



Connect: クラスタからクラウドへの ETL

すべてのエンタープライズプラットフォームに対応するパフォーマンスと拡張性に優れたデータ統合



ETL を最新化し、企業全体の全データを統合する

Connect は、RDBMS、メインフレーム、NoSQL、クラウドなどから 1 つの企業全体の全データを統合します。絞り込んだデータセットを抽出、ソート、ブレンド、準備、変換し、すべての主要エンタープライズデータウェアハウスや、BI および最先端の分析プロジェクト向けのクラウドデータウェアハウスに読み込みます。Connect を使用すると、以下に示すさまざまな ETL タスクを実行できます。

- レガシーデータ(メインフレーム、Teradata、Netezza)をクラウドプラットフォーム(Cloudera、Snowflake、Databricks、Microsoft Synapse など)に数分で接続
- 1つのインターフェイスでバッチデータとストリーミングデータのワークフローを視覚的に設計
- メタデータおよび分析のために複雑な COBOL およびメインフレームコピーブックを活用
- データ量とデータパイプラインが拡大しても、レポートとコンプライアンスのためにデータ系列に関する詳細なインサイトを提供
- Snowflake、Teradata、Vertica、Oracle、Tableau、NoSQL などの多種多様な分析データベースおよび BI プラットフォームにアクセス
- Cloudera、Databricks、Amazon EMR、Azure HDInsight などのクラスタ化されたコンピューティングプラットフォームの水平拡張性を活用し、ETL プロセスをスケールアウトしてテラバイトからペタバイト規模のデータセットを処理

より簡単・迅速にデータパイプラインを最新化するソリューション

ETL プロセスはビジネスのデータパイプラインに欠かせない要素です。しかし、従来の ETL ツールでは、今日の企業で基準となっているデータの増大し続ける量と種類に追い付くことができません。その結果、企業は貴重なリソースを手動調整で無駄にしたり、トランスフォーメーションをデータベースに押しつけたりしています。

Connect の特徴は、高速・柔軟性を実現する高パフォーマンスのセルフチューニング ETL 機能です。Connect ユーザーは、データ統合環境の開発とサポートについてコストと複雑性を大幅に低減できます。Connect を使用すると、コストのかかるデータベースステージングエリアや手動プロセスは不要になります。オンプレミス型でもクラウド型でも、スタンドアロンサーバーまたはクラスタ化されたコンピューティング環境でも、Connect は臨機応変にトランスフォーメーションを処理し、一貫して安定したパフォーマンスを発揮するためにデータブレンドを作成できます。

©2022 Precisely. 本書で使用されている会社名およびプロダクト名は、各社の商標である可能性があります。

影響を最大限に高めるために改善されたパフォーマンス

Connect は、実行時にデータ構造およびシステム属性に基づいて最も効率的なアルゴリズムを動的に選択するセルフチューニングエンジンを備えた唯一のデータ統合ソリューションです。ジョブをオンプレミスで実行する場合でもクラウドで実行する場合でも、可能な限り最高のパフォーマンスを得ることができます。従来の ETL ツールに比べ、最大 10 倍の高速パフォーマンスを実現し、開発期間を数百時間も短縮します。

一度の設計であらゆる場所に展開

Connect を使用すると、一度ジョブを視覚的に設計してしまえば、オンプレミスでもクラウドでも、Spark、MapReduce、Linux、Unix、Windows など、どんな環境にでも展開できます。Connect のインテリジェント実行テクノロジーを使用すると、データ統合フレームワークを拡張し、将来にわたっても使用可能な状態を保証できます。ジョブを作成し直したり、新しいスキルを身に付けたりする必要はありません。Connect は、高パフォーマンスの ETL、SQL 移行/データベース ELT オフロードおよびレガシーデータの移動用にテンプレートに基づいた設計を提供します。



データベース全体のクラウドへの移行

クラウド展開の利益を最大限に活用するためには、最初にデータをクラウドに移行する必要があります。従来の ETL ツールを使用すると、データベース内で数百または数千ものテーブルのインポートジョブを定義するのに数百時間もかかってしまいます。Connect なら、データベースからスキーマ全体を数分単位でポイント、クリックし、配布準備できます。Connect を利用すると、次のことが可能です。

- Db2、Oracle、SQL Server、Teradata、Netezzaなどのソースからデータをインポート
- Amazon Redshift、Amazon S3、Hive などのクラウドターゲットにクラウド上で書き込み
- Amazon Redshift から SQL Server や Postgres などのローカルデータベースに数百ものテーブルをコピー
- クラウド上で Hadoop 内の Hive テーブルを自動作成
- 必要なデータしか移動しないようにテーブル、コラム、データ型をフィルタで絞り込み
- 転送中にデータを変換・クレンジングできるので、データを保存する必要がない

サポートされるソースとターゲット

- | | |
|-----------------------------|--|
| • Amazon Redshift | • メインフレーム (固定長ファイル、シーケンシャルファイル、VSAM ファイル)* |
| • Amazon S3 | |
| • Amazon Web Services (AWS) | |
| • Cloudera | • Microsoft Azure |
| • Databricks | • MPP |
| • Db2 | • Netezza |
| • フラットファイル | • NoSQL |
| • HDFS | • Oracle |
| • Hive | • Postgres |
| • Impala | • RDBMS |
| • JSON* | • Snowflake |
| • Kafka | • SQL Server |
| • Kudu | • Teradata |
| | • XML |

*ソースのみ

