INFOSYS CLOUD DATA VALIDATION SOLUTION

Highly scalable cloud native automation solution designed to address modern day data validation challenges

Over the last few years enterprise data has undergone massive changes in terms of its size and source. As a result, in the current world, data is boundary-less, generated from infinite sources and increasing multifold. The technology world is experiencing a digital disruption with more and more customers opting for tailor made cloud applications to handle these massive amounts of data. In such a scenario, cloud native solutions are a game changing technology. With

traditional technology reaching a tipping point, customers across the world are migrating to cloud native data warehouses and analytics like Snowflake, Amazon Redshift and Google BigQuery. These migrations open an entirely new set of challenges in data testing as the existing traditional tools and data automation solutions have their limitations. This is where cloud native data automation solutions come into the picture to address these challenges.
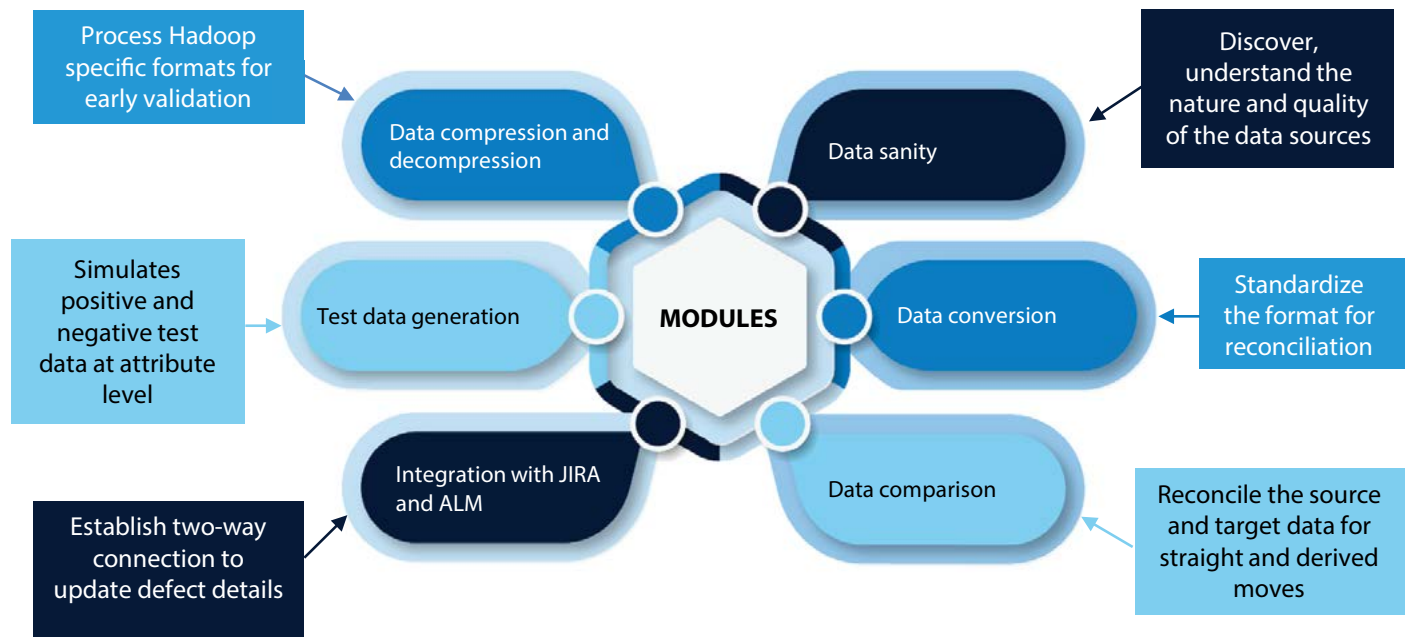
Key challenges

The QA challenges in a cloud migration journey can be summarized as below:

- Huge volumes of data and variety of data validation use cases
- Validation of legacy systems data while moving to cloud
- Lack of data testing solutions to support new technology
- Ensuring data integrity during movement across numerous technical layers
- Ensuring data accuracy between visualization layer and backend system

The Infosys solution for Cloud Data Validation

Infosys Cloud Data Validation Solution (ICDVS), part of Infosys Cobalt, is a best-in-class cloud native data automation solution, currently available both on AWS and Azure marketplaces, which helps streamline and accelerate testing of data integration / data analytics platforms. The solution uses a UI-based automated data quality and data integrity approach for accelerated validation. By bringing open source technologies and Azure native services together, the solution provides comprehensive technical approach for data validation of source and target database/files. These include data quality checks (duplicates records, missing records), data reconciliation checks, and data validation checks (duplicates records, incorrect pattern, missing records).

The figure below shows core modules of ICDVS along with high level features.

Process Hadoop specific formats for early validation

Data compression and decompression

Data sanity

Discover, understand the nature and quality of the data sources

Simulates positive and negative test data at attribute level

Test data generation

MODULES

Data conversion

Standardize the format for reconciliation

Establish two-way connection to update defect details

Integration with JIRA and ALM

Data comparison

Reconcile the source and target data for straight and derived moves

Solution features

Performs exhaustive validation of data	Ensures data quality	Enables data standardization	Integration of result reporting	Efficient execution of scripts
Ensures that data movement between any source to target is 100% validated. Supports File vs File, File vs DB, and DB vs DB validation.	Understands the nature and quality of the data sources. Performs data sanity test as a first step to ensure no bad data is being used for further processing which might lead to incomplete results. Scans and screens inbound files and tables for any quality issues.	Standardizes the format for reconciliation of data. Converts non-readable formats (AVRO, Parquet) and semi-structured formats (XML, JSON) to .csv to facilitate reconciliation. Objective is to do early validation to ensure files are not corrupted before being processed further.	Establishes two-way connection to update defect details. Can connect to most of the defect management tools like Jira, ALM etc. At the same time, result reports can be customized to the needs of the project.	Executes multiple scripts in a serial manner by utilizing the batch mode feature

Why choose Infosys?

Validation of Data analytics platforms can make or break the data journey of any organization. Infosys has created this cloud native automation solution to ensure quality data at the right time for both strategic planning and day to day operations. The solution covers key elements across the ecosystem of QA environments helping clients gain significant cost and time savings through:

- Robust automated solutions and continuous process improvements
- Improved productivity and overall QA efficiencies through a cost-effective model

Benefits

Infosys Cloud Data Validation Solution ensures data quality throughout ETL and Big Data ecosystem:

- Reduces overall effort by 25%
- Ensures 100% data testing coverage
- Accelerated test execution – 50% cycle time reduction
- Improves productivity by 40% by supporting concurrent testing

Success Story

For a US based private mutual fund company, Infosys was involved in the data lake migration to AWS. Real time ingestion of 8500+ tables from varied sources and transforming them into complex nested structured subject areas made the test scenarios complicated. Infosys solution helped validate millions of records in few minutes. It performed fully automated data comparison resulting in up to 50% savings in overall testing effort and reduced time-to-market up to 25%.

Discover how Infosys can help you automate data validation leading to reduced testing life cycle. Connect with us at IVS_Marketing@infosys.com

Infosys®
Navigate your next

For more information, contact askus@infosys.com

Infosys.com | NYSE: INFY

Stay Connected