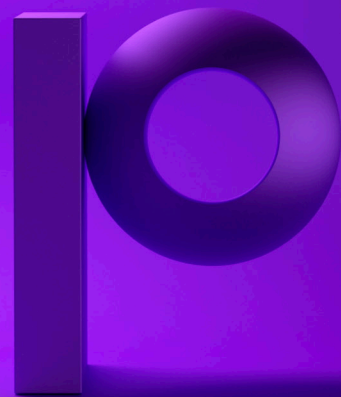




データサイロを打破し、データ活用を加速させる



precisely

Databricks と Precisely Connect

従来のメインフレームや IBM i のデータを Databricks

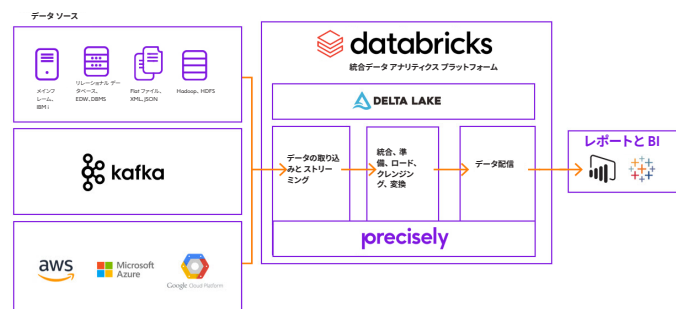
統合データ アナリティクスプラットフォームに統合することで、データサイロを分解してクラウドベースの AI および ML プロジェクトに活用

組織内の膨大なデータは、メインフレーム、IBM i システム、Hadoop、Oracle、Db2 などに分散して保存されることが多いです。データサイロは、増大の一途を辿るデータ量とデータ形式の不適合問題と相まって、革新的なプロジェクトのためにデータを活用する際の妨げとなってしまいます。こうした障壁のせいで、お客様のアナリティクス、AI、ML プロジェクトが頓挫したり遅延したりすることがあってはなりません。

Databricks と Precisely を併用すると、企業データの保存場所に関係なく、高性能アナリティクス、AI、機械学習を促進しやすくなります。データソースの統合によりお客様のビジネスでイノベーションを実現すれば、特別なスキルがなくても成功を手に入れることができます。

レガシー データとクラウドを結ぶパイプラインの構築

Connect を利用してデータパイプラインを構築すると、Databricks 統合データアナリティクス プラットフォームと Delta Lake 内でレガシー ソースからデータを自由に活用できます。Connect は、ETL ワークフローに「Design once, deploy anywhere(たった一度の設計でどこでも使える)」アプローチを提供します。このインターフェースを使用して、同じ画面で ETL ワークフローのバッチ処理とストリーミングを両方定義できます。このデータ統合を可視化するアプローチは、高性能データパイプラインの構築にフォーマット変換やコード生成が不要であることを意味します。ポイントアンドクリックで変換できるので、急激に拡大・進化し続けているテクノロジースタックに時間と労力を集約できます。



データ レイクハウスの構築

Connect を利用すると、レガシー データ ストアから Databricks 統合アナリティクスプラットフォームにデータを効率的にオフロードすることでデータ レイクハウスを構築できます。

- **メインフレーム データ:** VSAM、COBOL Copybooks、メインフレーム固定およびシーケンシャル ファイル
- **RDBMS:** Oracle、SQL、Db2、MySQL、Sybase、PostgreSQL
- **半構造化データ:** JSON、XML
- **エンタープライズ データ ウェア ハウス:** Teradata、IBM Netezza、Vertica、Greenplum
- **クラウド:** Amazon AWS、Microsoft Azure、Google Cloud Platform
- **ビッグ データ:** Hadoop、Hive
- **ストリーミング プラットフォーム:** Apache Kafka
- **フラット ファイル:** 固定長、可変長、区切り記号付き

Connect は、データのオフロードに対してエンドツーエンドマネージド アプローチを取ります。どのソースを選ぶかに関係なく、数百または数千のテーブル (データベース スキーム全体を含む) を Databricks に迅速かつ容易にレプリケートできます。

機械学習プロジェクトのスケーリング

機械学習プロジェクトでは、複雑なレガシー データストアを大規模に収集し、元の形式のまま統合する必要があります。Connect は、すべてのレガシー データストアから必要なデータを収集し、Databricks に送信します。これにより、機械学習用の拡張可能なフレームワークが実現します。Connect は Databricks Runtime と元の形式のまま統合します。Databricks クラスターで直接 Connect のインテリジェント エンジンを使用すると、厳格な SLA 内でペタバイト規模のデータセットをクレンジング、変換し、分析用に準備できます。

