Required capabilities for enterprise-scale data quality and observability

Data teams are often constrained by manual rule writing and management, with limited data coverage and a siloed view of data quality. To make things worse, data producers and data consumers often operate in silos and are unable to identify the opportunities to improve data quality in a business context. As a result, organizations lack an enterprise data quality foundation to respond to regulatory, analytics and AI demands in a scalable and cost-efficient way. Organizations need an enterprise-scale data quality and observability solution that:

- Auto-discovers issues in data without requiring domain experts and rule writers
- Scales data quality across large and diverse databases, files, and streaming data
- Automatically uncovers data drift, outliers, patterns and schema changes
- Reduces the risk and cost of data migrations via robust data reconciliation
- Puts business users at the forefront of building data trust via self-service data quality
- Traces data movement and helps data quality teams narrow the focus of root cause investigations
- Initiates remediation workflows with the right data owners and data stewards
- Focuses data quality efforts on business-critical data for maximum impact

Required capabilities	Features	Collibra	Other vendors
 **Auto discovery to figure out issues in your data without requiring domain experts and rule writers**	Associative ML-driven, autogenerated, SQL-based, non-proprietary, explainable and adaptive data quality rules	Yes	No
	Automatically put all or select columns under quality control including null checks, empty checks, statistical profiles, variance analysis and sketches	Yes	No
	Automated rules engine to design, create and deploy business rules for specific data values both in batch or real-time mode	Yes	No
	Automated rule inference and generation for a single attribute, for example, inferring that a social security number is always numeric and has nine digits	Yes	No
	Automated rule inference and generation for multiple attributes, for example, inferring that a start date occurs before the end date	Yes	No
	Utilize machine learning to automatically find exact and similar matches in data records	Yes	Yes
	Configure and customize the algorithms for matching, merging, linking and de-duplication	Limited	Yes
	Ability to write custom rules for simple and complex scenarios	Yes	Yes

Required capabilities	Features	Collibra	Other vendors
 **Horizontal and vertical scalability for establishing enterprise-wide trust in data**	Scan large and diverse databases, files and streaming data with Spark-based parallel processing	Yes	Limited
	Native connectors to a vast variety of data sources and systems for easy data source registration and metadata ingestion	Yes	Yes
	Deliver suitable throughput and response times to satisfy performance SLAs in both batch and real-time modes	Yes	Limited
	Scan data where it resides (at the edge site), without needing to move or extract the data	Yes	Limited
	Profile data outside of data sources by extracting and importing the data into the data quality solution or third-party repositories such as Hadoop	Yes	Yes
	Connect to the existing stack quickly (REST APIs) with no need to modify pipelines or write new code	Yes	Limited
	Apply data quality functions against real-time data streams (e.g., securities trading data and telematics)	Yes	Limited
 **Time series anomaly detection for end-to-end data observability across pipelines**	Automatically detect behaviors (drift and shift) i.e. if any column, schema or cell value has suddenly broken its past trend	Yes	Limited
	Automatically detect numeric/categorical outliers and patterns such as cross-column correlations and anomalies	Yes	Limited
	Automatic schema change detection i.e. if data fields, columns and data types are added, removed or changed	Yes	Limited
	Profile data sets and provide metrics on actual and inferred data types, minimum and maximum values, value frequencies, null value counts and unique values	Yes	Yes
	Profiling to provide descriptive statistics about the current run overlaid with the past runs for trend analysis	Yes	Limited
	Unsupervised anomaly detection model for datasets without any predefined normalcy values	Yes	No

Required capabilities	Features	Collibra	Other vendors
 **Robust data reconciliation to reduce the risk and cost of data migrations**	Automatically validate the source and target dataset in timeline snapshots	Yes	Limited
	Data profiling and cataloging on the source systems to understand the quality of data on the source system	Yes	Yes
	Continuously monitor data movement and automate data quality checks at every point in the DataOps journey to build high quality data pipelines and data products	Yes	Limited
	Identify missing records, values and broken relationships across tables or systems as you move data	Yes	Limited
	Reconcile data between two data objects on a record by record and attribute by attribute basis to validate that data has been replicated accurately	Yes	Limited
 **Self-service data quality to put business users at the forefront of building data trust**	A unified DQ scoring system across files, databases, data warehouses and data lakes, providing a single pane of glass for DQ across all data sources	Yes	No
	Personal alerts to proactively identify and assign the data quality issues, putting business users in charge of data quality	Yes	Limited
	Democratize data quality for all technical and business users to proactively identify and assign the data quality issues	Yes	Limited
	Out-of-the-box and customizable reporting templates and dashboards (by industry, business lines, and users) to gain insight into data quality	Yes	Yes
	Interactive reports, scorecards and dashboards to help business and IT users identify, understand and monitor data quality issues	Yes	Limited
 **Enterprise data governance to build trust and compliance**	Certification of any data assets, such as data sets, reports, tables, columns and policies, to establish trust and confidence	Yes	No
	Business glossaries for standardized definitions of business terms, rules and regulations, linked to technical metadata	Yes	Limited
	Capture, reconcile and interoperate metadata relating to the data quality processes to unite data creators and data consumers	Yes	Yes
	Centralized policy management to create, maintain and implement policies and compliance across the enterprise	Yes	Limited
	Reference data management to reconcile data between systems for more accurate analysis and reporting	Yes	Yes
	Share data quality metrics with governance and analytics applications through direct integration or APIs	Yes	Yes

Required capabilities	Features	Collibra	Other vendors
 **Business-driven, collaborative workflows to resolve data quality issues faster**	Rate and comment on any data asset to provide crowdsourced context and feedback	Yes	Limited
	Collaboration for end-users to down-train, annotate and audit each DQ item	Yes	Limited
	Continuously monitor data objects for violations of business rules, generate alerts and initiate remediation workflows with the right data owners	Yes	Limited
	Platform extensions (including mobile and desktop applications) to drive adoption by delivering approved content to users where and when they need it most	Yes	Limited
	Enable business users to easily identify, quarantine, assign, escalate and resolve data quality issues	Yes	Limited
	Conditional execution of profiles for root cause analysis, for example, a profile on a production table is only run when a data quality issue is detected while loading a data warehouse	Limited	Limited
 **End-to-end data lineage to prioritize data quality issues by business impact**	Native, automated and detailed technical lineage at the table, column, transformation and SQL query level	Yes	No
	Auto-linking of technical metadata, business terms and concepts, policies and processes to add business context at scale	Yes	Limited
	End-to-end business lineage to trace data flows from source to report, helping teams narrow the focus of root cause investigations	Yes	Limited
	Capture data lineage to allow business and IT users to view changes made to data using data quality processes, for example, which upstream sources and downstream assets were impacted	Yes	Limited
 **Enterprise data catalog to focus data quality efforts on business-critical data**	Automatically classify physical data and add business context at scale	Yes	No
	Detect sensitive information (including PII, PHI) automatically across the data landscape	Yes	Limited
	Gain visibility into all data assets with full business context, helping business users to easily discover and understand data	Yes	Limited
	Data profiling statistics displayed for each data set, table and column, along with relationships	Yes	Yes
	Connect business context of critical data elements with data quality issues	Yes	Limited

Required capabilities	Features	Collibra	Other vendors
Architected for secure, enterprise-wide adoption to implement a robust data strategy	Federated, flexible operating model that adapts the data strategy to meet the unique needs of any department or business	Yes	No
	Role-based permissioning to control users' access to data assets, resources and capabilities	Yes	Yes
	Multi-tenancy security, compliance and privacy approaches with enterprise-level standards, such as SAML single sign-on, LDAP integration and role-based access management and encryption	Yes	Yes
	SaaS cloud offering for a scalable, accessible and resilient solution	Yes	Yes
	FedRamp authorized for secure cloud deployment at government agencies	Yes	Limited
	One single platform to enable a robust data strategy	Yes	Limited
	Certified APIs to allow interoperability of data quality processes or tasks with other data management tools or applications	Yes	Yes

Data quality is core to data intelligence

Collibra Data Intelligence Cloud

Govern. Trust. Access.

Data Catalog
- Catalog
- ML Curation
- Ratings / Reviews
- Data Marketplace

Data Lineage
- Business Lineage
- Technical Lineage

Data Governance
- Glossary
- Stewardship
- Data Dictionary
- Policy Manager
- Reference Data
- Data Helpdesk

Data Quality
- Continuous Monitoring
- Adaptive Rules

Data Privacy
- GDPR
- CCPA

Active Metadata Graph

Foundational Services

Flexible operating model | Native connectors | Reports and dashboards | Collibra everywhere | Workflow engine | APIs

For additional questions, contact us at:

United States
+1 646 893 3042

United Kingdom
+44 203 695 6965

All other locations
+32 2 894 79 60

By email
info@collibra.com