Apache Iceberg Training

COURSE CONTENT

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 is a high-performance table format for data lakes, designed to handle petabyte-scale data with ease and efficiency. Multisoft Systems offers a comprehensive Apache Iceberg training program that equips data engineers, analysts, and IT professionals with the knowledge and skills needed to optimize data management and analytics in modern data lake environments.

Module 1: Understanding Apache Iceberg

- ✓ Evolution of Data Platforms
- ✓ Understanding Data Lakes and Technologies available
- ✓ Challenges with Data Lakes
- ✓ Introduction to Apache Iceberg
- ✓ Benefits of Apache Iceberg
- ✓ Apache Iceberg vs Delta Lake vs Hudi
- ✓ When to choose Apache Iceberg over other formats for data lake storage?

Module 2: Apache Iceberg Architecture

- ✓ Overview of Apache Iceberg architecture
- ✓ Various Apache Iceberg Components
- ✓ How does Apache Iceberg handle metadata and data versioning?
- ✓ Integration of Apache Iceberg with key data processing engines like Starburst, Spark

Module 3: Setting Up Apache Iceberg

- ✓ Installation and setup of Apache Iceberg
- ✓ Configuring metadata storage for Apache Iceberg tables

Module 4: Creating Apache Iceberg Tables

- ✓ Apache Iceberg table structure
- ✓ Step-by-step guide to creating Apache Iceberg tables using

  - Apache Spark

  - Presto/Trino on Starburst

  - Hive

Module 5: Writing and Reading Data

- ✓ Inserting data into Apache Iceberg tables

  - Batch inserts

  - Streaming inserts

  - Upserts

- ✓ Efficiently querying Apache Iceberg tables
- ✓ Demonstrating how Apache Iceberg's data layout optimization enhances query performance

Module 6: Internals of Apache Iceberg

- ✓ The Iceberg Catalog
- ✓ The Metadata Layer

  - Metadata File

  - Manifest List

  - Manifest File

- ✓ The Data Layer
- ✓ A look under the covers when CRUDing

Module 7: Management, Monitoring and Optimization

- ✓ Managing schema evolution
- ✓ Enabling partitions in Iceberg Tables

  - Hidden Partitioning

  - Partition Layer Evolution

- ✓ Understanding Time Travel

- ✓ Version Rollback
- ✓ Data Compaction
- ✓ Metrics and Alerts
- ✓ Monitoring Iceberg Tables