

Advanced Apache Iceberg

COURSE CONTENT

GET IN TOUCH



Multisoft Systems
B - 125, Sector - 2, Noida



(+91) 9810-306-956



info@multosoftsystems.com



www.multosoftsystems.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

Advanced Apache Iceberg Training is designed for experienced data professionals who want to implement, optimize, and manage Apache Iceberg in enterprise data lake and lakehouse environments. The course explores advanced table architecture, metadata optimization, streaming workloads, governance, performance engineering, and production deployment strategies, enabling learners to build scalable and reliable analytical data platforms.

Module 1: Apache Iceberg Architecture Deep Dive

- ✓ Internal Architecture
- ✓ Metadata Layers
- ✓ Table Specifications
- ✓ Manifest Structure
- ✓ Snapshot Lifecycle
- ✓ Enterprise Architecture

Module 2: Advanced Catalog Management

- ✓ Hive Catalog
- ✓ Hadoop Catalog
- ✓ REST Catalog
- ✓ AWS Glue Catalog
- ✓ Nessie Catalog
- ✓ Catalog Configuration
- ✓ Multi-Catalog Strategy

Module 3: Enterprise Table Design

- ✓ Advanced Table Modeling
- ✓ Data Organization
- ✓ Partition Planning
- ✓ Hidden Partitioning
- ✓ Table Properties
- ✓ Storage Strategies

- ✓ Enterprise Design Standards

Module 4: Advanced Schema Evolution

- ✓ Complex Schema Changes
- ✓ Nested Structures
- ✓ Data Compatibility
- ✓ Schema Validation
- ✓ Evolution Strategies
- ✓ Production Migration
- ✓ Best Practices

Module 5: Snapshot, Branching and Tagging

- ✓ Snapshot Management
- ✓ Branch Creation
- ✓ Data Tagging
- ✓ Rollback Operations
- ✓ Version Comparison
- ✓ Historical Analysis
- ✓ Release Management

Module 6: Performance Engineering

- ✓ Metadata Optimization
- ✓ File Size Optimization
- ✓ Manifest Optimization

- ✓ Query Planning
- ✓ Predicate Pushdown
- ✓ Partition Pruning
- ✓ Cost Optimization

Module 7: Streaming Data with Iceberg

- ✓ Streaming Architecture
- ✓ Apache Flink Integration
- ✓ Structured Streaming
- ✓ Streaming Writes
- ✓ Incremental Reads
- ✓ Change Data Capture (CDC)
- ✓ Real-Time Analytics

Module 8: Apache Spark Integration

- ✓ Spark SQL
- ✓ DataFrame Operations
- ✓ Batch Processing
- ✓ Spark Optimization
- ✓ Data Writing
- ✓ Job Optimization
- ✓ Performance Monitoring

Module 9: Security and Governance

- ✓ Data Governance
- ✓ Access Policies
- ✓ Encryption
- ✓ Authentication
- ✓ Authorization
- ✓ Audit Logging
- ✓ Compliance Management

Module 10: Maintenance and Operations

- ✓ File Compaction
- ✓ Snapshot Expiration
- ✓ Metadata Cleanup
- ✓ Table Optimization
- ✓ Maintenance Scheduling
- ✓ Operational Automation
- ✓ Health Checks

Module 11: Cloud and Lakehouse Integration

- ✓ Cloud Object Storage
- ✓ Lakehouse Architecture
- ✓ Multi-Cloud Deployment
- ✓ Data Sharing
- ✓ Enterprise Integration

- ✓ Analytics Platforms
- ✓ Scalability Planning

Module 12: Monitoring and Troubleshooting

- ✓ Performance Monitoring
- ✓ Resource Utilization
- ✓ Log Analysis
- ✓ Error Resolution
- ✓ Bottleneck Identification
- ✓ Diagnostic Tools
- ✓ Operational Best Practices

Module 13: Enterprise Deployment

- ✓ Production Deployment
- ✓ Migration Planning
- ✓ High Availability
- ✓ Disaster Recovery
- ✓ Backup Strategies
- ✓ Upgrade Management
- ✓ Enterprise Readiness

Module 14: Enterprise Capstone Project

- ✓ Large-Scale Lakehouse Implementation
- ✓ Performance Optimization Project

- ✓ Streaming Pipeline Deployment
- ✓ Governance Implementation
- ✓ Multi-Engine Integration
- ✓ Production Validation
- ✓ Enterprise Best Practices
- ✓ Final Capstone Project