Dataflow
Data Sheet

DATAGAPS
THE GREAT DATA COMPANY

"Next generation ETL/ELT tool to move large datasets across systems using Apache Spark."

DataOps Dataflow
Free ETL/ELT Tool for SMBs

Dataflow is a module within DataOps Suite designed to efficiently handle next generation ETL or ELT processes. It complements ETL Validator and other modules within the DataOps Suite to provide one unified platform to streamline data intensive operations.

Built on patented architecture, it leverages Apache Spark as the underlying engine to move data from source to target systems. By leveraging simple UIs and wizard-based capabilities, Data Engineers, Testers & Business Analysts can streamline data intensive operations. Management can save tremendous amount of costs by leveraging Dataflow as it is open source and sufficient for most of the data migration projects.

SQL : Is the only language required to automate flows using Dataflow. This brings down the total cost of ownership as SQL knowledge is a common skill in Engineering, Testing and Business Analysis teams. If you choose to leverage Python or Scala, then, Dataflow has the provision to support your needs.

Scheduling : At the scheduled time, Dataflow automatically executes the tests and emails summary reports to the right stakeholders.

Benefits

Simple to Use	Efficient Data Movement	Zero to Minimal costs dependent on the use cases	Seamless integration with other modules within the DataOps Suite

Features

Flat Files Ingestion

Organizations across industries receive flat files from vendors, partners, customers, or other stakeholders. Dataflow supports ingestion of flat files and moving the data into target systems with ease. In addition, rules can be defined to validate the accuracy and quality of data being ingested into the enterprise.
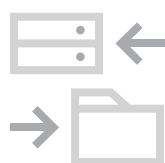
Big Data Ingestion

In addition to traditional sources, data formats of data can be ingested from big data sources such as Hadoop and synchronized into data lakes or other database systems.

Component based Development

Dataflow has in-built components that make it easy for Data Engineers to execute ETL or ELT related jobs. Popular ones include Source Components, Processor Components and Sink Components.

Data Migration Wizard

The Data Migration Wizard in Data Dataflow simplifies the process of migrating a large number of tables to the target system.

Optional Programming

SQL is the only language required to automate test cases using Dataflow. This brings down the total cost of ownership as SQL knowledge is a common skill in Engineering, Testing and Business Analysis teams. If you choose to leverage Python or Scala, then, Dataflow has the provision to support your needs.

Container-Based Security

Dataflow provides container-based security for managing access to test data, test cases and test results. This feature facilitates the restriction of data and features to select users.

CI/CD Enablement

Dataflows and Data Pipelines can be version controlled to GIT repositories and invoked from CI/CD orchestration tools such as Jenkins.

Scheduling Capabilities

Dataflow provides multiple options to assemble and schedule runs. At the scheduled time, DataOps suite Dataflow automatically executes the Dataflows and emails summary reports to the right stakeholders.

Data Pipelines

Data Pipeline provides a DAG interface for grouping multiple Dataflow tasks together with dependencies and executing them in parallel.

Enterprise Collaboration

All Dataflows are stored in Dataflow's enterprise repository. This enables users to reuse Dataflows across the organization.

Other Features

• Data Analysis
• Multiple Connectors
• Open Source Platform for ELT/ELT

• Data Mapping
• Data Integration
• Scheduling Notification

• Data Transformation
• Data Synchronization
• Simple Interface Visual

SUPPORTED DATA SOURCES

ORACLE

Microsoft® SQL Server®

IBM DB2

MySQL®

teradata.

N NETEZZA

VERTICA

GREENPLUM DATABASE

PostgreSQL

SYBASE®

hadoop

X

snowflake

salesforce

APACHE DRILL

Email us at **contact@datagaps.com** for a price quote, pilot or to know more about how we can help you solve your enterprise data testing issues.