

DataOps Tools :

1. Genie

Developed by Netflix, the DataOps tool is an open-source engine that offers distributed job orchestration services. This tool provides RESTful APIs for developers who wish to run a wide range of jobs with Big Data, such as Hive, Hadoop, Presto, and Spark. Genie also provides APIs for metadata management in distributed processing clusters.

2. Piper

Piper is a package of Machine Learning based DataOps tools that enable organizations to read data more smoothly and efficiently. This solution exposes data through a set of APIs which integrate easily with digital assets of the organization.

3. Airflow

Apache Airflow is an open-source DataOps platform that manages complex workflows in any organization by considering data processes as DAG (Directed Acyclic Graphs). This tool was first designed by Airbnb to schedule and monitor their workflows.

4. Automated Testing Tools

The second category of DataOps tools covers automated testing. Simply put, automated testing tools test and compare the actual outcomes of a software technique, versus the expected outcome. These tests are applied to repetitive tasks to identify the best methods.

5. Naveego

Naveego is a cloud data integration platform that allows businesses to reach accurate business decisions by integrating all company data in a regular business-centric format. This tool cleans stored data and makes it analytics-ready for data scientists.

6. FirstEigen

FirstEigen is a platform including Machine Learning tools that provide big data quality validation and data matching on the basis of self-learning. This platform learns about data quality behaviors and models using advanced ML techniques and then tests big data with just three clicks.

7. Right data

Right data is a self-service group of applications designed for achieving data quality assurance, integrity audit and continuous control along with automated validation. This suite is best suited for organizations seeking tools with automated testing and reconciliation capabilities.

8. Badook

Badook is a popular tool among data scientists since it allows them to write automated tests for datasets used in training/testing data models. This tool not only allows them to validate data automatically but also reduces turnaround time for generating insights.

9. Lentiq

This data model deployment tool works in a service environment for smaller teams. With Lentiq, you can run data science and data analysis at the scale of your choice in the clouds so your team can ingest real-time data, process it, and share useful insights.