

SCIKI 

ALL-IN-ONE DATA PLATFORM

Data analytics **without boundaries**

Enabling businesses and enterprises to deliver data-driven decisions in a fraction of the time compared to other data platforms

SCIKIQ DQ Overview

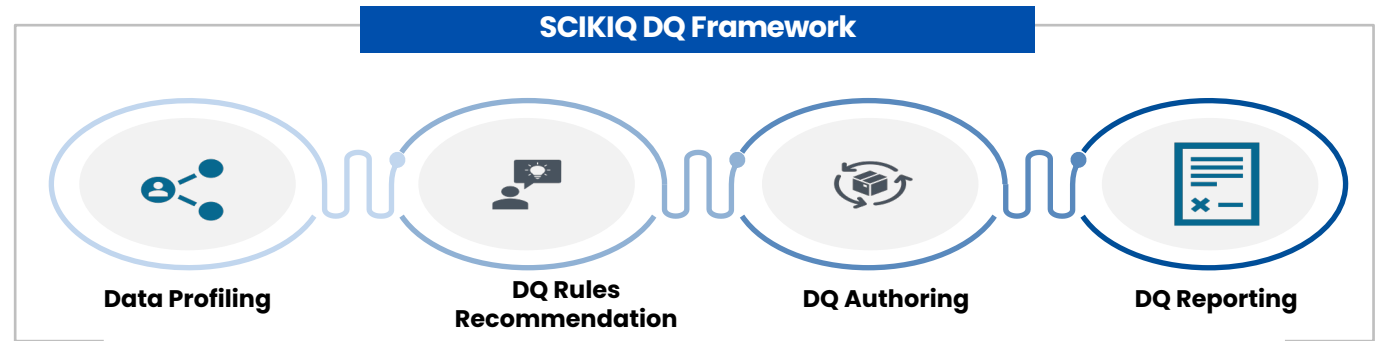
High Level Deck

OUR APPROACH TO DQ AND FRAMEWORK

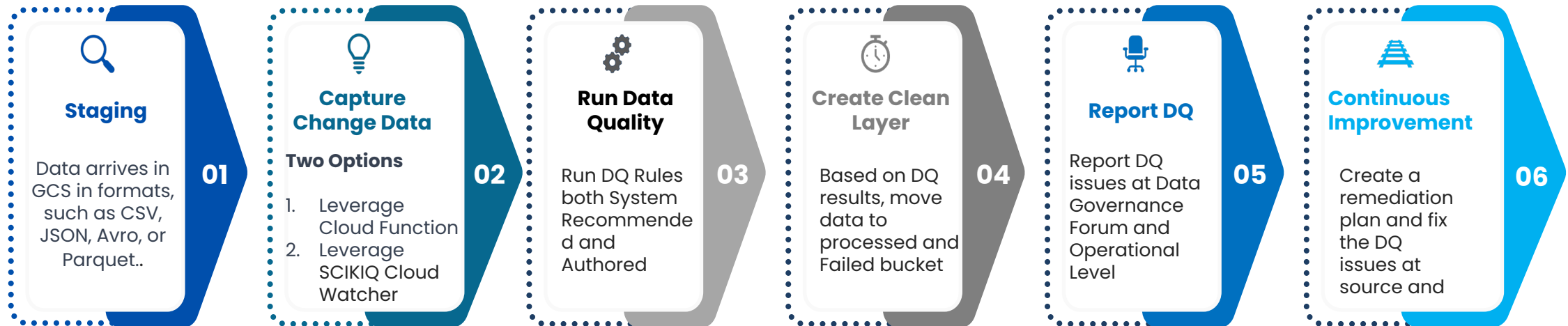
3% data in businesses meets quality standards

We take Data Quality Seriously

Our Framework

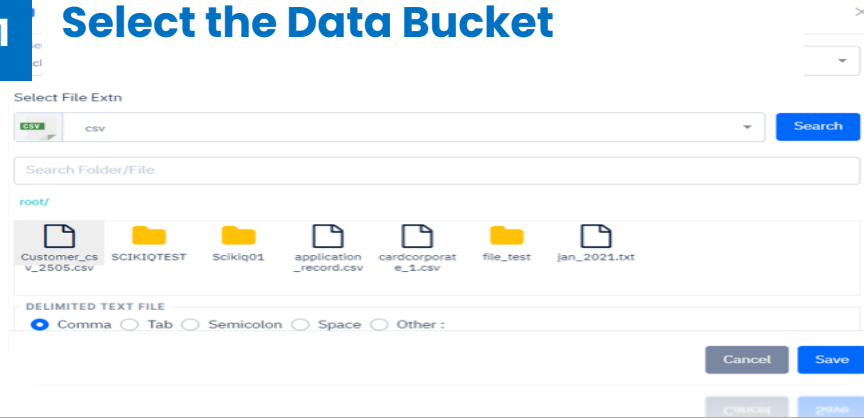


Our Approach

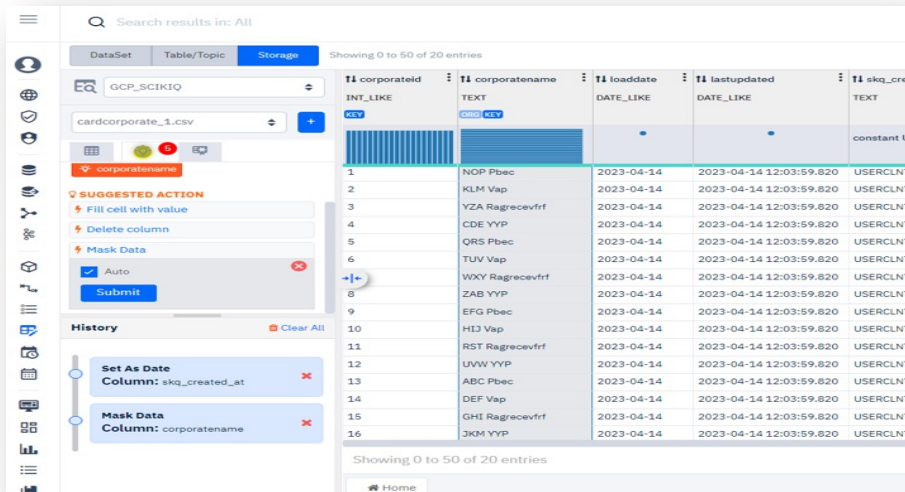


OUR DQ FRAMEWORK – UNDER THE HOOD

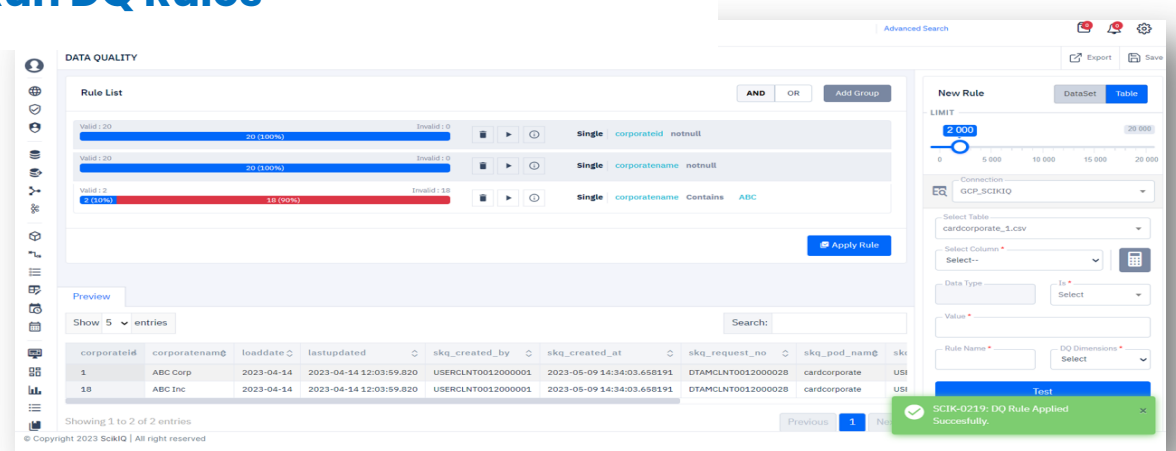
01 Select the Data Bucket



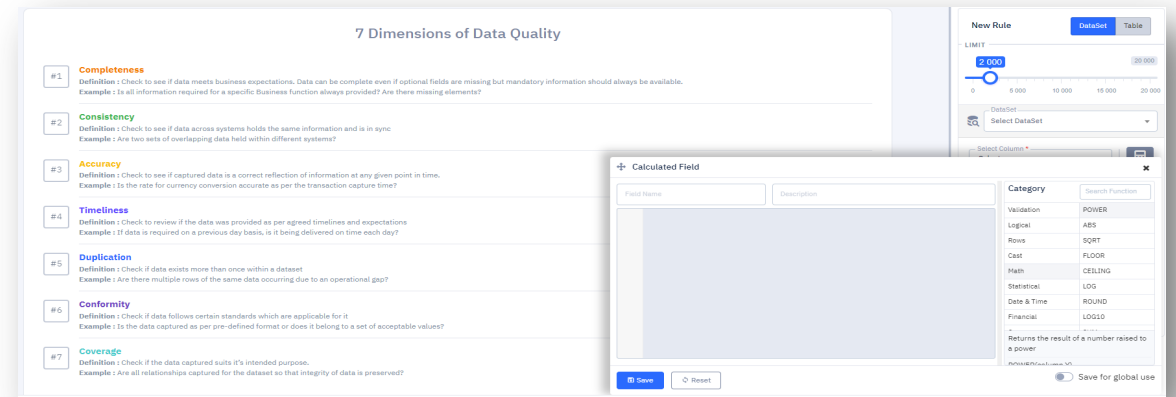
02 Profile the Data



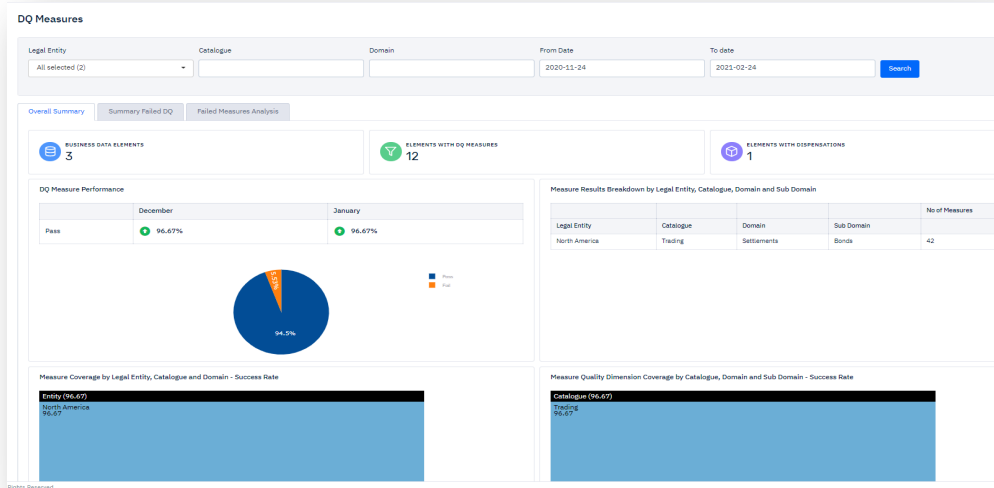
04 Run DQ Rules



03 Author your Rules



OUR DQ FRAMEWORK – UNDER THE HOOD



Data Quality Reporting at Board Level

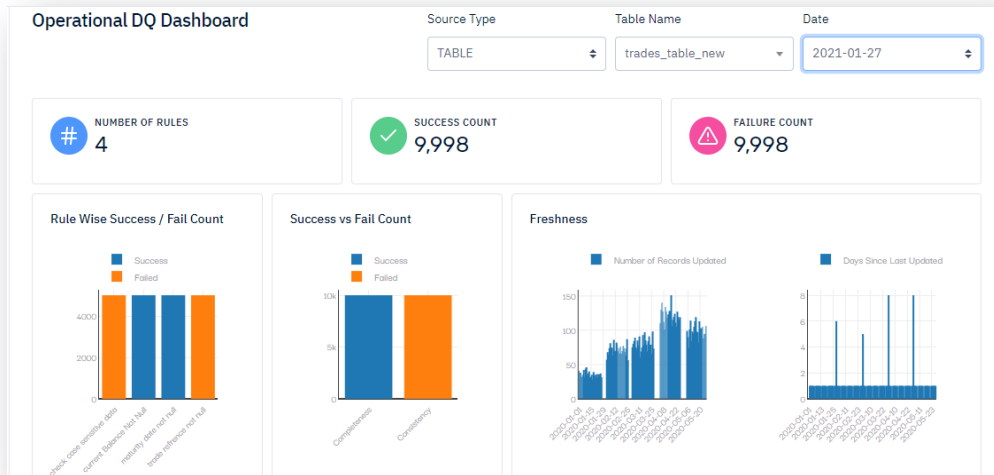
With in ScikIQ we have 3 dashboards to help stakeholders understand over all data quality structural diagnostics and data quality assessment reliability.

- Overall Summary
- Summary Failed DQ
- Failed Measure Analysis

Data Quality Reporting at Operational Level

This dashboard helps answer the below data observatory questions:

- Freshness – Is the data up-to date or are there any gaps where data has not been updated
- Distribution – Is the data within expected ranges
- Volume – is the data intake meeting expected thresholds
- Schema – Identify change in formal structure of data management system



GCP DEDICATED VS PARALLEL SPOT INSTANCES

SCIKIQ DQ can process data in both the cases of Vertical Process or Parallel Spot instances

Vertical Dedicated Process: Utilizing a single, dedicated server for your application.

Advantages:

- Predictable Performance: No contention with other workloads ensures stable performance.
- Reliability: Resources are not shared, reducing the risk of performance degradation.
- Security: Enhanced isolation from other users' workloads.
- Flexibility: Full control over server configuration

Use Cases

- Critical Applications: Mission-critical workloads requiring stable performance.
- Sensitive Data: Applications demanding high security and isolation.

Spot Instances: Leveraging unused computing resources at significantly reduced costs.

Advantages:

- Cost-Effectiveness: Significant cost savings, up to 90% compared to regular on-demand instances.
- Flexibility: Suitable for non-critical and fault-tolerant workloads.
- Batch Processing: Ideal for tasks with flexible scheduling and time tolerance.
- Termination Notice: 30-second notice before instance termination.

Use cases

- Non-Critical Workloads: Non-essential tasks with flexible scheduling.
- Cost-Optimized Processing: Large-scale data processing with cost constraints.

PRICING & TENTATIVE SYSTEM REQUIREMENTS

Tentative System requirements with various File types

File/Type	Data Source Size	Passed Bucket Size	Filed Bucket Size	Total Capacity Required
CSV/JSON	50GB	45 GB	5 GB	100 GB
Parquet	50GB	36 GB	4 GB	80 GB
Avro	50GB	40.5 GB	4.5 GB	90 GB

System requirements

Minimum

2 Core 4/16

Ideal recommendation is

2 Core 8/32

Pricing:

Tailored specifically for legal entities under the licensing agreement.
Valid for 3 distinct entities and 6 specialized departments for Tech Mahindra.

Cost of License/every year: \$120k



We look forward to partnering with you

- ✓ Data Acquisitions takes > 6 months.
- ✓ Need to build fully governed Data Lakehouse
- ✓ Need single tool which empowers business teams to take control of their data
- ✓ Need Integrated Tool for Data Governance & Data Quality
- ✓ Need Full Record to Report Lineage to satisfy Regulators
- ✓ Need help pulling Data from SAP/EBS or APIs and on real time basis

SCIKIQ

FORRESTER

Source:Forrester Research, Inc.

**SCIKIQ (DATA AND ANALYTICS SERVICES) RECOGNIZED
AMONG 34 AUGMENTED BUSINESS INTELLIGENCE PLATFORMS**



Sciq.com