Updated on 12/07/2024

Sign up

Starburst training

3 days (21 hours)

Presentation

Our Starburst training course will enable you to master the functionalities of this data lakehouse and greatly improve your data analysis processes. Starburst's high-performance data architecture enables rapid processing and analysis.

This training program will help you understand StarBurst's architecture and terminology, as well as its use with Trino, the SQL analysis engine.

Our training will put your knowledge to the test with practical exercises. You'll have a secure, scalable data warehouse and the governance you need to ensure that your teams of engineers and analysts never run into problems.

Like all our training courses, it will be run on the latest version of the tool: Trino 448.

Objectives

- Configuring Starburst for your data flows
- Understanding the use of SQL in Starburst
- Managing and exploiting your Starburst database

Target audience

- **Data Analysts**
- Data scientists

Prerequisites

- Basic knowledge of SQL and data manipulation
- Basic knowledge of Linux/Unix commands

STARBURST TRAINING PROGRAM

INTRODUCTION TO STARBURST

- Starburst overview
- Key components and architecture
- Benefits and use cases
- Introduction to platforms
- Starburst connectors overview

SQL FUNDAMENTALS AND PARALLELIZATION

- SQL basics in Starburst
- Query parallelization techniques
- Using aggregations and conditional queries
- SQL syntax and joins
- Understanding data definition language (DDL)

ADVANCED STARBURST FEATURES

- Explore Starburst feature tutorials
- Fault-tolerant execution and high availability
- Data federation and cluster scaling
- Introduction to data products and access control
- Schematics discovery and universal search

DATABASE MANAGEMENT AND STORAGE

- Database types: relational vs. non-relational
- Data warehouse and data lake management
- Row- and column-oriented file formats
- Object storage in the cloud and metadata
- Comparison of modern table formats: Hive vs Iceberg

DESIGN AND AUTOMATION OF DATA PIPELINES

- ETL principles (Extract, Transform, Load)
- Automating and integrating data pipelines with Starburst
- Construction of data sets organized as products
- Best practices for managing data pipelines

INTEGRATION WITH OTHER SOLUTIONS

- Understanding of ETL tools such as dbt and Informatica
- Using BI tools such as Tableau and Power BI
- Basics of governance and cataloguing tools
- Practical integration with configuration examples

TROUBLESHOOTING AND SAFETY

- Identify and resolve query performance problems
- Cluster scaling strategies for optimization
- Data security and role- and attribute-based access control
- Best practices for system maintenance and safety

PRACTICAL WORKSHOPS AND CASE STUDIES

- Setting up a Starburst environment
- Complex query execution and optimization
- Troubleshooting simulation based on real-life scenarios
- Create analysis dashboards with BI integration

REVIEW AND FINAL ASSESSMENT

- Review of Starburst's main concepts and features
- Knowledge test with exam questions based on real cases

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.