![syncsort]

Connect for Big Data

High Performance Data Integration Optimized for Big Data Frameworks

A Faster, Simpler Way to Build Modern Data Pipelines

Connect for Big Data is high-performance data integration software specifically designed for building modern data pipelines with highly-scalable distributed compute frameworks like Spark and MapReduce.

With a single software environment for accessing and integrating all your enterprise data - mainframe to cloud, batch and streaming -- Connect for Big Data lets you quickly and efficiently turn raw data into the insights that drive your business.

Its game-changing architecture gives you the flexibility to design your jobs once and deploy them across Spark, MapReduce and single servers -- in the cloud and on premise -- without any coding or tuning. This architecture, plus native integration with Spark and MapReduce, enables Connect for Big Data to quickly evolve with the Hadoop ecosystem, so you can stay current on the latest technologies without rewriting your jobs or acquiring new skills.

Finally, you can feel confident processing your most critical data with Connect for Big Data, as it packages the comprehensive support you need to manage, secure and govern the entire process.

Access All Enterprise Data

Simple, high volume data access and ingestion to Big Data repositories. Collect application data from every source, regardless of complexity.

Don't miss out on vital insights by ignoring complex data sources such as the mainframe, or newer sources like cloud or streaming platforms. Use Connect for Big Data to connect securely to virtually any source using native drivers to extract data for guaranteed optimal speed and efficiency.

- Onboard data from virtually any source, including:
  - Microsoft Azure
  - Amazon Web Services (AWS)
  - Snowflake
  - Kafka
  - Mainframe
  - IBM i
  - MPP
  - Flat files
  - RDBMS
  - XML
  - JSON
  - NoSQL
  - HDFS
  - Impala
  - Kudu
  - Hive

- Move hundreds or thousands of tables – including whole database schemas – into your data lake at once, with the press of a button
- Access both batch & streaming from the same interface
- Access, re-format and load data directly into Avro & Parquet, with no staging required
- Load more data into Hadoop in less time. Let Connect for Big Data dynamically split the data and load it to HDFS in parallel

Intelligent Execution allows users to design sophisticated data transformations, focusing solely on business rules, not on the underlying platform or execution framework.

- Visually design your jobs once, and deploy them anywhere – Spark, Hadoop, Linux, Unix, Windows – on premise or in the cloud. No changes or tuning required.

- Easily move applications from standalone server environments and from MapReduce to Spark – as easy as clicking on a dropdown menu

- Future-proof applications for emerging compute frameworks

Integrate and Achieve the Fastest Path from Raw Data to Insights

Design Streaming & Batch processes in a single interface

Using the Syncsort Connect graphical interface, you can easily design workflows that include both batch and streaming data sources. Quickly cleanse, blend and transform this data to give it context and meaning so your organization can execute fast.

- Integrate data at rest and data in motion

- Enrich your data on-the-fly at lightning speeds before loading into Hadoop

- High-performance connectivity to Big Data and NoSQL databases, such as Cassandra, HBase & MongoDB

- The fastest parallel loads to Amazon Redshift, Greenplum, Netezza, Oracle, Teradata and Vertica

- Create Tableau and Qlikview files with one click

Use Connect for Big Data's graphical interface to subscribe, transform and enrich data from real-time Kafka queues. Connect for Big Data can also publish these enriched datasets to Kafka, simplifying the creation of realtime analytical applications by cleansing, pre-processing and transforming data in motion.

Comply with Security, Management and Governance Protocols

Secure and govern with seamless Hadoop integration – Improve metadata management and track data lineage

Connect for Big Data enables secure access to critical enterprise data and metadata lineage across platforms. due to Syncsort's ongoing contributions to Apache™ Hadoop projects, Connect for Big Data is natively integrated into the Hadoop data pipeline, providing interoperability and scalability. Connect for Big Data offers:

Management: Full integration with Cloudera Manager and Apache™ Ambari for monitoring, maintenance and deployment across hundreds of nodes

Security:

- Native LDAP and Kerberos support

- Apache™ Ranger and Apache™ Sentry certified

- Secure mainframe data access through FTPS and Connect:Direct

Governance:

- Cloudera Navigator certified

- Tight integration with HCatalog for metadata management and data lineage

- Work directly with mainframe data in its native format – preserving data lineage across platforms

The Best Solution for Your Most Complex Mainframe Data

Syncsort combines cutting-edge technology and decades of experience with both mainframe and Big Data platforms to offer the best solution for accessing and integrating mainframe data with Hadoop. Syncsort provides the experience so you don't have to.

Get mainframe data into Hadoop - in a mainframe format - and work with it like any other data source! Integrate mainframe data with emerging data sources!

- Preserve your data exactly as it was on the mainframe to meet governance & compliance mandates

- Enable non-mainframe developers to work with native mainframe data on the cluster

Integrate mainframe data with emerging data sources

- Directly access and understand VSAM files, mainframe fixed & variable files, and DB2 data

- Give your data meaning with COBOL Copybooks mapped directly to the mainframe data

- Stop wasting weeks of development time just to understand the data

Simplify Automate and Optimize your Data Pipeline

Single interface to design all your jobs once and deploy them anywhere, with no coding or tuning

Connect for Big Data delivers the flexibility and optimization you need when creating data pipelines. It is a single tool that supports all platforms in your modern data architecture -- across cloud, multi-cloud, on-premise or hybrid environments.

Connect for Big Data is uniquely built with Intelligent Execution, which which enables you to execute the same data integration jobs on Spark, MapReduce, EMR, Linux or Windows, without making any changes.

Because there are no required code changes when deploying on different frameworks, users can design sophisticated data transformations focused solely on business rules, without worrying about the underlying platform or execution framework—and without needing a new set of skills. This unique architecture insulates users from changes in the underlying technologies, "future-proofing" the process of collecting, blending, transforming, and distributing data. The user enjoys a consistent experience while still taking advantage of the powerful native performance of the evolving compute frameworks.

syncsort