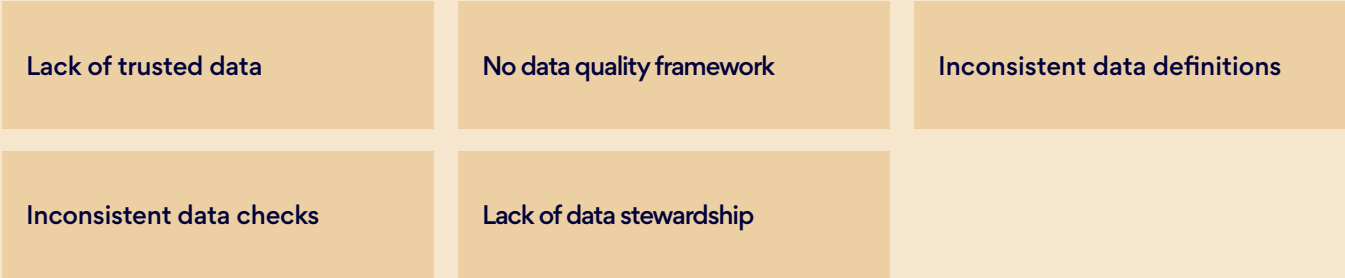**Persistent**

Solution Sheet

AI-Driven
Data Quality

For businesses, the effectiveness of informed decision making is directly proportional to the quality of their data. Trusted data instills confidence in decision making. Yet much of the data today is often un-useable across the business or even within a function. Astonishingly few businesses meet even basic data quality standards or know what criteria define those standards.

Lack of trusted data	No data quality framework	Inconsistent data definitions
Inconsistent data checks	Lack of data stewardship	

Persistent's AI-Driven Data Quality Framework

Built on the pillars of data quality monitoring, reporting and stewardship, Persistent's AI-driven data quality framework streamlines data quality checks every step of the way. The framework establishes the data quality process, automatically identifies data issues with machine learning, monitors data quality metrics in real-time and ensures rapid data correction.

Data Quality Methodology

1\ Define

Define data quality process, identify data stewards and streamline tools required for data quality

\ Identify data stewards to drive and own DQ initiative

\ Prioritize business area or pain area to be taken-up first

\ Define raw data zone, data processing zone and verified dataset zones to ensure correct expectations of data quality

\ Define KPI & metrics that need to be measured

2\ Measure

Measure data quality metrics using AI and business rules across all steps of the analytics

\ Use AI-driven accelerators to spot data quality issues

\ Use config driven business rules framework to spot data quality issues

\ Leverage data quality data model effectively store measured data quality results

\ Integrate data quality pipelines into data processing scheduler

3\ Analyze

Analyze quality metrics to identify areas of improvements & causes of data quality issues

\ Realtime and batch monitoring of data quality metrics

\ Data quality dashboard and reports for individual metrics

\ Trends, anomalies and outliers shown through explainable dashboards

\ Rate each LOB or subject area on quality benchmark

4\ Improvise

Change process, fix data issues, track historical trends to improve overall data quality

\ Add data correction, data clean-up logic at appropriate steps

\ Capture rate of improvement of the data quality

\ Track impact of improved data quality on business processes

\ Ensure data quality process overheads eventually become lesser than effort of fixing issues manually

Components Within Persistent's AI-Driven Data Quality Framework

DQ Dashboards

\ Quality metrics dashboards
\ Complete visibility into data issues

\ Different widget types for different DQ checks
\ One widget per quality metric
\ One dashboard is collection of widgets per one subject area
\ Role-based access to DQ dashboards.

DQ Pipeline

\ Quality data pipelines
\ Pre-process, in-process, post-process
\ Capture trends in metrics

\ Pre-process — DQ on raw data zone
\ In-process — DQ during data processing workflows
\ Post-process — DQ on verified data zone
\ Save DQ observations in DB
\ Data pipeline using airflow

DQ Rules

\ Define quality metrics
\ AI driven quality rules
\ Organize based on lobs

\ AI-based approach automatically spots hidden data issues
\ Define quality metrics by specifying table, columns, expected values, lower/upper bounds, range leveraging existing templates
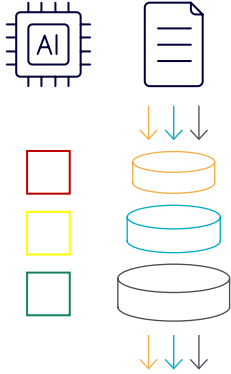\ Focus on all DQ dimensions — completeness, accuracy, timeliness, consistency, uniqueness, and validity

DQ Framework

\ Simple configurable framework
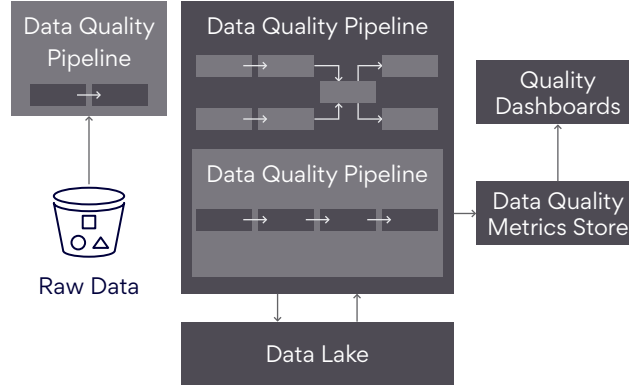\ Ready to use templates
\ Extensible with custom SQL

\ Python-based framework to configure quality check
\ Templates for — count verification, conformity, consistency, trends, completeness, control limits
\ Extensible to include custom SQL and to integrate with existing DQ tools

How It Works

AI-driven multi-tier process

Data Quality Pipelines

Data Quality Pipeline

Raw Data

Data Quality Pipeline

Data Quality Pipeline

Quality Dashboards

Data Quality Metrics Store

Data Lake

Data Quality Dashboards

DQ Pre-Process

\ Ensure raw data is checked for quality before it's taken for processing

\ DQ process such as pre-file validations, file format checks, CRC checks some examples

\ AI-driven metadata driven framework

DQ In-Process

\ DQ process during loading and processing step

\ Checks such as constraint validation, format check of a particular field, master ID validations, duplicate checks are in process DQ checks

\ AI-driven anomaly detection, outlier detection

DQ Post Process

\ DQ process after ETL loads

\ Checks such as post load validations, like control total validation, thresholds, schema checks

\ AI-driven framework for anomaly detection, outlier detection, and emerging trends

Make your data work for you by improving data quality across the business.

Request Demo

About Persistent

We are a trusted Digital Engineering and Enterprise Modernization partner, combining deep technical expertise and industry experience to help our clients anticipate what's next. Our offerings and proven solutions create a unique competitive advantage for our clients by giving them the power to see beyond and rise above. We work with many industry-leading organizations world-wide including 14 of the 30 most innovative US companies, 80% of the largest banks in the US and India, and numerous innovators across the healthcare ecosystem. Our company fosters a values-driven and people-centric work environment. Our strength of over 22,500+ employees is spread over 18 different countries across the globe.

USA
Persistent Systems, Inc.
2055 Laurelwood Road, Suite 210
Santa Clara, CA 95054
Tel: +1 (408) 216 7010
Fax: +1 (408) 451 9177
Email: info@persistent.com

India
Persistent Systems Limited
Bhageerath, 402
Senapati Bapat Road
Pune 411016
Tel: +91 (20) 6703 0000
Fax: +91 (20) 6703 0008

Persistent

www.persistent.com