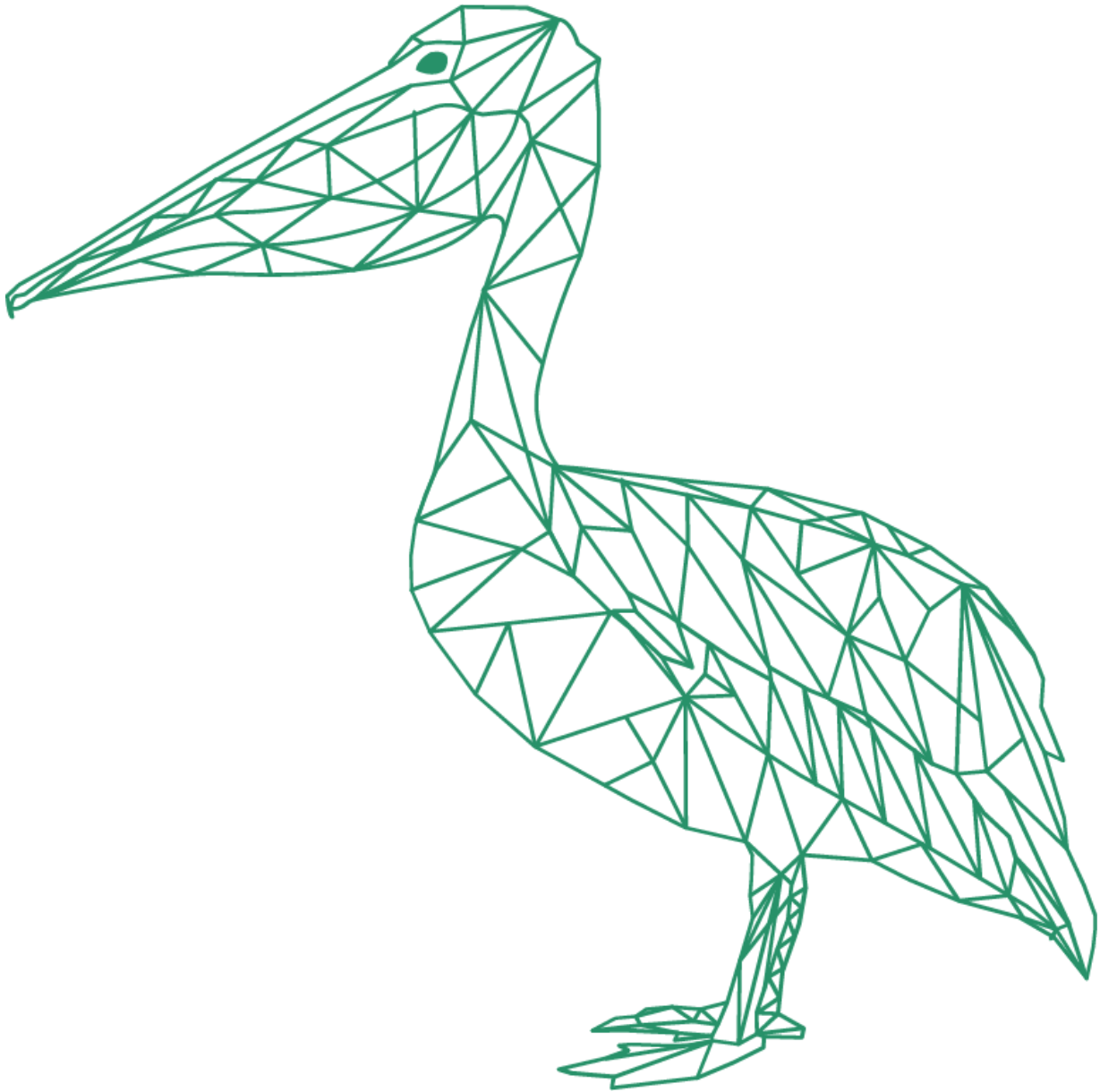# Pelican Tool Usage Guide

Hello, and welcome to an overview of Pelican, an automated validation and reconciliation tool powered by AI developed by Datametica.

Pelican is designed to work in parallel with data migration processes. It enables simultaneous data validation while the migration is in progress, eliminating the need to wait for migration completion before initiating the validation phase. It is an innovative enterprise-ready product that accurately checks/validates data between different data stores with zero data movement and coding. Pelican is equipped with distinctive functionalities that include cell-level comparison, zero coding requirement, zero data movement, bulk mapping, and Lineage tracking. These features enable Pelican to validate peta-byte scale of data with 100% accuracy and high data security; while saving cost and time.

Before proceeding further, let us undertake a comprehensive examination of Pelican Architecture.
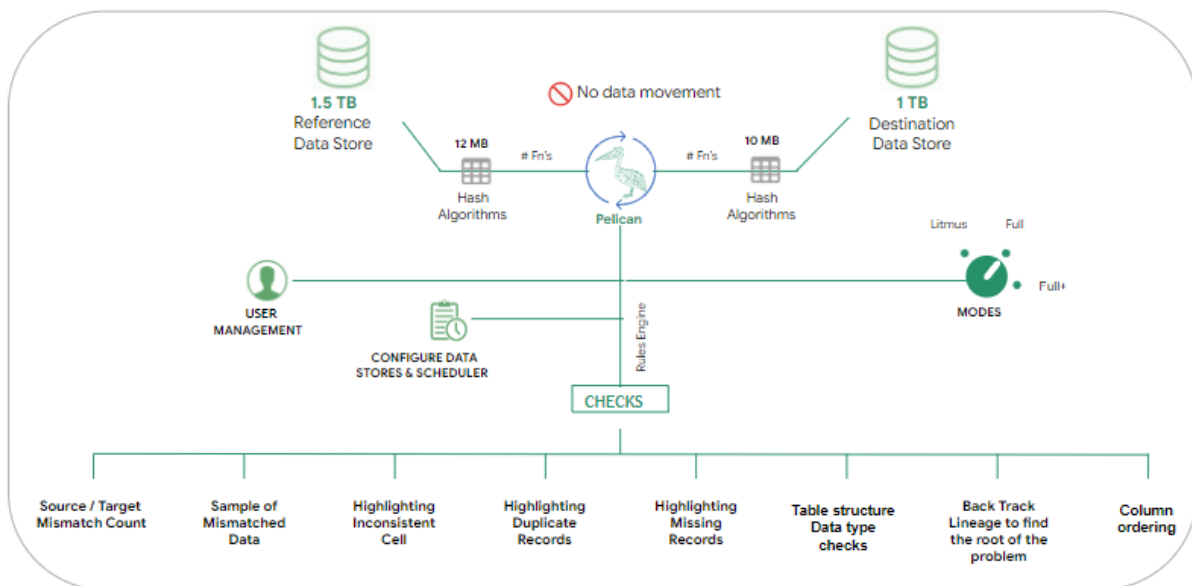
This architecture diagram illustrates the interaction between PELICAN and both the source and target systems, showcasing its extensive capabilities. PELICAN employs a meticulous approach to data validation, performing cell-level verification through the generation of unique fingerprints. Utilizing built-in algorithms, these fingerprints are compared without the need for data movement across the network. Furthermore, PELICAN is equipped with a rule engine that

facilitates data validation, incorporating both default and custom expressions.

PELICAN offers a user-friendly interface to configure organizational LDAP/AD settings, enabling seamless authentication using LDAP/AD credentials. Within its framework, PELICAN incorporates a wide range of validation checks, including source/target mismatch count, highlighting of mismatched cells, presentation of sample records with discrepancies, identification of missing and duplicate records, and verification of column ordering, among others.

Now, let us delve into the Pelican validation and reconciliation process in detail.

The Pelican validation process encompasses three crucial steps:

1. **Source/target Connectivity:** This initial step involves establishing a secure and reliable connection between Pelican and the source/target systems also called datastores. Pelican establishes connectivity with the source and target datastores using JDBC. To create a datastore please navigate to Administration->Configure->Datastores. Fill in all the details and a successful test connection confirms the connectivity.

2. **Mapping Creation:** The creation of mappings forms a pivotal part of the Pelican validation process. Mappings link the source and target tables. Pelican also enables the user to create hundreds of mappings simultaneously with the help of the bulk mapping feature. The Bulk mapping feature permits you to create a new mapping and also edit the existing expressions, filters and columns. To use the Bulk Upload feature enter details into the appropriate CSV file and upload it by clicking on the 'Click to upload' button. The sample CSV files can be downloaded by clicking on the Download button.

3. **Scheduler Creation & Execution:** Once the mappings are created the next step is to create and configure the schedulers. The scheduler executes the process and validates the source table with the target table.

The schedulers can be executed in three modes -

LITMUS: This mode offers core validation functionality. LITMUS mode is suitable for quick and simple validations and validates whether there is a complete cell level match. It creates a fingerprint of the complete source table and compares it to the fingerprint of the target table, and shows a pass/fail decision.

FULL: As the name suggests, this mode provides a comprehensive validation process. It thoroughly validates the tables by performing in-depth checks and validations to ensure its reliability and stability. As an initial step, it counts # of rows in Source and Target tables. It validates if there is a count match, extra rows on the target, missing rows on the source. It creates a fingerprint of the complete source table and compares it to the fingerprint of the target table, and shows a pass/fail decision with samples. FULL mode picks 1000 random mismatched rows and displays 10 samples in the validation output by default irrespective of the # of rows present in the Source and Target tables.

FULL+: This mode represents an enhanced version of the FULL mode which identifies the mismatches at a column level. By default this will scan 80% of the total columns identifying 1 mismatch per column and 1 million rows.

Pelican offers the capability to validate entire tables as well as perform incremental validation. Multiple schedulers can be executed with the help of the Group Scheduler feature.

4. **Report Viewing**: Pelican offers a robust reporting system that enables users to conveniently access and analyze validation outcomes. PELICAN incorporates various types of reports, as outlined below:

   - Scheduler Status Report
   - Sample Download Report
   - Validation Result Report (Table/View)
   - Validation Report (Detailed)

Pelican's reporting mechanism ensures that users have access to a wide range of reports tailored to meet their specific needs, thereby enhancing their ability to make informed decisions based on the analysis of validation outcomes.

Key Features:

1. **Lineage**: Lineage is an exceptional feature that helps to visualize the failures in the data pipeline. It helps to validate all tables from a data load pipeline in parallel and to traverse data load pipeline lineage to identify the genesis of the data error.

2. **Custom SQL Execution**: PELICAN supports execution of custom execution to create views on source and target datastores and use them for validation.

3. **User Management**: Pelican allows configuring users by using either of the below approaches:

   a. SMTP: The SMTP configuration(Email Configuration page) enables the superuser to configure SMTP server host details which will help to configure the local users to the PELICAN.

   b. LDAP and AD: The User Management(Authentication) screen allows the superuser to configure the LDAP/AD authentication details in PELICAN.
   Pelican facilitates effective User management by leveraging features such as User Group, Workspace and Roles.

4. **User Groups**: The User Group feature enables the administrator to give access rights (roles and workspaces) to all the users belonging to the same AD or LDAP group in one instance.
   a. Workspace - This feature enables the user to group the mappings and give access to the specific users to access those mappings. By default the mappings are grouped under PELICAN workspace. However a user can create custom workspaces.
   b. Roles - This feature comes with two default roles i.e. ADMINISTRATOR and VALIDATOR. However users can create custom roles and assign permissions to them as per the business needs. Superuser can then associate users to this custom role.
5. **Export/Import**: Export/Import feature is used to import/export data from one Pelican instance to another instance of the same version along with PELICAN metadata viz. table mappings and their schedulers.

Conclusion:

So in this video we have glanced upon Pelican and all its major functionalities. For further queries please write to us on PelicanCustomerSuccess@datametica.com