

DATA SHEET | DataMover for Hadoop

A highly efficient and easy to use data movement tool for Hadoop

There is no easy, efficient and inexpensive way to import and export data for Hadoop today. Middleware ETL platforms are not optimal as they are complicated, expensive, inefficient and require a significant hardware and software investment. Pig and Sqoop scripts introduce complexity, development and maintenance issues, and require specialized skills.

DataMover

Diyotta DataMover is an intuitive and easy to use tool built to develop, monitor and schedule data movement jobs to load into Hadoop from disparate data sources including RDBMS, flatfiles, mainframes and other Hadoop instances. DataMover enables users to graphically design data import or export jobs by identifying source objects and then schedule them for execution.

Diyotta DataMover eliminates the need to develop scripts by providing out-of-the-box functionality to perform data movement tasks. Ideal for both technology teams and non-technology teams including data scientists, business analysts and power users.

Diyotta DataMover is purpose-built for big data platforms with a lightweight, metadata-driven and easy to use development environment. Our integration with big data platforms provides direct data movement from source to target systems, eliminating the need for intermediate ETL servers, significantly reducing costs and providing a high degree of scalability. The solution deploys in minutes and users can immediately begin designing data movement jobs to move data in and out of any Hadoop distribution.

“Diyotta was very easy to deploy. We were operational in less than a day and developing workflows 10 times faster than hand-crafted code.”

- IT Project Manager, System Integrator

Move Data into or out of Hadoop in 3 Easy Steps

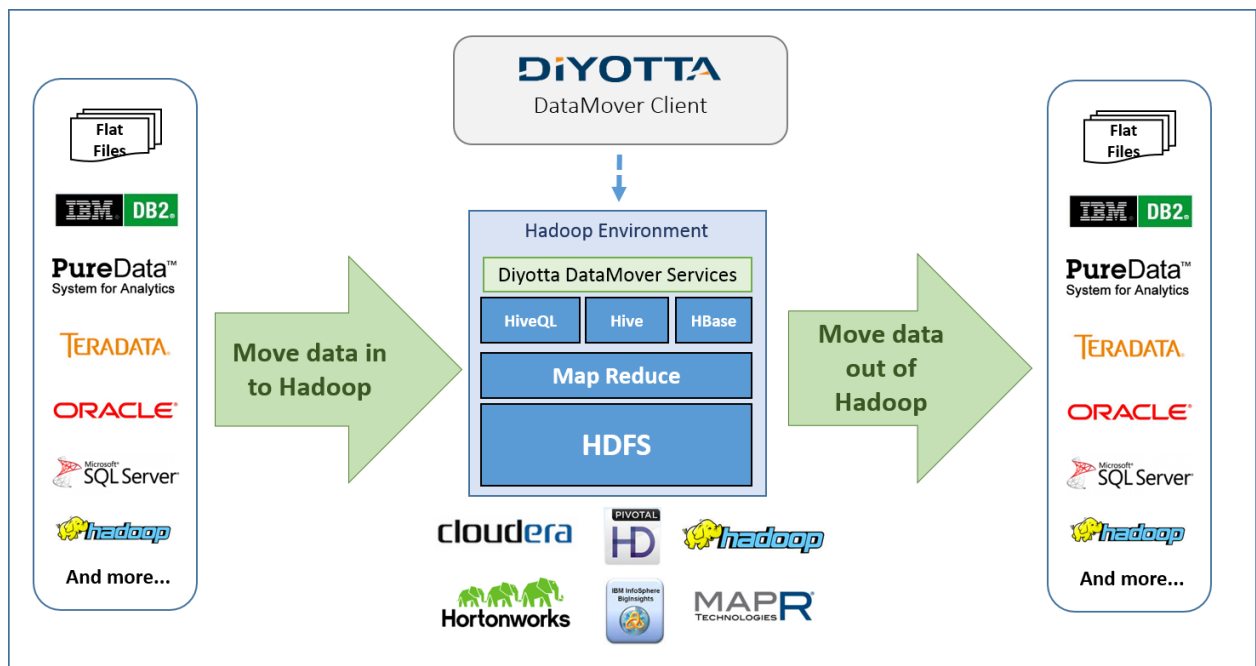
In an effort to simplify data load and extract processes for Hadoop, Diyotta provides a web-based interface where users can build import or export jobs using 3 simple steps:

- **Acquire:** Identify the source from which the data needs to be extracted. DataMover allows users to readily connect to these sources and define extraction rules like filter criteria or derived columns.
- **Ingest:** Mapping columns between the source and target objects. Diyotta even creates target objects for you if they do not exist.
- **Execute:** Data movement jobs can be immediately executed or scheduled for later execution.

Easy and Efficient

- The GUI is simple to navigate and users can create data movement jobs for full or incremental data extraction within minutes and then immediately execute or schedule these jobs for future execution. The extraction rules allows users to insert incremental logic and other rules to extract data to meet specific requirements.
- Diyotta DataMover has a lightweight footprint requiring minimal system resources and is installed directly on the name node or any data node within the Hadoop system.
- Dashboard, operational statistics and multi-level logs provide monitoring capabilities of job execution and help to identify and troubleshoot any errors.

DIYOTTA DATA MOVER LOGICAL ARCHITECTURE



Benefits:

- Rapid development of data movement jobs for Hadoop environment
- High-speed extraction, transfer and ingestion of large volumes of data
- Significantly reduces costs and eliminates the complexities of middleware ETL tools
- Eliminates the need to develop custom code for data movement in Hadoop environment
- Operational statistics and monitoring features for data movement jobs

Diyotta DataMover makes it easy to move data between Hadoop and a variety of sources including flat files, streams, EBCDIC files, Netezza, IBM DB2, Oracle, Teradata, SQL Server and other databases. Supported Hadoop distributions include Cloudera, Hortonworks, IBM BigInsights, Pivotal HD and MapR.

DIYOTTA
Frictionless Data Integration