Informatica
from Salesforce

28 July 2026

Ensuring Data Trust: Cloud Data Validation Strategies with Informatica IDMC

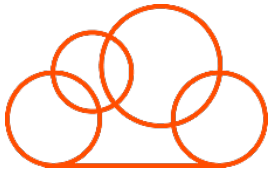
- Chinnari Damerla, Senior Success Guide
- Dihitha Sai Nadella, Success Guide

Last edited by Creative in February, 2026

Housekeeping Tips

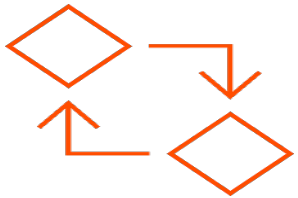
- Today's session is scheduled for 1 hour and will feature a presentation followed by a Q&A session at the end.
- To ensure a smooth presentation without interruptions, all participants will be muted.
- Please submit your questions to "All Panelists" using the Q&A panel. We look forward to answering them during the live Q&A.
- This session is being recorded. The recording and slide deck will be available on our Success Portal, and a recorded link will be emailed to you after the event.
- We value your input! Please take a moment to complete the post-webinar survey to share your feedback and suggest future topics.

Informatica from Salesforce



Bootstrap trial and POC Customers

Enriched Customer Onboarding experience



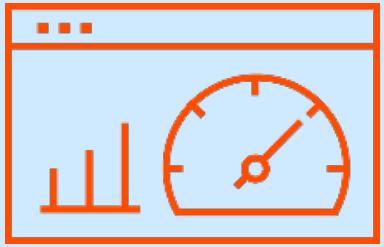
Product Learning Paths and Weekly Expert Sessions

Informatica Concierge

Tailored training and content recommendations

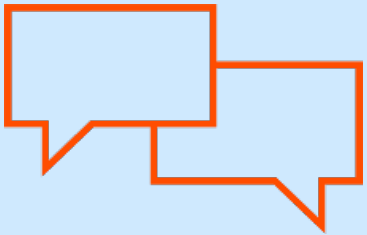
3

More Information



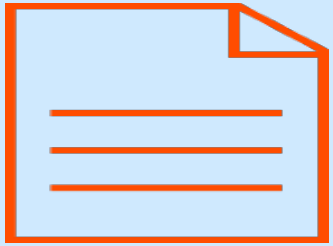
Success Portal

https://success.informatica.com



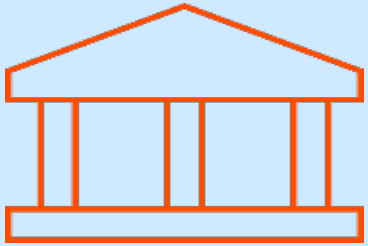
Communities & Support

https://network.informatica.com



Documentation

https://docs.informatica.com



University

https://www.informatica.com/in/services-and-training/informatica-university.html

Safe Harbor

Disclaimer: The information being provided herein is for informational purposes only. The development, release and timing of any Informatica product, service or functionality described herein remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made herein are based on information currently available, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products, services or functionality in the future.

Forward-looking statements

This presentation contains forward-looking statements about, among other things, trend analyses and statements regarding future events, anticipated growth and industry prospects, and our strategies, expectation or plans regarding product releases and enhancements. The achievement or success of the matters covered by such forward-looking statements involves risks, uncertainties and assumptions. If any such risks or uncertainties materialize or if any of the assumptions prove incorrect, results or outcomes could differ materially from those expressed or implied by these forward-looking statements. The risks and uncertainties referred to above include those factors discussed in Salesforce's reports filed from time to time with the Securities and Exchange Commission, including, but not limited to our ability to consummate the pending acquisition of Informatica on a timely basis or at all; our ability to meet the expectations of our customers; uncertainties regarding AI technologies and their integration into our product offerings; the effect of evolving domestic and foreign government regulations; regulatory developments and regulatory investigations involving us or affecting our industry; our ability to successfully introduce new services and product features, including related to AI and Agentforce; our ability to execute our business plans; the pace of change and innovation and our ability to compete in the markets in which we participate; and our ability to maintain and enhance our brands.

Agenda

01 Introduction

02 Data Validation Techniques

03 Test Cases & Test Suites

04 Business Outcomes

05 Demo

06 Q&A

Introduction

The Modernization Bottleneck

Moving on-premises critical applications and logic is complex, lengthy, and expensive. Traditional manual testing leads to severe bottlenecks:

01 Transcription Mistakes: Manual recreation of testing logic leads to human copy-paste errors.

02 Complexity Risks: On-premises legacy infrastructures are heavily entangled, making it difficult to detect mapping alterations.

03 Inconsistent Rules: Validation is applied differently across separate QA, Dev, and Prod teams.

THE CHALLENGE

74%

Of legacy modernization projects fail due to complex environments & data platform gaps.

TIMELINE IMPACT

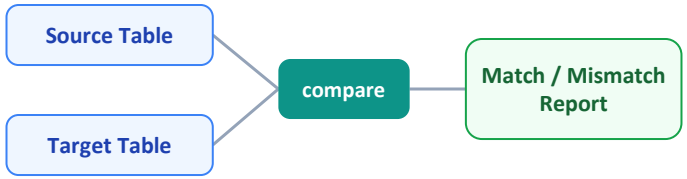
16 Months

Average duration of migration projects, heavily prolonged by manual data reconciliation.

Introduction

What is Data Validation?

Data Validation is a service in Informatica Intelligent Data Management Cloud (IDMC) that you can use to compare two data sets. It verifies the accuracy and completeness of data integration operations by comparing and validating source and target data.



fact_check

What It Does

- Validates data in source and target systems
- Compares source vs target row-by-row or aggregate
- Identifies bad records and generates exception reports
- Runs tests inline within CDI pipelines or standalone

settings_suggest

How It Works

- You create a Data Validation task in IDMC
- Define one table pair (source + target)
- Uses identical connections as Cloud Data Integration
- Task generates underlying mappings automatically
- Run the task — mappings execute on Hosted Agent
- Review bad record reports and dashboards

hub

Where It Fits in IDMC

- Service in IDMC alongside CDI, CDGC, MDM
- REST API enables CI/CD pipeline quality gates

Common Use Cases:

- Prove CDI pipeline loaded all rows correctly
- SUM(revenue) source = SUM(revenue) target
- No orphaned FK records in target after load
- All source IDs present in target (no drops)
- MD5 checksum key column comparison

Data Validation Techniques

Informatica from Salesforce

Comparison Approaches

table_chart

RAW RECONCILIATION

Table-Level Comparisons

Compare the entire table directly based on actual underlying data values to ensure exact row-by-row consistency.

functions

AGGREGATE METRICS

Aggregation Functions

Apply mathematical and aggregation functions to compare datasets at scale without checking every single row.

storage

VERSATILE SYSTEMS

Flexible Data Sources

Validate seamlessly across database tables, database views, or pre-saved custom SQL queries.

Data Validation Techniques

Informatica from Salesforce

Table level comparison - Map source and target columns side-by-side to verify logic consistency

Mapping Configuration

- **Source-to-Target Column Pairing:** In the CDV test case, developers pair the columns of Connection 1 (Oracle) with Connection 2 (S3). This forms the foundation for data-level reconciliation.
- **Exact Match Rules:** Configure string-level matching, handling trailing white spaces, case-sensitivity differences, and string-to-numeric comparisons.
- **Isolating Schema Mismatches:** The CDV column mapping engine parses underlying field datat ypes, precision, and scale. Any incompatible schemas are instantly flagged in the test case editor to prevent execution failures.
- **Key Matching Constraints:** Allows the selection of unique primary keys across endpoints to ensure matching logic runs on corresponding records.

Mapped?	Connection 1 (Oracle)	Connection 2 (S3)
No	**ID** Integer (10,0)	*Select a column...* Integer (10,0)
Yes	**Name** String (10,0)	**Name_1** String (10,0)
Yes	**Currency Code** Integer (10,0)	**Currency Code_2** Integer (10,0)
Yes	**Date** Date (29,9) *	**Date_1** String (29,9) *
Yes	**Last Modified Time** Time (29,9)	**Last Modified Time_1** Time (29,9)

Data Validation Techniques

Informatica from Salesforce

Parameterized test cases and saved SQL queries maximize validation reuse

history_edu

REUSABLE CUSTOM CODE

Saved SQL Queries

Allows data architects to save complex, custom SQL validation queries once and reuse them across multiple test suites, removing manual setup and reducing coding errors.

tune

DYNAMIC ENVIRONMENTS

Parameterized Test Cases

Write a single validation rule and execute it dynamically across different environments (Dev, QA, Prod) or different partitions/shards of data.

api

AUTOMATED PIPELINES

CDV REST API

Triggers validation runs programmatically and enables conversational interactions with integration platforms, minimizing graphical UI dependence.

Test Cases

Informatica from Salesforce

Types of Test Cases using Saved SQL Queries

Row Count

Compares the number of rows between source and target.

```
src_count = tgt_count
```

Use: SLA minimums

Null Check

Validates that specified columns contain no null or empty values.

```
customer_id IS NOT NULL
```

Use: Required fields

Duplicate Check

Detects duplicate values across one or more columns.

```
COUNT(DISTINCT id) = COUNT(*)
```

Use: PK uniqueness

Value Range

Checks that column values fall within min/max bounds.

```
age BETWEEN 0 AND 120
```

Use: Numeric/date bounds

Pattern Match

Validates values against a regex pattern for format consistency.

```
REGEXP_MATCH(email, '...')
```

Use: Emails, zip codes

Set Membership

Checks if values belong to an approved list or reference table.

```
status IN ('ACTIVE','CLOSED')
```

Use: Enum columns

Aggregate Match

Compares aggregates (SUM, AVG) with configurable tolerance.

```
ABS(SUM(src) - SUM(tgt)) <= 0.01
```

Use: Financial reconciliations

Custom Expression

Write any SQL expression or complex rule not covered by templates.

```
end_date > start_date
```

Use: Complex business rules

Test Cases

Execution of Test Cases

▶️ Run On Demand

Trigger a single run immediately from the IDMC console. Use for ad-hoc validation, testing new rules, or investigating a data issue.

When: Dev & testing, incident investigation, one-time scans

🗓️ Scheduled Run

Configure a cron-style schedule to run the validation task automatically at defined intervals — hourly, daily, weekly, or custom cron.

When: Daily quality monitoring, nightly ETL validation, SLA checks

⚡ Inline (Within CDI)

Embed CDV validation rules directly inside a Cloud Data Integration pipeline mapping. Records are validated inline during the load.

When: Real-time pipelines, streaming CDC, zero-tolerance gates

🔌 REST API Trigger

Trigger CDV jobs programmatically via IDMC REST API. Integrate into CI/CD pipelines, orchestration tools (Airflow, ADF), or event-driven setups.

When: CI/CD quality gates, automated orchestration workflows

Queued → Initializing → Running Tests → Generating Reports → Completed / Failed

Test Cases

Result of Test Cases

Test Case Report Analysis

• **Row Count Reconciliation:** CDV validates overall dataset scale. In our execution test, Connection 1 processed 13 records and Connection 2 processed 11, resulting in an overall row count failure.

• **Detailed Match Analytics:** Out of 14 total evaluated records, 9 matched, 1 mismatched, 1 extra record was found in Connection 2, and 3 records were missing in Connection 2.

• **Column-Level Status Checks:** Identifies exactly which column-level rules passed or failed. In our run, the core 'ID' comparison failed (unsuccessful), while Name, Currency Code, Date, and Last Modified Time successfully validated.

• **Actionable CSV Logging:** Enables immediate downloading of full reconciliation error logs. Teams can audit precise record keys that failed column matching for swift remediation.

Testcase 1_05 Execution Summary

• Overall Test Status: ❌ UNSUCCESSFUL
• Duration: 6 minutes | Total Records: 14

Data Distribution Overview:
• Connection 1 (Oracle) Records: 13
• Connection 2 (S3) Records: 11
• Matched: 9 | Mismatched: 1 | Extra: 1 | Missing: 3

Matching Rule Assertions:
❌ COUNT_ROWS(Table A) = COUNT_ROWS(Table B) (Failed)
❌ ID (Table A) = ID (Table B) (Failed)
✓ NAME (Table A) = NAME (Table B) (Success)
✓ DATE (Table A) = DATE (Table B) (Success)
✓ LASTMODIFIEDTIME (Table A) = LASTMODIFIEDTIME (Table B) (Success)

Test Cases

Connections — Supported Sources & Targets

📌 **Data Validation** uses the same connections as Cloud Data Integration. Create connections in IDMC Administration > Connections. The Hosted Agent handles all connectivity — no additional configuration needed.

Cloud Applications

- Salesforce

Cloud Warehouses

- Snowflakes
- Google BigQuery
- Amazon Redshift
- Azure Synapse
- Databricks
- Netezza
- Teradata

Relational Databases

- Oracle DB
- MS SQL Server
- PostgreSQL
- MySQL
- IBM Db2
- SAP HANA

Files & Object Storage

- Amazon S3
- Azure ADLS Gen2
- Flat Files

Test Suites

Informatica from Salesforce

Use test suites to generate reports of multiple test cases simultaneously. Test suite reports display the matching results of each test case job.

Key Benefits

assessment Consolidated Reporting

View success states and detailed metrics for all jobs in a single place, eliminating the need to open each test case individually.

visibility Simultaneous Monitoring

Effortlessly track and oversee multiple database validation checks running at the exact same time.

Business Outcomes

Use Case 1: Validate converted mappings side-by-side to de-risk cloud warehouse cutovers

Automated Parallel Run Verification

When organizations migrate their legacy PowerCenter mappings, sessions, and workflows to IDMC, they run both environments in parallel during the testing phase. Cloud Data Validation (CDV) acts as the central comparative framework:

- **Side-by-Side Interface**
  Allows users to easily compare dataset outputs side-by-side to spot differences instantly.
- **Functional Integrity**
  Verifies that converted mapping assets process identical input data and produce exactly matching outputs, preserving historic business logic.
- **Outlier Detection**
  Rapidly exposes missing records, extra records, or value mismatches in the target lakehouse (Databricks, Snowflake) before going live.

How Side-by-Side Validation Works

Source (Legacy Oracle/SQL)
Extracts expected data outputs generated by the legacy on-premises mapping engine.

Conversion & Repointing
PC2CDI automates over 90% of mapping logic conversion, redirecting target endpoints to Databricks/Snowflake.

Target (Modern Lakehouse)
CDV executes parallel comparative tests, matching rows across columns on the newly loaded targets.

Validation Outcome
Summary and detailed test reports flag any missing records, extra rows, or cell value discrepancies.

Business Outcomes

Use Case 2: Endpoint-aware rules prevent false failures across diverse cloud targets

Managing Dialect & Target Divergence

A typical multi-cloud strategy introduces diverse data engines. Reconciling data across different physical tiers requires intelligent translation layers:

- **File-to-Database & DB-to-DB Checks**

CDV natively bridges comparisons between flat files (e.g., in Amazon S3) and structured tables (e.g., Snowflake).

- **Endpoint Difference Awareness**

Target modern platforms handle date/time formats, data types, and case sensitivity differently than on-premises databases. CDV is designed to handle these endpoint formatting variations natively to prevent false testing failures.

- **Exact Matching**

Empowers data teams with flexible rules—choosing strictly exact matches for finance tables, or matches for customer names.

✓ Endpoint Difference Awareness

Adapts rule execution to expected variations in modern target structures, such as date-time format conversions or column casing shifts, without throwing false positives.

✓ Exact Matching

Strict cell-by-cell validation. Crucial for financial numbers, mathematical calculations, primary key integrity, and absolute data replication audits.

✓ Aggregate & Sampling Options

Avoids heavy compute bills and data egress by running aggregate function checks (like sum or count) or sampling subsets of very large multi-terabyte datasets.

Business Outcomes

Use Case 3: REST APIs embed parameterized validation directly into active dev pipelines

Transitioning to API-Driven QA

In modern, agile development teams, manual interface navigation is a critical bottleneck. CDV offers robust features that move data validation from a static UI task to a continuous, automated workflow:

- **CDV REST API**
  Enables data engineers to programmatically trigger validation runs, manage test cases, and capture success/failure outputs.
- **Parameterized Test Cases**
  Developers can write a single validation rule and execute it dynamically across different environments (Dev, QA, Prod) or partition slices.
- **Saved SQL Queries**
  Teams can store and reuse complex, custom-written SQL validations, drastically accelerating the testing of highly transformed datasets.

01 | Saved SQL Queries
Allows developers to execute custom SQL directly in CDV, saving time on complex or custom transformation tests.

02 | Parameterized Tests
Maximizes rule reuse. A single test rule can run dynamically across Dev, QA, and Production by swapping connection parameters.

03 | CDV REST API
Bypasses manual UI navigation by triggering execution programmatically from external CI/CD pipelines.

Business Outcomes

Best Practices

🗂️ Organize with Groups

Use test groups within each table pair to categorize tests by dimension (Completeness, Accuracy, Uniqueness). This makes summary reports easier to read and lets you set group-level thresholds independently.

📐 Use Constraints to Scope

Always apply a date constraint (e.g. WHERE load_date = CURRENT_DATE) on large tables. Validating only the incremental load is faster, cheaper, and more actionable than scanning the entire table every run.

⚠️ Set Severity Levels

Use Error for tests that must block the pipeline (null PKs, FK violations). Use Warning for tests that need review but should not block (format inconsistencies, minor null rates). Use Info for trend monitoring only.

🔁 Start Simple, Then Layer

Begin with Row Count and SUM reconciliation tests — they catch 80% of ETL bugs. Add column-level tests in the next sprint. Add statistical anomaly tests after you have 4+ weeks of baseline run history.

🔌 Automate via REST API

Do not rely on manual task runs in production. Wire CDV into your CI/CD pipeline using the REST API. Trigger runs automatically after every CDI pipeline execution and block deployments when quality drops below threshold.

🔐 Manage Connections Centrally

Create and test all CDV connections in IDMC Administration > Connections — not inside individual tasks. Central connection management makes credential rotation, permission audits, and connection reuse much simpler.

Business Outcomes

Modernize up to 8x faster while reducing migration costs by 50%

speed

ACCELERATE DELIVERY

8x Faster Value

Experience an 8x faster time-to-value by modernizing legacy PowerCenter workloads 'as is' without business logic migration.

payments

REDUCE TCO

50% Lower Cost

Automate over 90% of assets migration via self-service modernization tools, eliminating manual upgrades and patching.

verified

ENSURE ACCURACY

100% Data Trust

Establish a completely verified, secure, and auditable data foundation to support critical enterprise AI and analytics.

Demo -Data Validation in Action

Informatica from Salesforce

What the Demo Covers

An interactive walk-through of the core capabilities and workflows for automated data validation and testing.

01 Creating & Running a Test Case

Define connections, select data sources, map columns, configure sampling and notifications, run the test case.

02 Running a Test Suite

Group test cases into a suite, run simultaneously, and view a consolidated report of all job results.

03 Using Saved SQL Queries

Create a saved SQL query to filter and combine data sets, then reference it across multiple test cases.

04 Reviewing Reports

View summary and detailed reports - identify unmatched, missing, and extra records between source and target.

05 Monitor Dashboard

View test case jobs, check run history, view logs for debugging failed jobs.

Q&A

Thank you