Connect: ETL from Cluster to Cloud

High performance and scalable data integration for all enterprise platforms

Modernize ETL, integrate ALL data across the organization

Connect integrates all data across an organization from RDBMS, mainframe, NoSQL, the cloud, and more. Extract, sort, blend, prepare, transform, and load refined datasets to all major enterprise data warehouses and cloud data warehouses for BI and advanced analytics projects. Perform a variety of ETL tasks with Connect such as:

- Connect legacy data – mainframe, IBM i, Teradata, Netezza – to cloud platforms such as Cloudera, Snowflake, Databricks, Microsoft Synapse, and more in minutes
- Visually design workflows for both batch and streaming data in once interface
- Leverage complex COBOL and mainframe copy books for metadata and interpretation
- Get full insight into your data lineage for reporting and compliance even as data volumes and your data pipelines grow
- Access a wide variety of analytics databases and BI platforms including Snowflake, Teradata, Vertica, Oracle, Tableau, NoSQL
- Leverage the horizontal scalability of clustered compute platforms like Cloudera, Databricks, Amazon EMR, and Azure HDInsight to scale out ETL process to handle terabytes to petabyte-scale datasets

A faster, simpler way to modernize data pipelines

ETL processes are an integral part of any business's data pipeline. However, traditional ETL tools cannot keep up with the growing volume and variety of data that is the norm in today's enterprise. As a result, organizations are wasting valuable resources on manual tuning or pushing transformations down to the database.

Connect features a high-performance, self-tuning ETL functionality that delivers speed and flexibility. Users of Connect experience significantly lower cost and complexity for developing and supporting data integration environments. Connect removes the need for costly database staging areas or manual processes. Whether on-premises or in the cloud, on a standalone server or clustered compute environment, Connect processes transformations on the fly and enables you to create data blends for consistent, sustainable performance.

Improved performance for maximum impact

Connect is the only data integration solution with a self-tuning engine that dynamically selects the most efficient algorithms based on the data structures and systems attributes at run-time. Get the best performance possible whether the job is on premises or in the cloud. Experience up to 10X faster performance and save hundreds of hours of development in comparison to traditional ETL tools.

Design once, deploy anywhere

With Connect visually design your jobs once and deploy them anywhere – Spark, MapReduce, Linux, Unix, Windows – on premise or in the Cloud. Use Connect's Intelligent Execution technology to scale and future-proof your data integration frameworks without rewriting your jobs or requiring new skills. Connect offers a template-based design for high-performance ETL, SQL migration/Database ELT offload and legacy data movement.

Move entire database schemas to the cloud

To get the most out of your cloud deployment, the first thing you must do is get all your data into the cloud. Defining import jobs for the hundreds or even thousands of tables in your databases can take hundreds of hours with traditional ETL tools. Connect lets you point, click and onboard entire schemas from a database in minutes. With Connect you can:

- Import data from sources like Db2, Oracle, SQL Server, Teradata, Netezza ...
- Write to cloud targets like Amazon Redshift, Amazon S3, or Hive on the cloud
- Copy down hundreds of tables from Amazon Redshift to a local database like SQL Server or Postgres
- Automatically create Hive tables in Hadoop on the cloud
- Filter tables, columns, or data types so you only move the data you need
- Transform and cleanse data in-flight, without the need to stage data

Supported Sources and Targets

- Amazon Redshift
- Amazon S3
- Amazon Web Services (AWS)
- Cloudera
- Databricks
- Db2
- Flat files
- HDFS
- Hive
- IBM i
- Impala
- JSON*
- Kafka
- Kudu
- Mainframe (fixed length files, sequential files, VSAM files)*
- Microsoft Azure
- MPP
- Netezza
- NoSQL
- Oracle
- Postgres
- RDBMS
- Snowflake
- SQL Server
- Teradata
- XML

*Source only