AMBIENT-IT

+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated 04/26/2024

Sign up

Apache Iceberg training

3 days (21 hours)

Presentation

Looking for a table format for all your analytical data? Our Apache Iceberg training course will help you understand how to use it for all your statistical analyses.

This open source framework offers efficient management of massive tables, with features such as transactional integrity and cross-platform compatibility.

During this course, you'll learn about data traceability, which enables you to track the complete history of modifications.

This training should give you expertise in large-scale data management, and you should be able to design and manage massive data.

As with all of them, the Apache Iceberg training course will be presented with its latest features (at the time of writing).

Objectives

- Understanding the fundamentals of Apache Iceberg
- Learn techniques for optimizing table performance
- Write and execute read and write requests in Iceberg
- Understanding catalog management and data security
- Master Iceberg architecture, data layers and metadata

Target audience

- Data scientists
- Data analysts

- Data engineers
- Architects

Prerequisites

- Understanding the fundamental concepts of data lakes and databases
- Knowledge of programming (Python) and SQL

APACHE ICEBERG TRAINING PROGRAM

Introduction to Apache Iceberg

- Initiation
- Data warehouse
- Data lake
- Discover what Hive is
- Architecture

table formats

- The data layer
- The metadata layer
  - Manifest files
  - Manifest lists
  - Metadata files
  - Puffin files
- Catalog

Read/write request life cycle

- Writing queries in Apache Iceberg
  - Create table
  - Insert query
  - Merge query
- Reading requests in Apache Iceberg
  - The SELECT
  - The time travel query Optimizing

table performance

- Compaction

- Practical with compaction
  - Strategies
  - Automation
- Sorting
- Order Z
- Partitioning
  - Hidden partitioning
  - Scores evolution
  - Other partitioning considerations
- Copy on write and merge on read

Catalogues

- Iceberg catalog requirements
- Catalog comparison
- Catalog migration
  - Using the Apache Iceberg catalog migration CLI
  - Using an Apache

Spark engine

- Configuration
- Data definition language operations
- Reading data
- Iceberg table maintenance procedures

Dremio's SQL query engine

- Iceberg table maintenance
- Rewriting data files
- Rewriting Colle AWS

manifests

- Configuration
  - Designing a Glue database
  - Glue ETL job configuration
- Table creation using the Glue data catalog
  - Read the table
  - Insert data

Governance and security

- Securing data files

- Securing and governing the semantic layer
  - Semantic layer best practices
  - Drémio
  - Trino
- Catalog security and control

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.