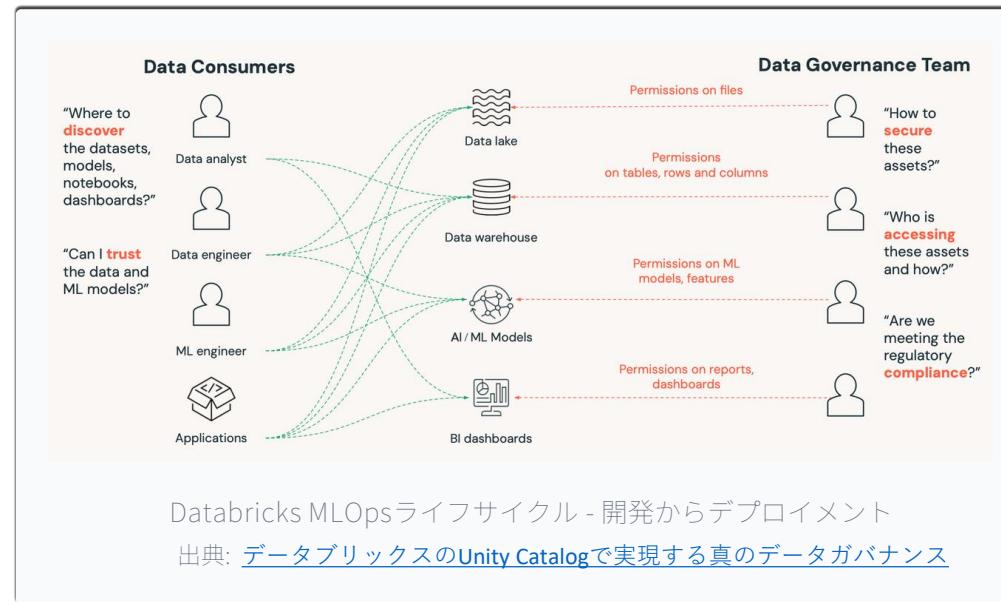
フィードバックフェーズ：モデル更新トリガーの自動化、A/Bテスト、改良サイクルの最適化

## MLOps自動化ポイント

- ✓ データ検証と品質チェック
- ✓ CI/CDパイプラインとの統合
- ✓ モデルの自動デプロイ
- ✓ パフォーマンス監視とアラート

詳細はAzure Databricks MLOps ドキュメントを参照

## Databricksでのモデル開発ライフサイクル



Databricksプラットフォーム上のMLOpsプロセスは、モデル開発から本番デプロイまでをシームレスに統合し、AIプロジェクトの価値実現を加速します。

実験と開発：Notebookでのコラボレーション開発、MLflowによる実験追跡、自動化されたハイパーパラメータ最適化

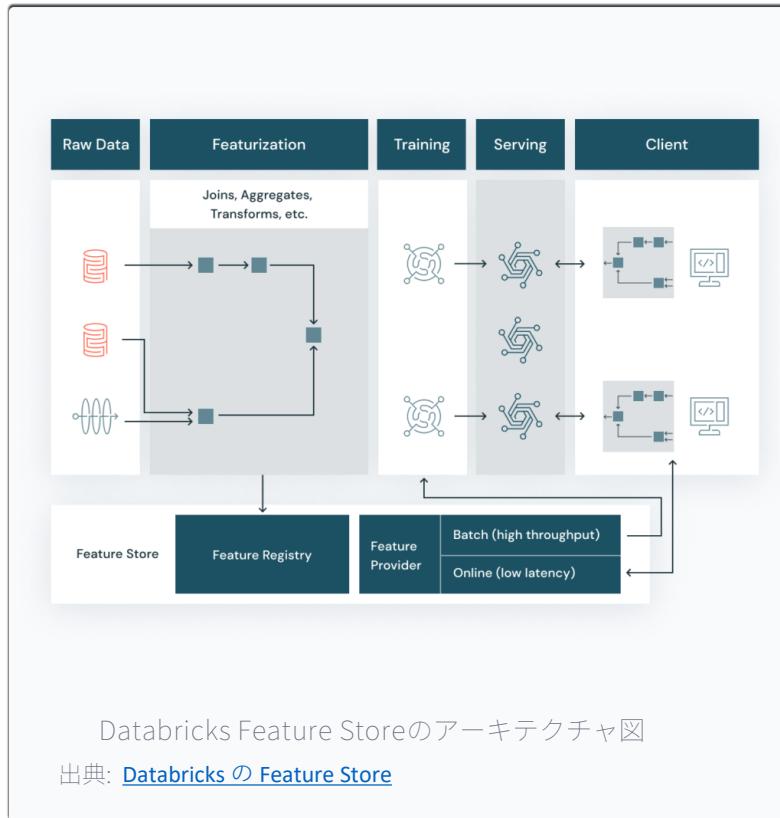
モデルの登録と管理：MLflow Model Registryでのバージョン管理、承認ワークフロー、Unity Catalogとの統合による全社的なガバナンス

デプロイメント：複数のデプロイオプション（バッチ推論、Mosaic AI Model Serving、REST API）によるスケーラブルなモデル配信

モニタリングとフィードバック：パフォーマンス監視、特徴量ドリフト検知、自動再トレーニングパイプラインの構築

### 2025年の強化ポイント

- ✓ LLMOps - 大規模言語モデルに特化した運用ワークフロー
- ✓ Auto-Scaling Model Serving - コスト効率とパフォーマンスの最適化
- ✓ 自動モデルドキュメント生成 - ガバナンス強化と透明性向上



特徴量の保存



特徴量の共有・再利用



リアルタイム提供

## Feature Storeの活用

Feature Store（特徴量ストア）は、機械学習モデルで使用する特徴量を効率的に管理・提供するための中央リポジトリです。Databricks Feature Storeは、Unity Catalogと統合され、開発から本番までのMLライフサイクル全体をサポートします。

特徴量の一元管理：全てのML特徴量を一か所で管理し、チーム間で共有・再利用が可能

オンライン/オフラインストア：バッチ処理とリアルタイム推論の両方に対応

バージョン管理とリネージ：特徴量の変更履歴と依存関係を追跡

MLflowとの統合：モデル学習から推論までのプロセスをシームレスに連携

### 主なユースケース

リアルタイム推論

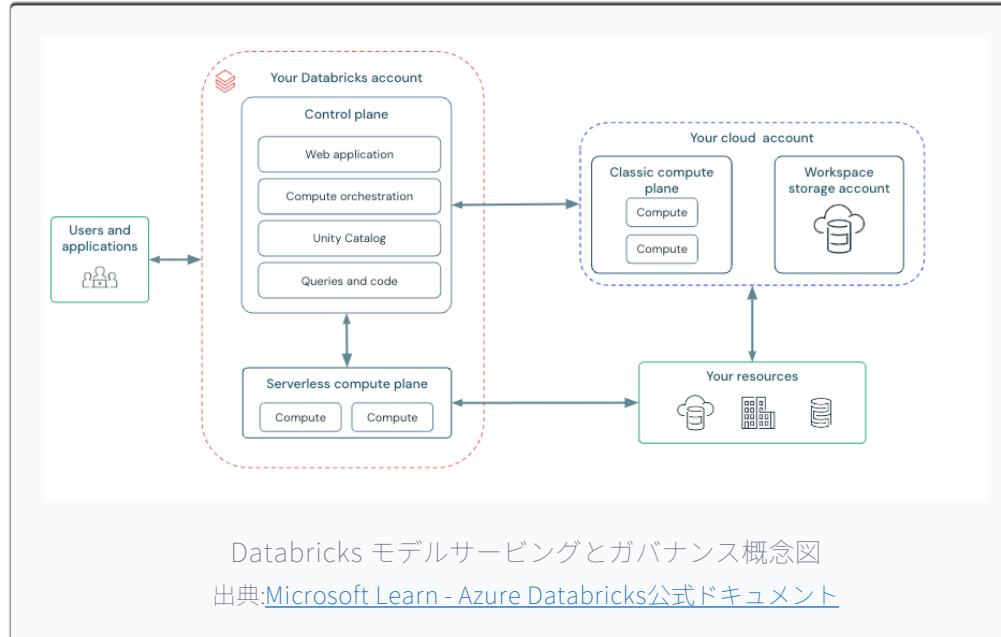
パーソナライゼーション

不正検知システム

予測分析モデル

Feature Storeの詳細は[こちら](#)

## モデルサービングとガバナンス



Databricks Mosaic AI Model Servingは、機械学習モデルを本番環境に効率的にデプロイし、一元的にガバナンスするためのエンタープライズグレードの仕組みを提供します。

スケーラブルなサービス：サーバレスインフラストラクチャによる自動スケーリングと高可用性

統合ガバナンス：Unity Catalogとの連携によるモデルのアクセス制御・監査・リネージ管理

モデル監視：モデルパフォーマンス・ドリフト検知・説明可能性の継続的なモニタリング

MLflow統合：モデルレジストリとの完全統合によるバージョン管理とロールバック機能



### モデルサービングの主な機能

⌚ リアルタイム推論

⌚ A/Bテスト

🔒 アクセス制御

⌚ モデル説明可能性

🔗 詳細はAzure Databricks Model Serving公式ドキュメントを参照

## 生成AI・AI Gateway新機能



Mosaic AI Gatewayは、多様なAIモデルとの接続を一元管理し、セキュリティとコスト管理を強化する2025年の新機能です。企業内のLLMやAIモデルの安全かつスケーラブルな活用を実現します。

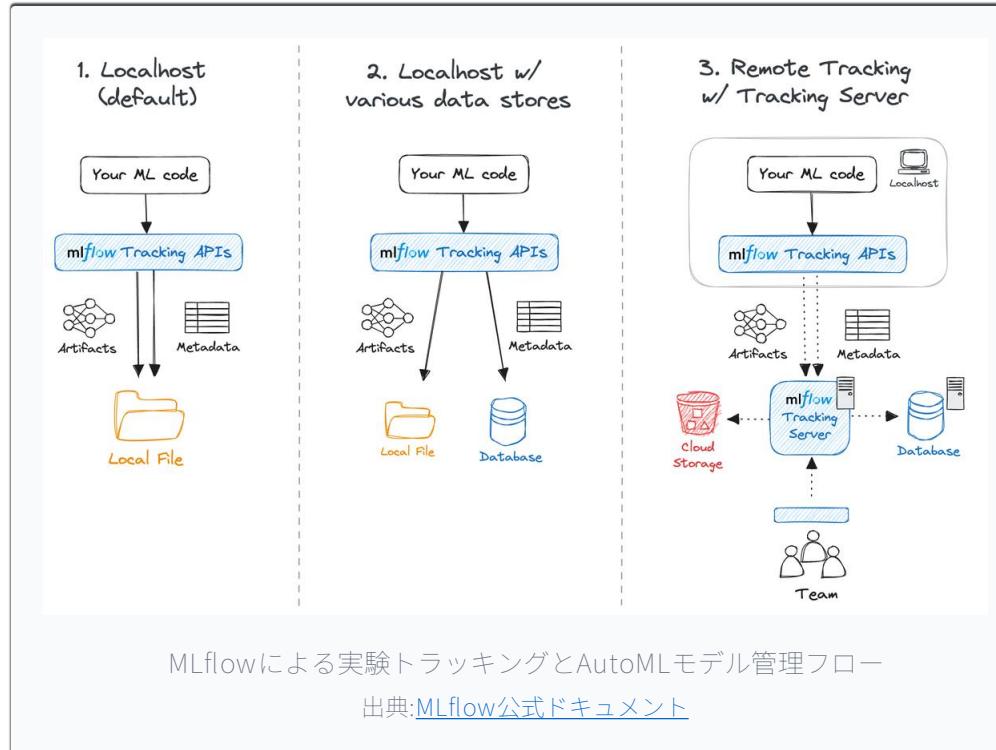
🔑 中央管理型認証・認可

⌚ 使用状況モニタリング

📋 監査ログ

🤖 RAG実装の簡素化

# AutoMLと実験管理



DatabricksのAutoMLとMLflow統合による実験管理機能は、モデル開発ライフサイクル全体を効率化し、ガバナンスを強化します。データサイエンティストの生産性向上と、企業レベルのAI実装を加速します。

AutoMLの主要機能：特徴量エンジニアリングの自動化、モデル選択、ハイパーパラメータチューニングの自動最適化

実験のバージョン管理：すべてのモデル、パラメータ、メトリクスを自動追跡し、再現性を確保

モデル比較と選択：異なるアルゴリズム・パラメータセットの性能を可視化し、最適モデルを選定

MLOpsとの統合：実験からプロダクション環境へのシームレスな移行を実現

## MLflowによる実験管理の利点

⌚ 実験履歴の完全追跡

⌚ 再現可能なワークフロー

🔗 チーム間コラボレーション

➡ モデル展開の簡素化

🔗 詳細はAzure Databricks AutoML公式ドキュメントを参照



## 第6部 ワークロード活用パターン

代表的な活用シナリオごとに構成例を提示します。

### ワークロード活用パターン

BI・アナリティクス ダッシュボードによる意思決定支援、Power BIやTableauとの統合

データエンジニアリング ETLパイプライン自動化、データ品質管理、メタデータ統合

データサイエンス・ML 予測分析、クラスタリング、推薦エンジン開発

ストリーミング分析 リアルタイムデータ処理、異常検知、イベント処理

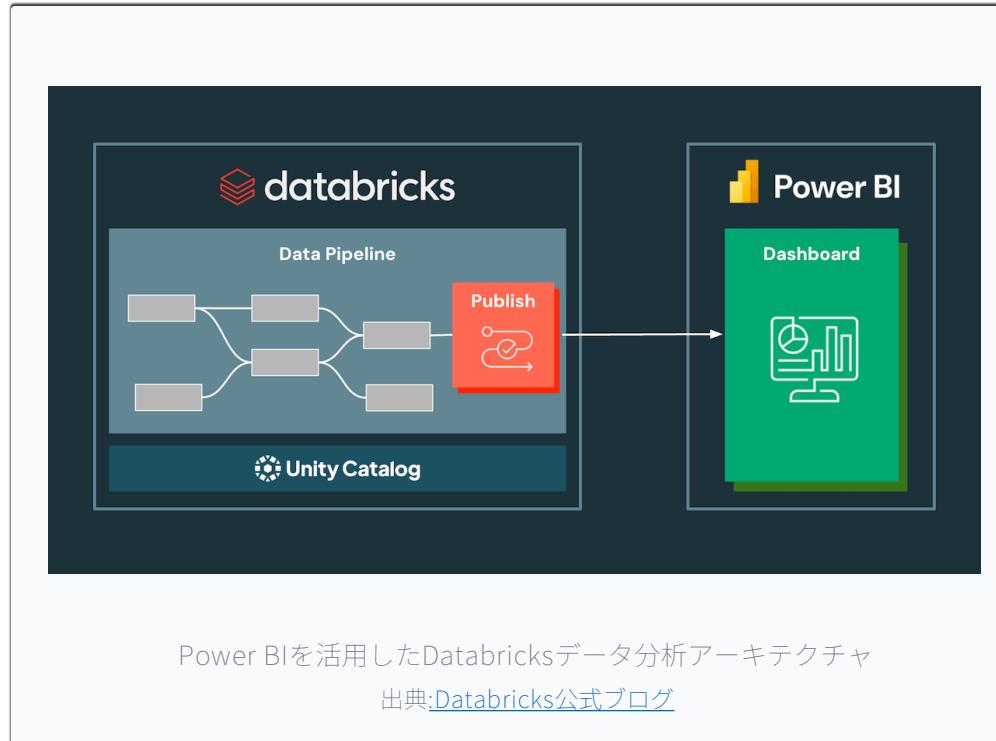
生成AI・RAGアプリケーション 大規模言語モデル活用、ナレッジベース検索拡張

外部システム連携 クラウドサービス統合、エンタープライズデータ連携

✉ 詳細情報: [Azure Databricksユースケース & ソリューション](#)



## BI・アナリティクスワークフロー



DatabricksのBI・アナリティクスワークフローは、SQLウェアハウスと主要なBIツールとの統合により、エンタープライズグレードの分析環境を提供します。Unity Catalogによる統一ガバナンスを基盤にして、データレイクハウス上のデータを直接可視化・分析できます。

Databricks SQL：高性能なSQLウェアハウス機能で、複雑なクエリも高速実行

ダッシュボードと可視化：内蔵のダッシュボード機能と外部BIツール連携による柔軟な可視化

Unity Catalogセマンティクス：統一されたメトリクス定義と意味づけで一貫したレポーティング

エンドツーエンドのガバナンス：データ品質、セキュリティ、系統追跡が組み込まれた環境

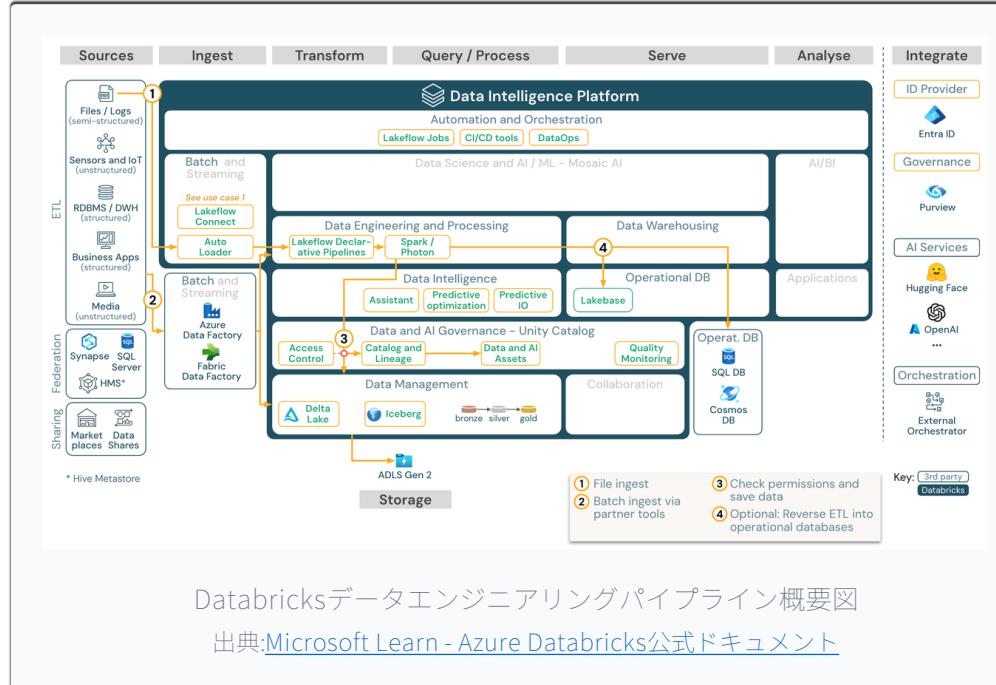


### 主なユースケース

- ↳ セルフサービスアナリティクス      ↳ リアルタイムダッシュボード
- ↳ データ探索と分析      ↳ エンタープライズレポートイング

↳ 詳細はDatabricks SQLのドキュメントを参照

## データエンジニアリングの最適化



Databricks プラットフォームは、高度なETL/ELTプロセスを設計・開発・運用するための強力な基盤を提供します。従来のデータエンジニアリングプロセスを自動化・最適化し、高速で安定したデータパイプラインを実現します。

**Lakeflow宣言型パイプライン：**コードを減らしメンテナンス性を高めるSQL/Python宣言型ETL

**Auto Loader：**クラウドストレージからの増分データ取り込みを自動化

**Lakeflow Connect：**SaaSアプリやデータベースからのノーコード連携

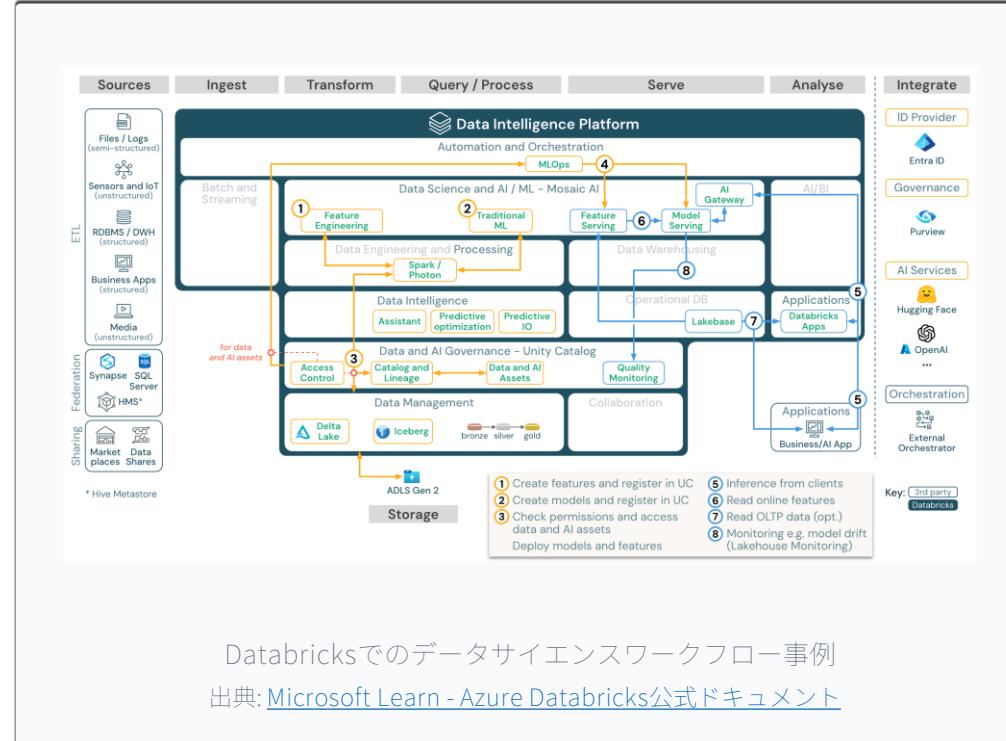
**メダリオンアーキテクチャ：**Bronze→Silver→Goldの階層的データ処理

### 主要な機能と効率化

- ✓ データ品質チェックの自動化
- ✓ スケーラブルなバッチ処理
- ✓ マルチタスクワークフロー制御
- ✓ メタデータ・リネージュ統合
- ✓ ジョブスケジューリング
- ✓ 統合モニタリング・通知

🔗 ETL クイックスタート - Azure Databricks公式ドキュメント

# データサイエンス・MLプロジェクト事例



需要予測



顧客セグメンテーション



異常検知



画像処理



自然言語処理



生成AI

Databricksを活用したデータサイエンス・機械学習プロジェクトでは、エンドツーエンドの実装が可能です。実際の業界導入事例と実装パターンを紹介します。

## 代表的なユースケース

需要予測：小売・物流企業での在庫最適化、売上予測モデル実装

顧客セグメンテーション：金融機関でのリスクスコアリング、購買行動分析

異常検知：製造業での品質管理、セキュリティ監視システム

画像・ビジョン：医療画像分析、商品認識システム、検査自動化

自然言語処理：感情分析、文書分類、質問応答システム

## 実装アプローチ

Delta Lakeでのデータ準備と特微量エンジニアリング

MLflowでの実験管理と再現性確保

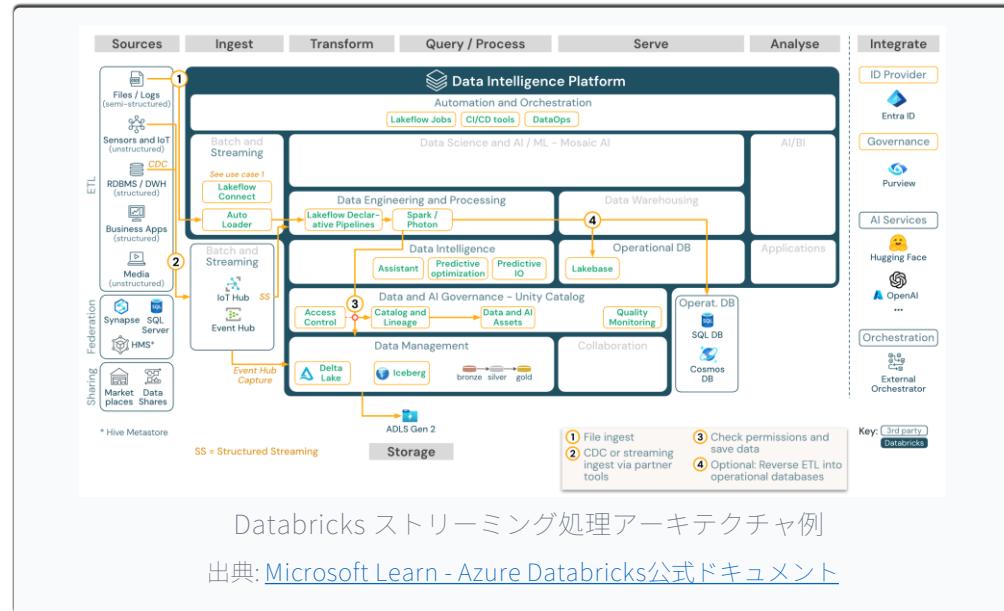
AutoMLによる迅速なプロトタイピング

Feature Storeでの特微量共有と管理

モデルレジストリによるバージョン管理

モデルサービスによる本番環境へのデプロイ

## リアルタイムデータ処理と意思決定



Databricksのストリーミング分析機能を活用すると、リアルタイムデータに基づく迅速な意思決定と自動化されたアクションが実現できます。Structured Streamingを活用した高性能・高信頼のストリーム処理が特徴です。

低レイテンシー処理：マイクロバッチまたは連続処理モードでミリ秒レベルの応答を実現

デルタテーブル連携：Delta Lakeとのシームレスな統合でACIDトランザクションを保証

エンド・ツー・エンドの信頼性：Exactly-onceセマンティクス、ウォーターマーク、ステートフル処理をサポート

MLモデルとの連携：リアルタイム推論と連携した自動アクション実行

### 主要ユースケース

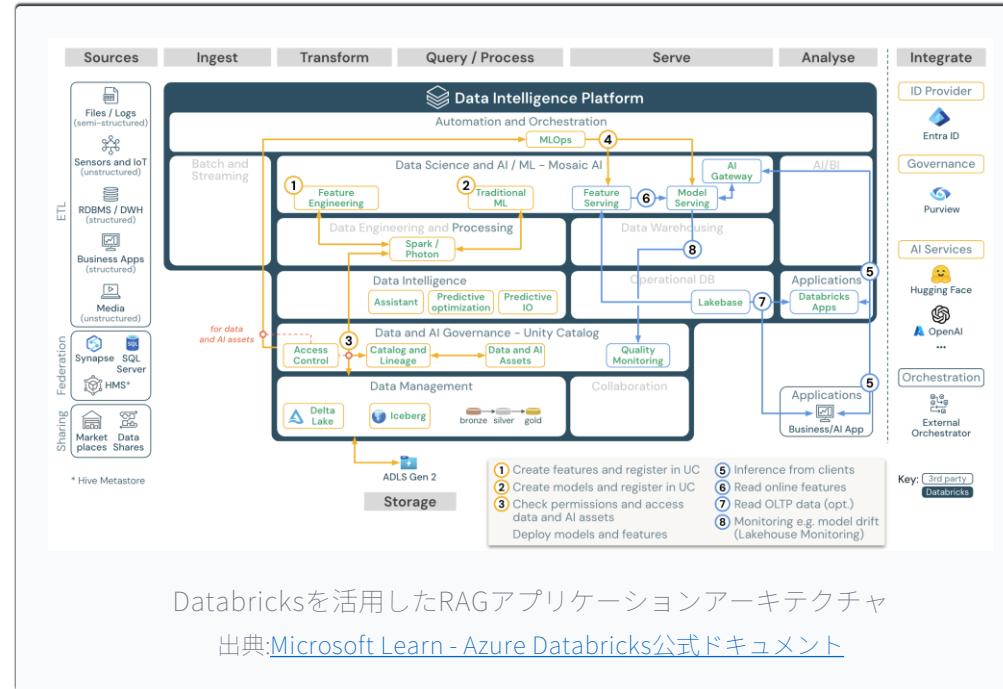
⚠️ リアルタイム異常検知

⚡ 繙続的なKPI監視

⚙️ IoTデバイス分析

🛡️ リアルタイム不正検出

↗ Structured Streamingの詳細はこちら



## 生成AI・RAG応用

RAG (Retrieval Augmented Generation) は、生成AIの回答精度と信頼性を高めるために、外部データソースから関連情報を検索・抽出し、プロンプトに組み込む手法です。Databricksプラットフォームでは、エンドツーエンドのRAGソリューションを効率的に構築できます。

データの鮮度確保：最新の社内データを活用した回答生成

幻覚の軽減：事実に基づいた情報提供による精度向上

ドメイン特化：業界・企業固有の知識ベース活用

マルチモーダル対応：テキスト、画像、表形式データの統合活用

### Databricksでの実装ポイント

- 1 Vector Searchを活用した高速検索 - Unity Catalogと連携したスケーラブルな検索基盤
- 2 Lakeflow Jobsによる自動更新 - ナレッジベースを継続的に最新化
- 3 AI Gateway & モデルサービング - LLMやエンベディングモデルの統合管理

詳細はVector Search ドキュメントを参照

### RAGの主要コンポーネント

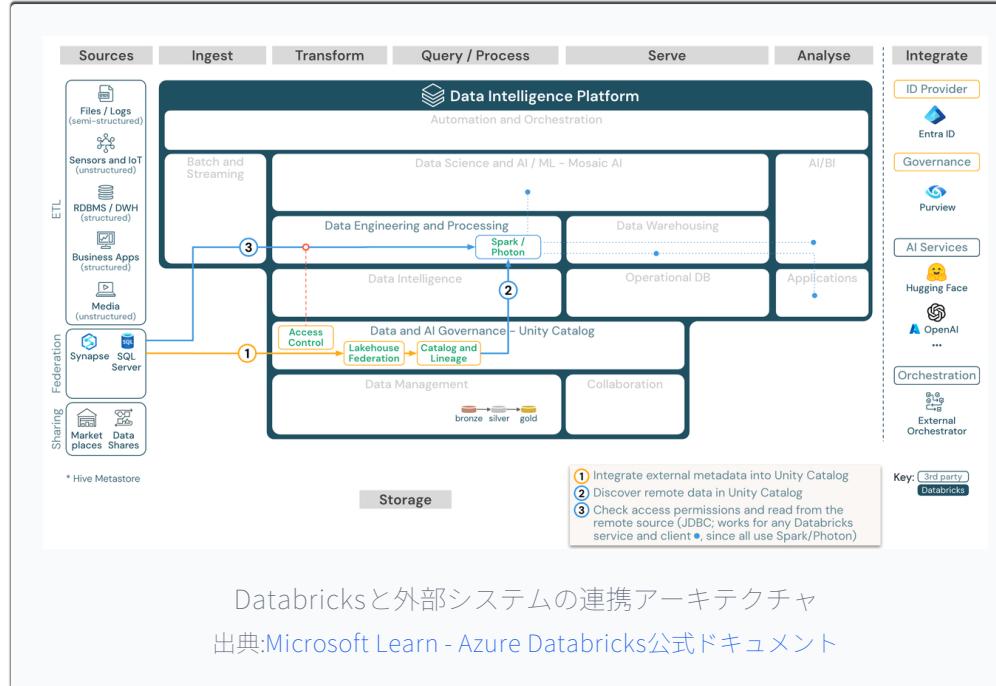
ベクトルデータベース

生成AIモデル

検索・検索拡張

パイプライン管理

# 外部システム連携



Databricksと外部システムの連携アーキテクチャ

出典: Microsoft Learn - Azure Databricks公式ドキュメント



Databricksは様々な外部システムと柔軟に連携し、データの統合・活用を実現します。データを移動せずに直接アクセスすることで、効率的なワークフローを構築できます。

Lakehouse Federation : 外部DBをUnity Catalogに統合し、移動なしで直接クエリ

カタログフェデレーション : 既存のHiveメタストアをUnity Catalogに統合

REST API連携 : 包括的なAPIを通じた自動化・外部アプリケーション連携

ETL/ELTコネクタ : 各種データソースからの効率的な取り込み

## 主要連携システム例

Azure Synapse

MySQL/PostgreSQL

Snowflake

SAP/Salesforce

詳細はAzure Databricks クエリフェデレーション公式ドキュメントを参照

## 第7部 セキュリティと ガバナンス

Databricksプラットフォームのセキュリティ要点まとめ。多層防御の考え方と実装方法を解説します。

### セキュリティベストプラクティス

多層防御アプローチによるセキュリティ設計

IAM統合とID管理のベストプラクティス

ネットワーク分離とPrivate Linkの活用

データ暗号化（保存中・転送中）の実装方法

### コンプライアンス要件への対応

GDPR、CCPA、HIPAAなど主要規制への対応方法

業界別コンプライアンス要件（金融・医療・公共）

監査ログとコンプライアンスレポート作成の自動化

### データプライバシーとユーザーガバナンス

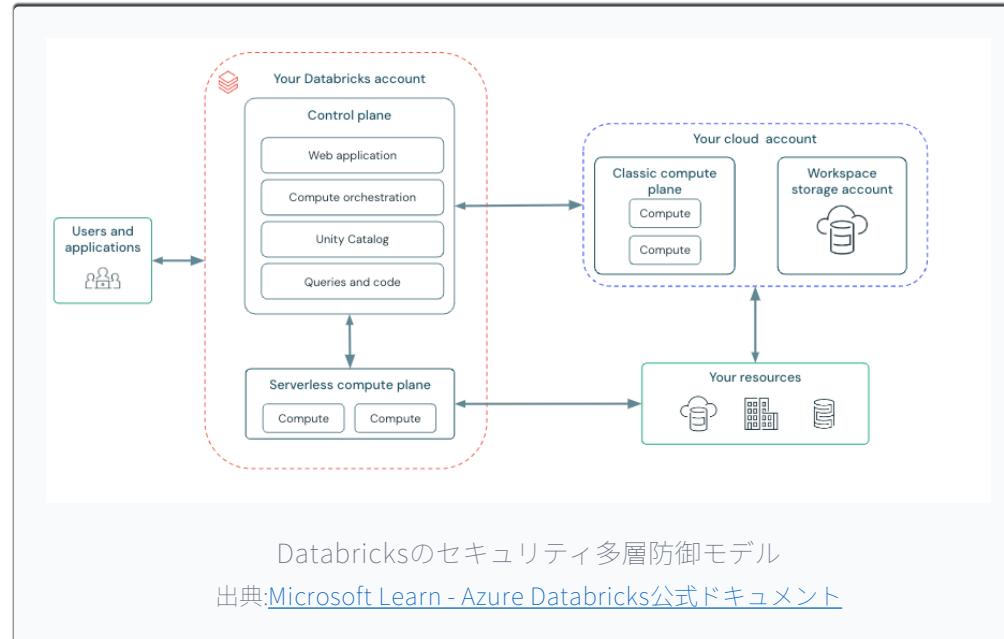
属性ベースのアクセス制御（ABAC）によるきめ細かな権限管理

データ分類・マスキング・匿名化技術

Unity Catalogによる統合ガバナンスの実現方法

SAIFフレームワークを活用したセキュリティ管理





## セキュリティベストプラクティス

Databricks プラットフォームを安全に運用するための主要なセキュリティベストプラクティスを紹介します。これらの対策を組み合わせることで、包括的なセキュリティ体制を確立できます。

アクセス管理：Unity Catalogによる細粒度アクセス制御の実装、IAM ロールの最小権限原則の徹底、SSO連携

データ保護：保管時と転送時の暗号化、顧客管理暗号化キー(CMEK)の活用、自動データ分類によるPII保護

ネットワーク分離：PrivateLink/VNet連携、IP許可リスト、Secure Cluster Connectivity (SCC) の有効化

監査と検知：包括的な監査ログの有効化、異常検知、Security Analysis Tool(SAT)による継続的なセキュリティ姿勢評価

### セキュリティの4つの柱

- |               |           |
|---------------|-----------|
| アイデンティティとアクセス | データ保護と暗号化 |
| ネットワーク分離      | 監査と可視性    |

### セキュリティの自動化

Terraform/AzureCLIを使用したセキュリティ設定の自動化と Security Reference Architecture(SRA)テンプレートの活用により、一貫したセキュリティ対策の実装が可能です。

詳細はDatabricksセキュリティベストプラクティスガイドを参照

## グローバルコンプライアンスへの対応



Databricksは様々な国際規制や業界基準に対応し、グローバル規模でのコンプライアンス要件を満たすための包括的なフレームワークを提供しています。

地域固有のデータ主権対応：EUデータ所在地要件、GDPR、米国CCPA/CPRAなど各地域特有のデータプライバシー法に対応

業界別コンプライアンス：金融（GLBA、SOX）、医療（HIPAA）、小売（PCI DSS）など、業界ごとの規制に準拠

第三者認証と監査：SOC 1/2/3、ISO 27001、FedRAMP（政府機関向け）などの国際認証を取得

共有責任モデル：プラットフォーム側と顧客側の責任分界点を明確化したコンプライアンスフレームワークを提供

### コンプライアンス対応のキーポイント

- ✓ Unity Catalogによる統合アクセス制御と監査ログ機能
- ✓ データの暗号化（保存時・転送時）とプライバシー保護機能
- ✓ クラウドプロバイダーとの連携による地域特定のコンプライアンス要件対応
- ✓ 繙続的なセキュリティアップデートと脆弱性管理

### 業界別コンプライアンス対応

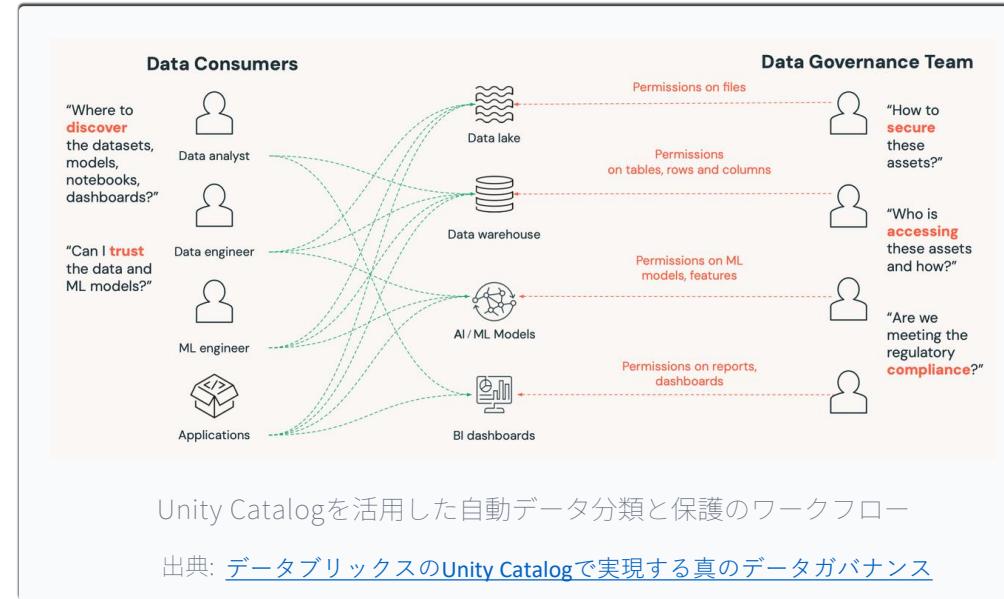
金融：FINRA, Basel III

医療：HIPAA, HITECH

小売：CCPA, PCI DSS

製造：ISO 9001, IATF

# データプライバシー保護と効率的ガバナンス



## ABAC (属性ベースアクセス制御) 適用例

- PII データへのアクセス制限
- 地域別アクセス権限
- 部門・役割別ポリシー
- データ機密度レベルによる制御

Databricksのデータプライバシーとユーザーガバナンス機能は、企業の厳格なコンプライアンス要件に対応しながら、データ活用を促進するための包括的なソリューションを提供します。

**自動データ分類**：機械学習を活用してPIIなどの機密情報を自動検出・タグ付け、保護ポリシーを自動適用

**動的マスキング**：ユーザーの権限に基づいた列レベル・行レベルでの動的なデータマスキング

**属性ベースアクセス制御 (ABAC)**：役割、部門、地域などの属性に基づく柔軟なアクセスポリシー

**タグポリシー**：一貫したデータ分類と保護のための標準化されたタグ付けシステム

**統合監査・証跡**：データアクセス、ポリシー変更、クエリ実行など全操作の詳細なログ記録

## 実装ベストプラクティス

- 🛡 最小権限の原則に基づいたアクセス設計
- 🛡 データ分類タクソノミーの標準化と自動適用
- 🛡 定期的な監査とアクセス権レビューによるプロセスの確立

# Databricks プラットフォームの全体像



アーキテクチャ  
Control/Compute Plane分離設計



Unity Catalog  
統一データ・AIガバナンス



メダリオンアーキテクチャ  
Bronze/Silver/Gold構造



Mosaic AI  
統合AI/ML開発・運用基盤



セキュリティ  
多層防御・コンプライアンス対応



多様なワークLOAD  
BI/ETL/ML/Gen AI統合



Databricksは統合データ・AI基盤として、データエンジニアリングからAI開発・運用まで一貫したプラットフォームを提供します。Unity Catalogによるガバナンスとオープンフォーマット対応で、データとAIの民主化を実現し、企業のデータ活用を加速します。

より詳細な情報 ご質問・疑問点 デモ・ハンズオン 次のステップ