

High Data Coverage and Data Quality with Smart Testing Solution

Customer Overview

The customer is a leading book, periodical, and trade publisher based in New York.

The Business Context

As part of a migration project, the customer wanted data validation and cleansing to ensure high data quality. Since data cleansing requires the validation of all migrated data between the source and target databases, manual validation based on random sampling techniques would not have sufficed. Validation of the entire dataset at various levels was required to avoid movement of redundant data and for maximum accuracy of the migrated data. To accomplish this, the customer required comprehensive testing with a reliable data validation and test automation tool.

Type of Service Provided

Functional/Performance/Automation Testing
Quality Assurance

Technologies Used

QuerySurge, Selenium, SQL Server, JMeter, Oracle SQL Developer

The Solution

The team of product engineering and data experts at GS Lab | GAVS decided to use QuerySurge – a smart data testing solution. A systematic approach was followed to achieve maximum accuracy. The team obtained access to the relevant databases. However, there were many instances of restricted access to databases and data in flat file formats such as CSV, XML, etc. QuerySurge's features made it possible to validate data across all databases and file formats. QuerySurge enabled effective use of time and resources by allowing analysis of validation results parallelly while automated validation schedules and test scenarios performed backend validations. Reporting mechanisms were developed to highlight data differences, mismatches, data flow failures, etc. that helped minimize memory implications. Audit trail was maintained on subsequent validations of the same data, for review and analysis. This end-to-end validation of the entire dataset delivered 100% data coverage and reliable data quality at speed.

To find out how GS Lab | GAVS can help your organization, please visit www.gavstech.com

Challenges

- Huge volumes of complex data to be validated
- Diverse data sources – from multiple databases and servers, with varied schema and formats
- Difficulties in manually finding missing data between tables with MINUS queries:
 - Not possible to find all missing values if there are duplicates or noise in data
 - Lack of a reporting mechanism or audit trail
 - Time consuming, resource intensive, and prone to errors
 - Does not provide complete data correctness or 100% data coverage
 - Time consuming validation of data types for each schema for the 100+ migrated tables
 - Cumbersome validation of millions of rows for most of the tables during DB refresh
 - Accuracy cannot be achieved quickly with production tables having 25k+ records each
 - Risky for the business to get only count matches in PROD for DB refresh and other patching

Solution Highlights

- End-to-end validation of the entire dataset
- Validation of data between different databases and varied file formats
- Data validation with 100% data coverage and accuracy
- Parallel analysis of results while automated schedules performed backend validations
- Reporting mechanisms to highlight data differences, mismatches, data flow failures, etc.
- Audit trail maintained on subsequent validations of the same data, for review and analysis

Solution Impact

- **~80% Increase in Efficiency** through automation and elimination of manual effort
- **High Resource Cost Savings, Reduced TAT, Increase in Data Coverage from <0.1% to 100%** through reduction in manual effort from 3-4 days for 30k records for example, to <2-4 hours
- **Effective Use of Time and Resources** through parallel execution of validation and result analysis
- **High Data Accuracy and Reliability, In Depth Error Detection** in just 18 months
- **Reduced Memory Implications** through reporting on duplication, mismatches, failures, etc.
- **Adoption of QuerySurge as Testing Solution** by customer after extensive research by the team