

Apache Iceberg Fundamentals

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

Apache Iceberg Fundamentals Training introduces professionals to a modern open table format that simplifies large-scale data lake management. The course explains how Apache Iceberg improves data consistency, supports evolving schemas, optimizes query performance, and enables reliable analytics across distributed data platforms. Learners gain practical exposure to building efficient and maintainable data lake architectures.

Module 1: Introduction to Apache Iceberg

- ✓ Apache Iceberg Overview
- ✓ Modern Data Lakes
- ✓ Open Table Format
- ✓ Architecture
- ✓ Core Components
- ✓ Business Benefits

Module 2: Iceberg Table Architecture

- ✓ Table Metadata
- ✓ Manifest Files
- ✓ Snapshot Structure
- ✓ Catalog Services
- ✓ Table Lifecycle
- ✓ Metadata Management
- ✓ Storage Architecture

Module 3: Creating and Managing Iceberg Tables

- ✓ Table Creation
- ✓ Table Configuration
- ✓ Data Loading
- ✓ Table Properties
- ✓ Partition Specification
- ✓ Table Maintenance

- ✓ Table Operations

Module 4: Schema and Partition Evolution

- ✓ Schema Evolution
- ✓ Column Operations
- ✓ Data Compatibility
- ✓ Hidden Partitioning
- ✓ Partition Evolution
- ✓ Metadata Updates
- ✓ Best Practices

Module 5: Snapshots and Time Travel

- ✓ Snapshot Management
- ✓ Snapshot Isolation
- ✓ Time Travel Queries
- ✓ Rollback Operations
- ✓ Historical Data Access
- ✓ Version Management
- ✓ Data Recovery

Module 6: Data Processing Integration

- ✓ Apache Spark Integration
- ✓ Apache Flink Integration
- ✓ Trino Integration

- ✓ Presto Integration
- ✓ Hive Compatibility
- ✓ Cloud Storage Integration
- ✓ Data Processing Workflows

Module 7: Query Optimization

- ✓ Predicate Pushdown
- ✓ Partition Pruning
- ✓ File Skipping
- ✓ Metadata Optimization
- ✓ Query Planning
- ✓ Performance Tuning
- ✓ Optimization Strategies

Module 8: Maintenance and Performance

- ✓ File Compaction
- ✓ Metadata Cleanup
- ✓ Snapshot Expiration
- ✓ Storage Optimization
- ✓ Resource Management
- ✓ Performance Monitoring
- ✓ Operational Best Practices

Module 9: Security and Governance

- ✓ Access Control
- ✓ Data Governance
- ✓ Catalog Security
- ✓ Audit Capabilities
- ✓ Data Integrity
- ✓ Compliance
- ✓ Enterprise Best Practices

Module 10: Enterprise Implementation Project

- ✓ Iceberg Deployment
- ✓ Lakehouse Architecture
- ✓ Performance Optimization
- ✓ Data Migration
- ✓ Operational Validation
- ✓ Enterprise Use Cases
- ✓ Final Project