A FirstEigen White Paper

Replacing Sampling-Based Migration Validation

Why Enterprises Are Standardizing on DataBuck for Audit-Grade Cloud Migration Confidence

Executive Summary

Large-scale data migrations fail when organizations cannot prove data correctness at scale. Sampling-based validation and manual configuration approaches do not work for enterprise migrations involving thousands of tables and billions of records.

DataBuck is purpose-built for audit-grade, scalable migration validation on modern cloud platforms. Pelican, while suitable for smaller datasets, does not scale operationally or technically for enterprise data modernization programs.

Executive Decision Criteria

1

Scalability

Ability to validate billions of records

2

Accuracy

Exact mismatch counts, not estimates

3

Speed

Faster migration completion and sign-off

4

Operational Simplicity

Minimal manual configuration

5

Audit Readiness

Deterministic, provable results

Where Sampling-Based Migration Validation Fails

Enterprise cloud migrations have accelerated dramatically. Organizations are modernizing legacy platforms such as Teradata, Oracle, and on-prem data warehouses to cloud-native environments like Databricks and Snowflake.

Yet despite these investments, a critical business risk persists: **leaders cannot confidently prove that migrated data is complete, accurate, and consistent before business sign-off.**
Traditional migration validation approaches rely on:

- Sampling
- Row-count estimates
- Partial checks
- Manual SQL comparisons

As data volumes and complexity increase, the limitations of traditional sampling-based tools such as Pelican become unavoidable due to:

Sampling-Centric Architecture

Sampling validates only a fraction of data, assuming that subsets represent the whole. In enterprise datasets, this assumption routinely fails.

- Misses long-tail errors
- Ignores rare but material records
- Cannot detect aggregate inconsistencies

Non-Auditable Results

Executives, auditors, and regulators require exact counts, not percentages.

- "99.9% matched" is not audit evidence
- No visibility into which records failed
- No defensible trail for compliance reviews

Manual Effort at Scale

Sampling tools demand significant human involvement to configure, interpret, and justify results.

- ~1 hour of setup per table
- Manual interpretation of mismatches
- High operational drag during migration windows

Business Impact: Sampling creates false confidence while leaving material risks undiscovered until after go-live.

Platform Comparison (Executive View)

Capability	DataBuck	Pelican
Proven scale	14+ Billion records, 150+ columns	Fails beyond ~2 million records
Validation approach	Full data comparison	Sampling-based
Mismatch reporting	All mismatched records	Limited samples
Mismatch accuracy	Exact mismatch counts	Percentage estimates
Aggregate validation	Supported	Not supported
Primary key detection	ML-based auto-detection	Manual
Column mapping	Intelligent auto-mapping	Manual
Average setup per table	~5 minutes	~1 hour
Enterprise readiness	Designed for cloud-scale migrations	Limited to small datasets

Why the Difference Matters

1 Scale is a hard requirement, not an optional feature.

Enterprise migrations routinely involve thousands of tables and hundreds of millions to billions of records.

2 Sampling creates business risk.

Executives require provable correctness for audit, compliance, and financial sign-off.

3 Manual configuration multiplies cost and risk.

At enterprise scale, manual setup results in thousands of hours of effort and increased error probability.

Case Study: Global Automotive Manufacturer

The Challenge

A global automotive manufacturer undertook a large-scale modernization initiative:

- **Migration:** Teradata → Databricks
- **Scope:** ~9,000 tables
- **Data Volume:** Average 100M rows per table (1M–500M range)

The organization initially evaluated a sampling-based validation tool (Pelican). The approach failed to scale beyond small tables and required extensive manual effort, delaying the migration timeline.

The Pelican Solution

By deploying Pelican, the outcome was:

- High manual effort even for small tables and failed proof-of-concept on tables >2M rows
- Consistent failures for tables larger than 10 million records
- High operational overhead even for smaller tables

The DataBuck Solution

By deploying DataBuck, the organization achieved:

- Full reconciliation across all tables
- Automated key detection and column mapping
- Exact row-level and aggregate validation

30	9K	100%
Days	Tables	Confidence
Complete validation delivered in under 30 days with DataBuck	Accurate reconciliation across billions of records with DataBuck	Confident executive and audit sign-off with DataBuck

> **Pelican failed proof-of-concept on most large tables and could not support enterprise-scale validation.**

Transform Migration Validation with DataBuck

Executive Recommendation

For enterprise-scale data migrations, DataBuck provides provable data trust, massive scalability, and automation that reduces risk, effort, and time-to-signoff. Sampling-based tools, like Pelican, by their very nature, cannot guarantee 100% data fidelity. They may be acceptable as high-level, directional indicators of data fidelity, but are not suitable for mission-critical modernization programs.

Ready to eliminate migration risk?

Join enterprises standardizing on DataBuck for cloud migration validation.

About FirstEigen & DataBuck

FirstEigen's DataBuck platform is purpose-built for enterprise-scale data quality, migration validation, and modernization assurance, delivering provable correctness across the most complex data environments.
Learn more: www.firsteigen.com/databuck