

Data Validator App - Documentation

Overview	2
Architecture	2
User Guide	2
Dashboard.....	2
Test Cases.....	2
Types of Test Case	2
Prerequisites.....	3
Steps to Add a Test Case	3
Executing a Test Case.....	3
Sample CRON Schedule	4
Viewing Test Results	4

Overview

The Data Validator is a Snowflake-based application designed to automate and standardize data quality validation between source and destination systems. It supports a range of rule-driven test cases to monitor data integrity, identify discrepancies, and ensure pipeline reliability.

- Supports custom test definitions and automated execution
- Captures and visualizes test results
- Enables both on-demand and scheduled test execution

Architecture

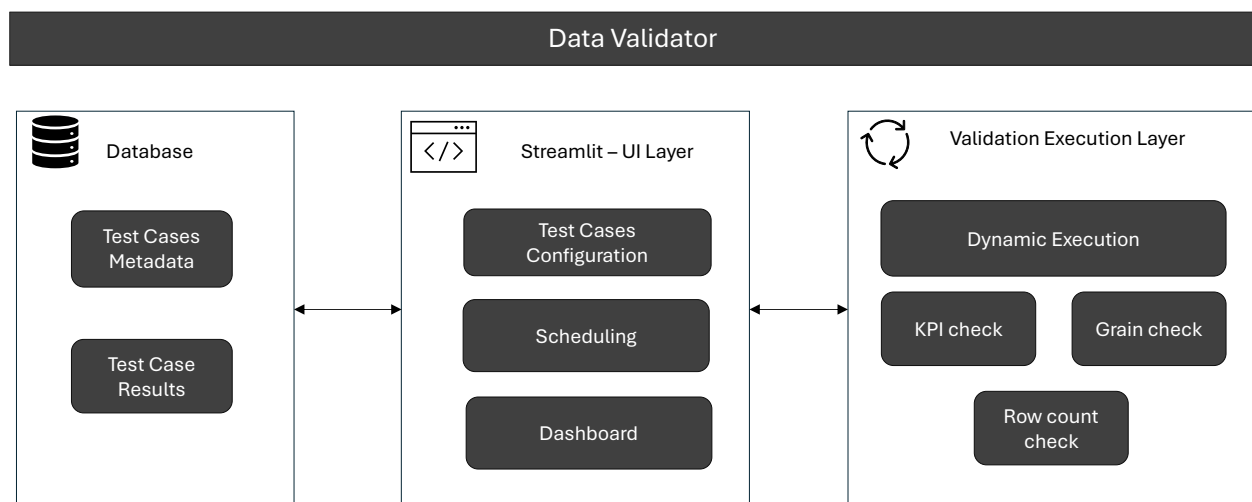


Image 1: Data Validator app architecture

User Guide

Dashboard

Display summary of:

- Total test cases created
- Recent run status by database

Test Cases

Types of Test Case

Test Type	Description
-----------	-------------

Row Count Check	Compares row count between source and destination. Passes within a $\pm 5\%$ margin
Grain Check	Validates uniqueness of primary keys in both tables
KPI Check	Compares aggregated/custom KPI values. Passes within a $\pm 5\%$ difference

Prerequisites

Ensure that the application has appropriate **SELECT** access before adding a test case:

```
GRANT USAGE ON DATABASE <DATABASE_NAME> TO APPLICATION <APP_NAME>;
GRANT USAGE ON SCHEMA <SCHEMA_NAME> TO APPLICATION <APP_NAME>;
GRANT SELECT ON TABLE <TABLE_NAME> TO APPLICATION <APP_NAME>;
```

Steps to Add a Test Case

1. Navigate to "Add Test Cases" from the sidebar
2. Choose a Test Type:
 - Grain Check
 - Row Count Check
 - KPI Check
3. Select the Source and Destination:
Provide Databases, Schemas, and Tables
4. Enter a Title for the test case
5. Additional Configuration:
 - Grain Check: Select columns that define grain on both sides. Test passes only if both maintain uniqueness
 - KPI Check: Choose between:
 - Custom SQL expression input
 - Columns with aggregation types (for simple metrics)
 - Row Count Check: Select source and destination tables

Executing a Test Case

Type	Steps
On-Demand Execution	<ul style="list-style-type: none"> • Go to the Execution Panel • Select the Test Case ID • Click Execute • View current and past results for the Test Case ID
Scheduled Execution	<ul style="list-style-type: none"> • Define a CRON schedule in the Scheduler Panel • Select: Minute, Hour, Day of Month, Month, Day of Week, Time Zone • Click Update Schedule

Sample CRON Schedule

Field	Value	Explanation
Minute	3	Job runs at the 3rd minute of the hour (i.e., hh:03)
Hour	1	Job runs at 1 AM
Day of Month	2,6	It runs on the 2nd and 6th day of the month
Month	4 (April)	April
Day of Week	WED	Runs on Wednesdays
Time Zone	Asia/Kolkata	CRON schedule is based on the Asia/Kolkata time zone

Result: Runs at 1:03 AM on April 2nd and 6th, and every Wednesday in April

Viewing Test Results

- Go to the Results Panel
- Select a Test Case ID
- View recent run details:
 - Execution status (Pass/Fail)
 - Execution time
 - Historical Trend to Monitor Data Consistency