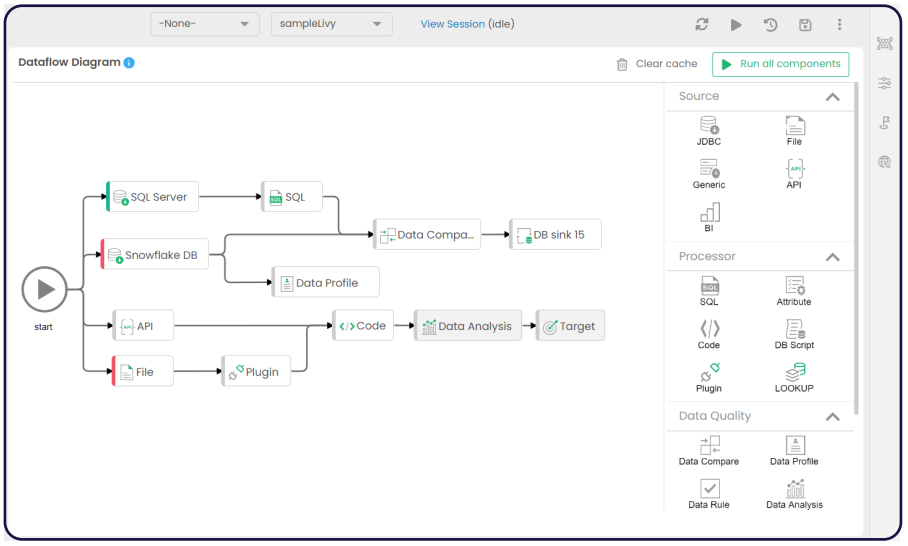
DATAGAPS
THE GREAT DATA COMPANY

ETL Validation Data Sheet

The most comprehensive ETL Validation Tool

"**Datagaps**' ETL Validator has improved the data quality at every stage in our pipeline and reduced the downtime at migrations, data structure updates and daily operations" - Data Architect at NYU

ETL Validator, a module within DataOps Suite to efficiently handle and validate Big Data ETL/ELT Pipelines. Built on patented architecture, it leverages Apache Spark as the underlying engine. Data Engineers, Tests, Analysts can easily and quickly be deployed into data intensive operations. With in-product as well as integrated reporting and notification systems, management can save tremendous amount operational and overhead costs.



60% Reduction in Migration Testing Time

35% Reduction in Migration Time

40% Reduction in Data Quality Testing Time

30% Reduction in Total Cost of Operation

Popular Supported Sources and Integration

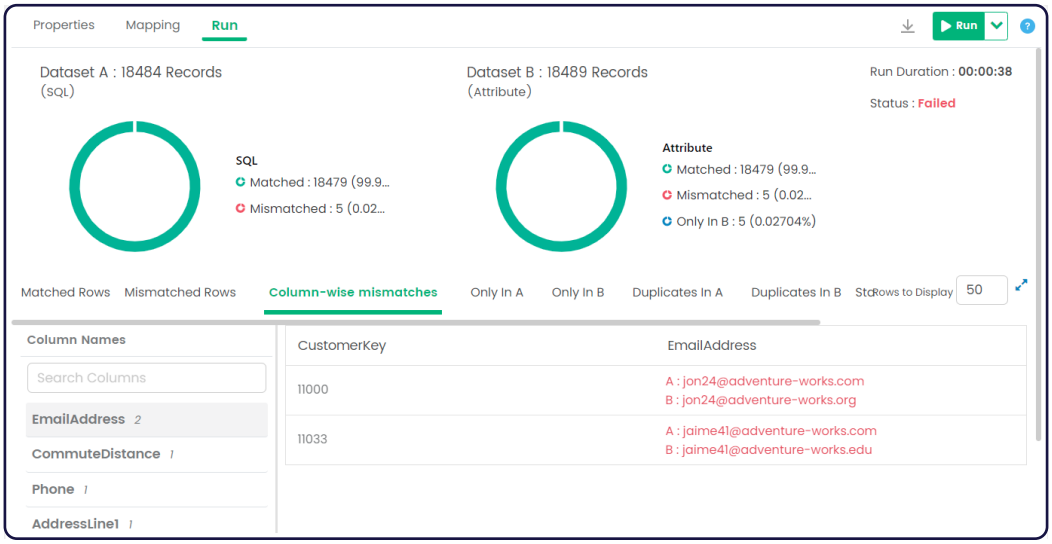
Feature List

- Zero Code Deployment
- Container-Based Security
- Enterprise Collaboration
- SQL, Python, Scala and Plugin Support
- Automated Reporting System
- Data Compare Node
- Metrics Comparision Node
- Metadata Comparision and Validation Nodes
- Data Migration Wizard
- Scheduling Capabilities
- API, Flat File and BI Connector
- Email and Slack Notification
- CI/CD Enablement
- Data Profiling Node
- Data Rules Node
- Data Pipelines and Conditional Processing

Data Compare

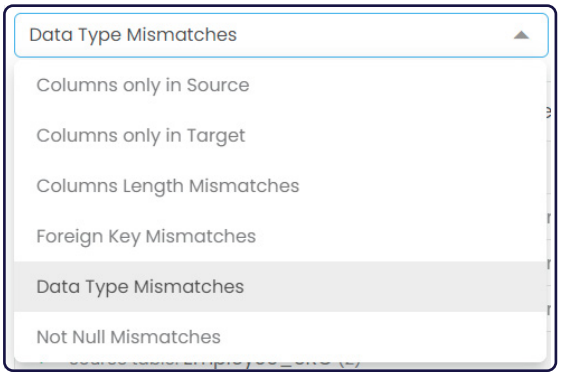
The Data Compare Node is an easy to deploy dataset / transformation comparison node that is setup in less than 5 clicks. With column-wise mismatches and thresholding parameters, every use-case of comparison can be validated easily.

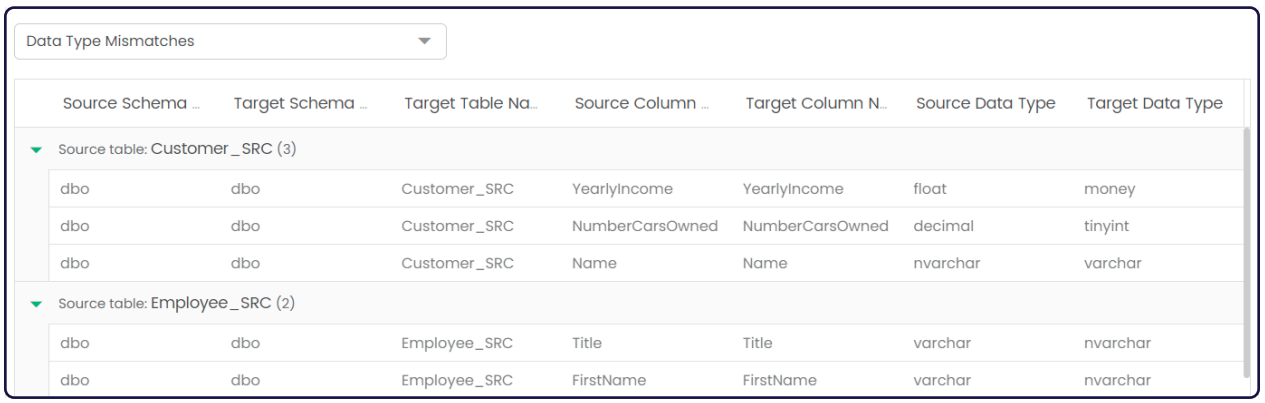
The suite comes with a Migration Wizard Tool that automates the creation of comparison test cases for easy migration testing.



Easy Metadata Compare

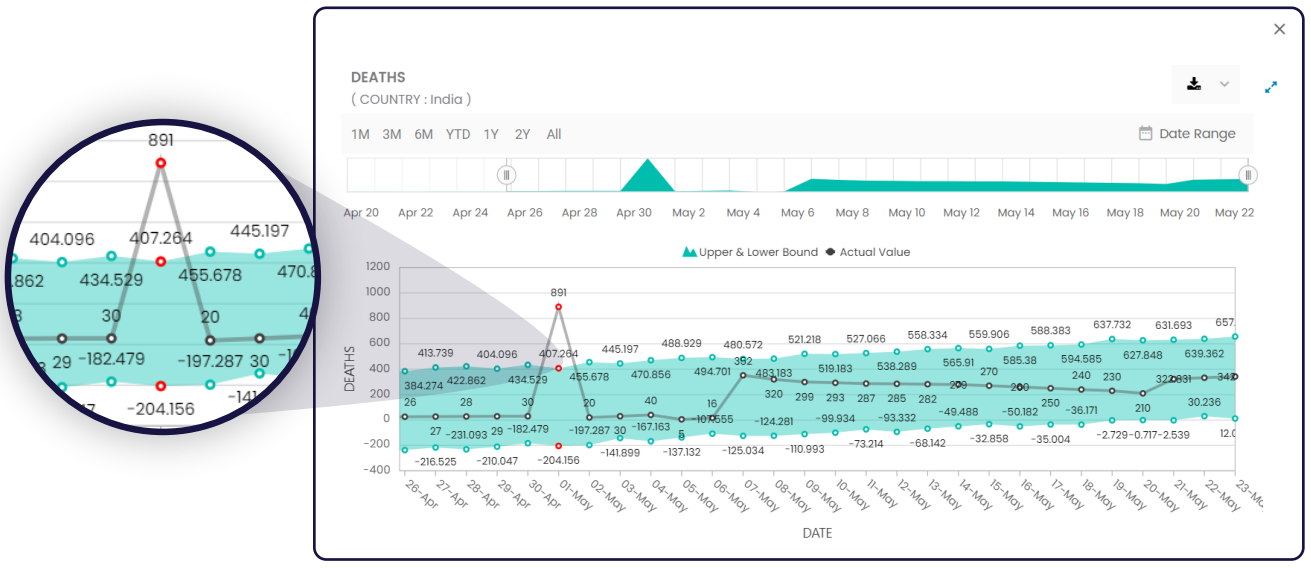
With a single node, 2 compete schemas' metadata can be compared instantly with detailed reporting on lengths, type, foreign key mismatch, non-null and exclusive column difference.





Anomaly Detection

Using ML and Prediction techniques, metrics can be tracked, and users can be notified in case of statistically significant anomalies are found the dataset.



Profiling

Within ETL Validator, a Profiling node can be applied to a dataset to track various aggregates, frequency and patterns within the data.

The aggregations help point out anomalies such as sharp changes in mean, lengths, min-max, decimal points and such. Frequency Analysis is used to keep track of the nature of distribution in the datasets. Pattern analysis keep track of the string patterns and # of digits in a particular dataset.

Aggregate	Profile Value	Average (from previous batch runs)	Trend
Row Count	50	50	
Null(%)	0.0	0	N/A
Null count	0	0	N/A
Distinct(%)	100.0	100	
Distinct count	50	50	
Constancy	0.02	0.02	
Min Value	N/A	N/A	N/A

Frequency Analysis

Value	Count
AB89294	1 *(2%)*
AB64352	1 *(2%)*

Pattern Analysis

Pattern	Count
XX99999	50 *(100%)*