IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 3, Issue 2, October 2023

Impact Factor: 7.301

# Enterprise Playbook: Validating Billions of Rows Safely into Redshift - Design Patterns and Anti-Patterns

**Maheshbhai K Kansara**

Mill Creek Seattle, Seattle, WA, USA

**Abstract:** *Validating billions of rows of heterogeneous systems like SQL Server, Oracle, PostgreSQL, and Amazon Redshift are common across migration of enterprise-scale data warehouses. This is not a trivial problem when it comes to assuring the correctness of the data when it comes to these large-scale Extract-Transform-Load (ETL) processes, especially where performance, reliability and cost-effectiveness are to be evaluated. The paper suggests an enterprise safe validation playbook of large datasets organized around effective design patterns and important anti-patterns. We dwell upon the purpose of column checksums, row counts and pipeline scaling mechanism as pillars of effective validation strategies. Based on customer evidence on the real-world implementation, we demonstrate some of the successful methods and traps in practice. The article does not just focus on the technical nature of the massive data validation, but gives a lot of attention to the operational relevance of the enterprises that are in the process of digital transformation. Finally, the presented playbook gives academic depth and practical effect, as it offers advice to both database architects and engineers, as well as decision makers, who are involved in managing enterprise data pipelines.*

**Keywords***:* Data validation, Redshift, enterprise pipelines, ETL, checksums

DOI: 10.48175/IJARSCT- 13555G