Data Quality Steward Certification

Industry-recognized certification for experienced Collibra Data Quality users

Using high-quality data to make informed decisions is essential for any business. Collibra Data Quality & Observability can help ensure your organization uses consistent and reliable data. Simply point it to your data source system, and it will scan and profile your data to identify data quality issues. Over time, Collibra Data Quality & Observability learns about your data and provides many profile insights, so you can automatically expand your awareness of the data. Historical profiling helps you gather a baseline. Then, the built-in machine-learning capabilities allow you to identify new data issues and generate new rules without writing a single line of code.

Collibra certifications verify your deep knowledge of Collibra products and prove you're ready to apply your expertise. Collibra certifications can help you to get hired and receive the recognition you deserve.

Who should get certified?

Those who work on the systems that provide data to data consumers and who continuously monitor datasets and pipelines to detect and remediate anomalies to deliver reliable and high-quality data.

Those responsible for:

- Proactively discovering, acknowledging, prioritizing and remediating data issues
- Implementing policies and procedures to ensure compliance with data privacy regulations such as GDPR, CCPA or HIPAA
- Setting up monitoring, profiling and rule execution for data sources
- Creating, reviewing and validating the data quality rules
- Assessing if you need to remediate datasets
- Managing the entire lifecycle of data, from validating results of initial data product creation to maintaining data retention policies and procedures for old or obsolete data
- Writing new custom rules that enforce quality across your organization
- Implementing rules across your data landscape to promote consistency and standardization in your data
- Reporting progress on data quality issues
- Creating scoring thresholds and automatically notify your team when the scores fall below these thresholds
- Leveraging pre-built reports that show data quality coverage across all technical systems and business units
- Using data quality workflows to inform multiple data owners across data sets to initiative remediation when data quality scores drop below the target threshold
- Working with stakeholders to understand their data needs, address concerns and ensure alignment with organization objectives

Who's eligible?

- 3+ months of experience using Collibra's Data Quality & Observability (Classic) product
- An understanding of the six data quality dimensions
- An understanding of your data, its metadata and how to query it using SQL
- Basic understanding of REST API usage
- Nice to have:
    - Python experience to wrap REST API invocations or invoke the DQ programmatic APIs
    - Regular Expressions [RegEx]
    - Basic understanding of Apache Spark SQL built-in function library functions (https://spark.apache.org/docs/latest/api/sql/index.html)

Exam cost
The exam costs $125 USD and is paid via credit card.

Premium Education members will have two exam attempts included as part of their membership.

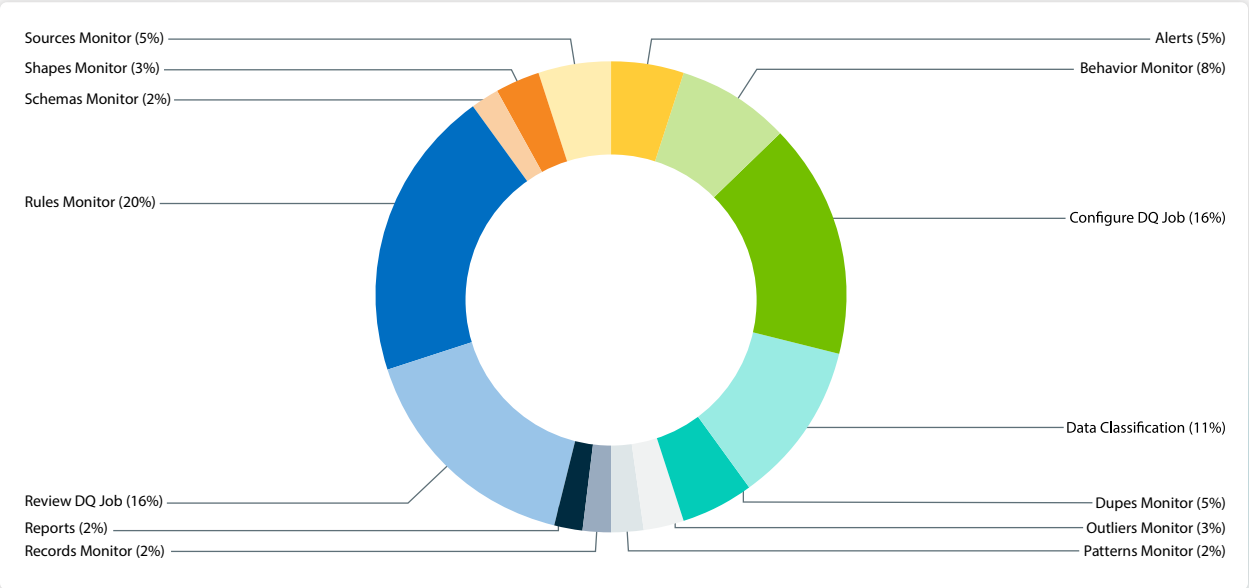
Exam details

During the exam, the test taker will be:

- Verified using a government-issued ID
- Given 90 minutes to complete 40 multiple-choice questions
- Proctored and recorded while taking the exam through Certiverse, our trusted certification vendor
- Required to utilize a lock-down browser provided by Certiverse that will limit computer/software access while taking the exam

What does the exam cover?

The exam covers 14 different domains, each with associated tasks.



Alerts
- Configure condition alerts for datasets
- Configure status alerts for datasets

Behavior monitor
- Enable or suppress manual verification for adaptive rules
- Interpret adaptive rules
- Review adaptive rules configurations

Configure data quality job
- Analyze data quality job run status and output
- Automate job runs via the scheduler
- Create dataset(s) from existing data source connections
- Execute a data quality job
- Perform data quality job back runs
- Review data quality command line to submit jobs

Data classification
- Define data classes
- Find data domains using data categories and data classes
- Find PII using data categories and data classes
- Navigate column manager
- Use data categories

Dupes monitor
- Enable duplicate detection
- Use fuzzy matching to catch duplicates

Outliers monitor
- Check for numerical and categorical outliers

Patterns monitor
- Identify uncommon patterns across columns with Patterns findings

Records monitor
- Confirm persistence of key values using Records findings

Reports
- Interpret out-of-the box data quality reports

Review data quality job
- Analyze dataset profile pages to gain insights on your data
- Describe the function of each data quality monitor
- Interpret data quality scorecard
- Resolve data quality issues on datasets
- Retrain datasets to acknowledge false positives in the results
- Validate data quality issues on datasets

Rules monitor
- Configure link ID to connect break records with source data
- Configure lookback (i.e. @t1) for dataset compare
- Create simple and freeform SQL-based rules
- Enhance and use templates to create shareable rules
- Implement complex rules with multiple datasets using joins
- Interpret data quality rule scores
- Use stat rules to get profile statistics for datasets or columns

Schemas monitor
- Review changes to underlying table structure using Schema findings

Shapes monitor
- Check for unexpected text formats using Shapes findings
- Perform sensitivity tuning for shape dimension on datasets

Sources monitor
- Configure source monitor to reconcile across connections
- Review source findings

Preparing for success

Those interested in obtaining this certification are strongly encouraged to take hands-on training in preparation for the exam. This can be completed either through self-paced learning (Path 1) or instructor-led training (Path 2). However, if you already have the recommended knowledge, you can enroll directly into the exam via university.collibra.com.
Check your readiness

Certification paths

If you decide to take hands-on training to prepare for a certification, you have two direct paths, or you can mix and match courses from each path if that better meets your goals and learning style.

Path 1: Self-paced learning

Customers: A Data Quality Steward learning path is available to all customers and partners for this certification. It takes approximately 4 hours to complete the learning path. We also recommend having had hands-on experience prior to taking the exam.

Premium Education: Use the Data Quality Steward certification path for Premium members to complete this certification. It includes the Data Quality Steward learning path, virtual lab and certification exam. These activities require approximately 6 hours to complete.

Partners: Use the Data Quality Steward certification path for Partners to complete this certification. It includes the Data Quality Steward learning path, virtual lab and certification exam. These activities require approximately 6 hours to complete.

Path 2: Instructor-led training (private, public or combination)

Recommended courses

1. DQ profiling and insights	3. DQ custom rules
2. DQ adaptive rules	4. DQ dataset management

View the course syllabi

Private training: 8 total credits are required for a group of up to 15 members from your organization to attend virtual or onsite.

Public training: 4 Education Credits are required per person to register; Premium Education members can sign up for free. View the schedule here. Note: Public training is available to customers only.

1. If you have Education Credits, view the schedule and enroll here
2. If you are a Premium Education member, view the schedule and enroll here

Ready to get started on your roadmap to success?
Talk with your Collibra representative or email certification@collibra.com to design the right certification path for you.