multisoft
Systems

Databricks Data Quality & QA Engineering Training

COURSE CONTENT

GET IN TOUCH

Multisoft Systems
B - 125, Sector - 2, Noida

(+91) 9810-306-956

info@multisoftsystems.com

www.multisoftsystems.com

About Multisoft

Train yourself with the best and develop valuable in-demand skills with Multisoft Systems. A leading certification training provider, Multisoft collaborates with top technologies to bring world-class one-on-one and certification trainings. With the goal to empower professionals and business across the globe, we offer more than 1500 training courses, which are delivered by Multisoft's global subject matter experts. We offer tailored corporate training; project Based Training, comprehensive learning solution with lifetime e-learning access, after training support and globally recognized training certificates.

About Course

Databricks Data Quality & QA Engineering training by Multisoft Systems is designed to help data professionals build, test, and maintain reliable data pipelines in modern Lakehouse environments. As organizations increasingly depend on analytics, AI, and real-time reporting, ensuring data accuracy, consistency, and completeness has become mission-critical.

Module 1: Introduction to Data Quality & QA Engineering

- ✓ Overview of data quality challenges in modern data platforms
- ✓ Role of QA engineering in data pipelines
- ✓ Key data quality dimensions: accuracy, completeness, consistency, timeliness
- ✓ Introduction to Databricks Lakehouse architecture

Module 2: Databricks Platform & Lakehouse Fundamentals

- ✓ Databricks workspace overview
- ✓ Lakehouse architecture concepts
- ✓ Understanding Delta Lake fundamentals
- ✓ Data ingestion patterns in Databricks

Module 3: Data Profiling and Quality Assessment

- ✓ Data profiling techniques
- ✓ Identifying anomalies, nulls, and outliers
- ✓ Data completeness and uniqueness checks
- ✓ Establishing baseline data quality metrics

Module 4: Data Validation and Business Rule Implementation

- ✓ Designing data validation rules
- ✓ Column-level and row-level quality checks
- ✓ Handling invalid and bad records
- ✓ Implementing reusable validation frameworks

Module 5: Delta Lake Quality Controls

- ✓ Delta Lake constraints and expectations
- ✓ Schema enforcement and evolution

- ✓ Handling late-arriving and corrupted data
- ✓ Auditability and version control using Delta Lake

Module 6: Testing Data Pipelines

- ✓ QA strategies for batch pipelines
- ✓ Testing streaming pipelines
- ✓ Validating transformations and aggregations
- ✓ Regression testing for data pipelines

Module 7: Monitoring, Observability, and Alerts

- ✓ Monitoring pipeline health and data freshness
- ✓ Detecting data drift and schema changes
- ✓ Setting up alerts and notifications
- ✓ Logging and error tracking best practices

Module 8: Automation and CI/CD for Data QA

- ✓ Automating data quality checks
- ✓ Integrating QA into CI/CD pipelines
- ✓ Version control and deployment strategies
- ✓ Environment management (Dev, Test, Prod)

Module 9: Performance Optimization and Scalability

- ✓ Optimizing quality checks for large datasets
- ✓ Managing performance overhead
- ✓ Scaling QA frameworks in enterprise environments
- ✓ Cost considerations and best practices

Module 10: Real-World Use Cases and Best Practices

- ✓ End-to-end data quality implementation scenarios

- ✓ Common production issues and resolutions
- ✓ Industry best practices for data QA
- ✓ Designing a robust data quality framework