



databricks

# Databricks Data Lakehouse & Data Intelligence Platform の理解

データレイクとデータウェアハウスの融合

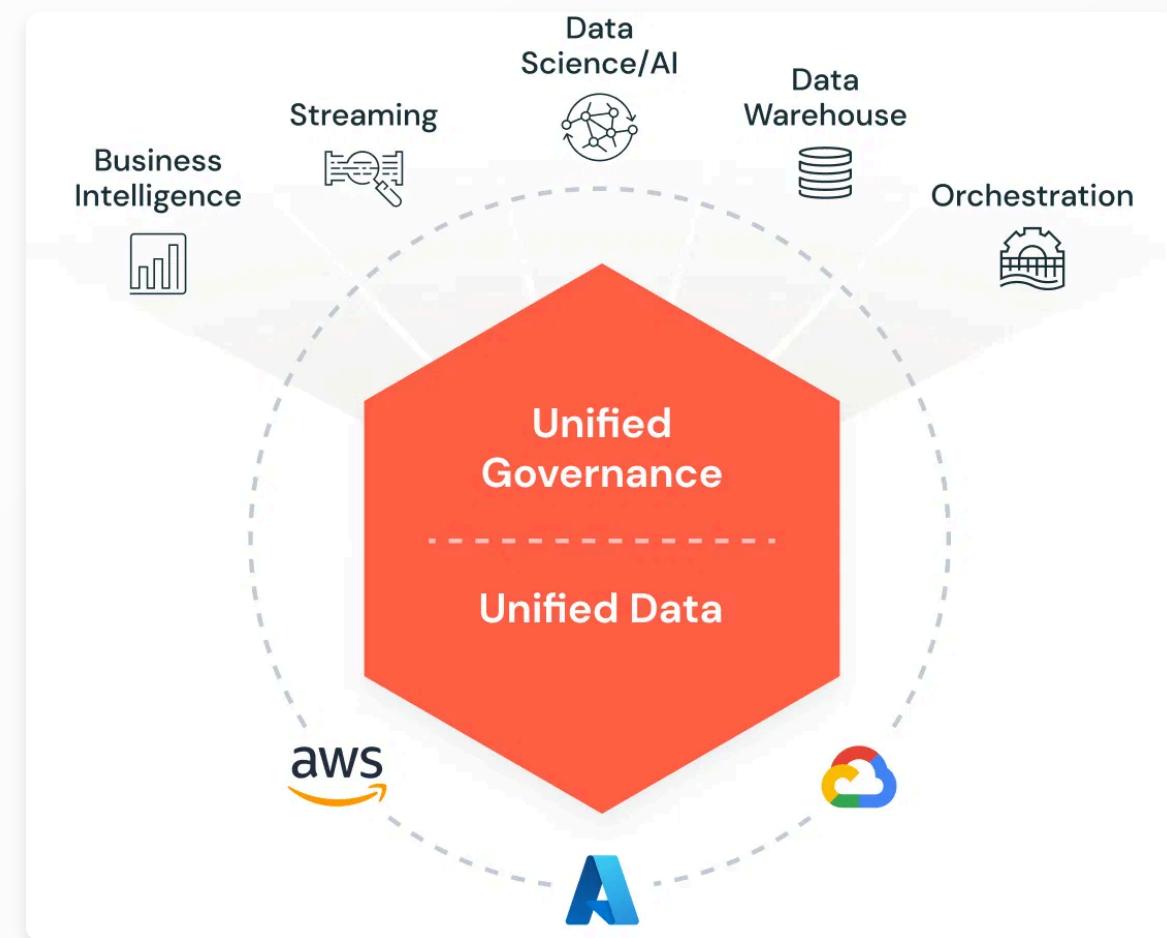
Databricksの最新データアーキテクチャとそのメリットについて学びましょう

# Databricks Unified Solution

**Databricks Unified Solution**は、データエンジニアリング、データウェアハウジング、機械学習を統合する単一プラットフォームです。

主なメリット：

- ・データアーキテクチャの簡素化
- ・データインシアチブの加速
- ・データの民主化



# Unified Solutionの構成要素

Databricks Unified Solutionは、以下の主要な領域を統合しています：

## データエンジニアリング

データの取り込み、変換、クレンジングを効率化し、高品質なデータパイプラインを構築

## データウェアハウジング

高速なクエリパフォーマンスと分析機能を提供し、ビジネスインテリジェンスをサポート

## 機械学習

モデルの開発、トレーニング、デプロイを簡素化し、AIアプリケーションの構築を加速

これらの要素が单一のプラットフォームで統合されることで、データサイロの解消とエンドツーエンドのデータライフサイクル管理が実現します

# Databricksとオープンソース: Apache Spark

**Apache Spark**は、大規模データ処理のための統合分析エンジンです。

## 高速処理

メモリ内処理により、従来のMapReduceよりも最大100倍高速

## 多言語サポート

Scala、Java、Python、R、SQLなど複数の言語をサポート

## 統合フレームワーク

バッチ処理、ストリーミング、機械学習、グラフ処理を単一のエンジンで実現



## Databricksにおける重要性

Databricksの創設者はApache Sparkの開発者であり、Sparkを中心としたエンタープライズ向けプラットフォームを提供しています。

# Databricksとオープンソース: Delta Lake

**Delta Lake**は、データレイクに信頼性をもたらす  
オープンソースのストレージレイヤーです。

## ACIDトランザクション

データの整合性と信頼性を確保するトランザクションサポート

## スキーマ進化と強制

データ品質を維持しながらスキーマの変更を柔軟に管理

## タイムトラベル

以前のデータバージョンへのアクセスと監査履歴の確認が可能



## DELTA LAKE

## データレイクの課題を解決

Delta Lakeは、従来のデータレイクの課題である信頼性、パフォーマンス、品質の問題を解決し、Data Lakehouseアーキテクチャの基盤となっています。

# Databricksとオープンソース: MLflow

MLflowは、機械学習ライフサイクルを管理するためのオープンソースプラットフォームです。

## 実験の追跡

パラメータ、メトリクス、結果、コードを記録し、実験を比較

## モデルの管理

モデルのパッケージ化、バージョン管理、共有を簡素化

## モデルのデプロイ

様々な環境へのモデルデプロイを標準化



## 機械学習の課題を解決

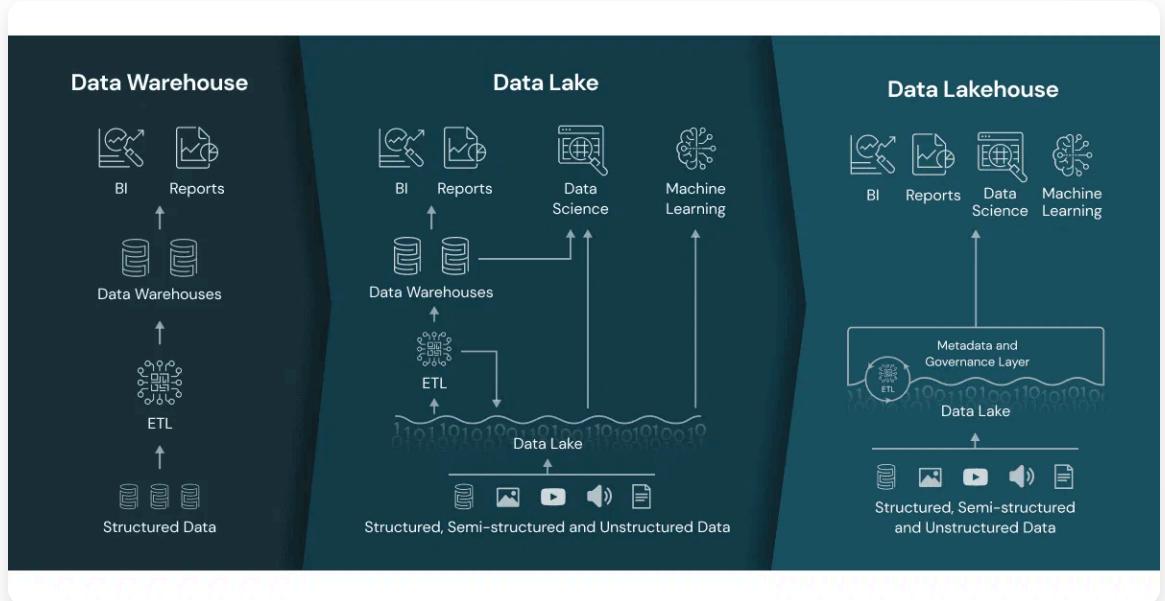
MLflowは、機械学習プロジェクトの再現性、追跡可能性、共有の課題を解決し、エンドツーエンドの機械学習ライフサイクル管理を実現します。

# Data Lakehouseとは？

**Data Lakehouse**は、データレイクとデータウェアハウスの最高の機能を組み合わせた新しいデータアーキテクチャです。

## Data Lakehouseの定義:

- ・データレイクの柔軟性とスケーラビリティ
  - ・データウェアハウスのデータ管理とパフォーマンス
  - ・構造化・非構造化データの統合ストレージ
  - ・トランザクションのサポートと一貫性の保証



# Data Lakehouseのメリット

## データアーキテクチャの簡素化

複数のシステムを統合し、データサイロを解消

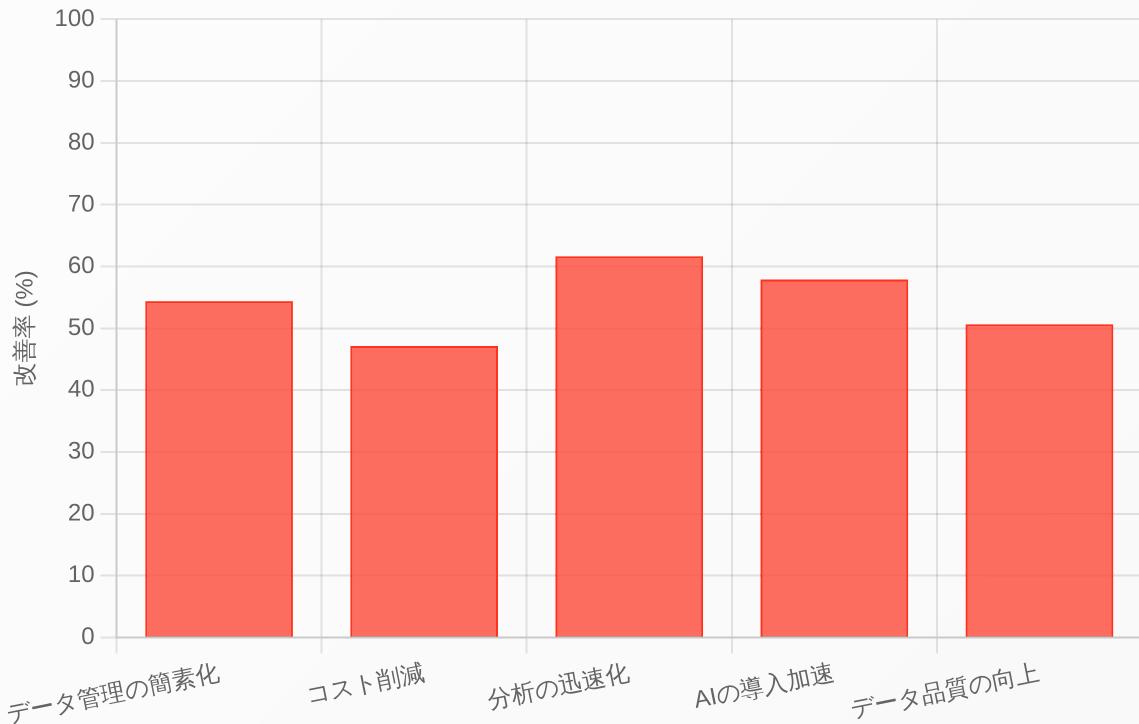
## データコストの削減

重複したシステムやデータ移動の必要性を低減

## データイニシアチブの加速

データエンジニアリングから分析、AIまでのプロセスを効率化

## Data Lakehouseによる改善効果



# Databricks Lakehouse Platform

**Databricks Lakehouse Platform**は、データレイクとデータウェアハウスの最高の機能を組み合わせた統合データプラットフォームです。

このプラットフォームは、すべてのデータ関連アクティビティを単一の場所で提供します。

- データエンジニアリング
- データウェアハウジング
- 機械学習
- ビジネスインテリジェンス

 データアーキテクチャの簡素化

 データイニシアチブの加速

 データの民主化

 データの可能性を最大限に引き出す

Databricks Lakehouse Platformは、データの収集から分析、機械学習まで、データライフサイクル全体をサポートします。

# Databricks Data Intelligence Platform

# まとめ

Databricks Data Lakehouse & Data Intelligence Platformの主要ポイント：

## Databricks Unified Solution

データエンジニアリング、ウェアハウジング、機械学習を統合

## オープンソースの活用

Apache Spark, Delta Lake, MLflowで柔軟なプラットフォームを提供

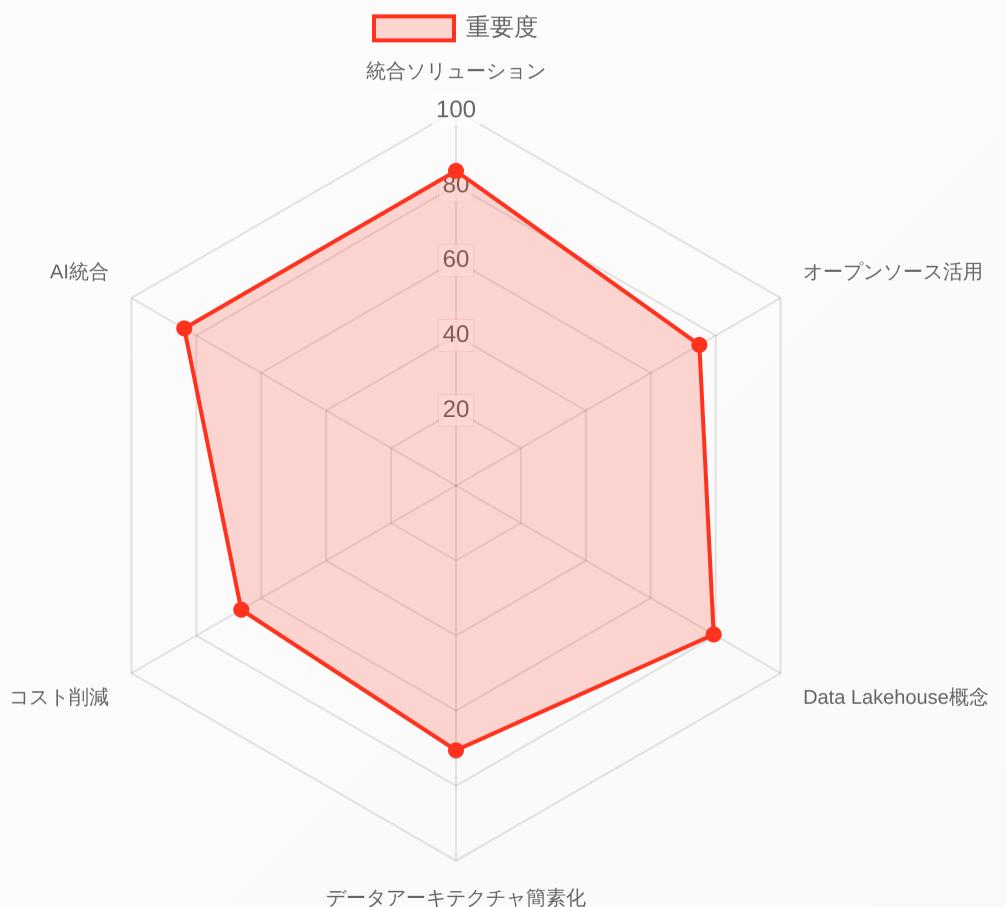
## Data Lakehouse

データレイクとデータウェアハウスの利点を融合した新アーキテクチャ

## Data Intelligence Platform

AIと生成AIを統合し、データの価値を最大化する次世代プラットフォーム

Databricks Data Lakehouse & Intelligence Platform 主要ポイント



## 詳細情報

Databricks公式サイト: [www.databricks.com](http://www.databricks.com)