Faysal Aziz

Introduction to Advanced ClickHouse

1-day to boost your Clickhouse expertise

FA-0254 · 1-day

Data & Analytics

Introduction

In today's data-driven world, businesses and organizations are constantly seeking ways to handle large volumes of data efficiently and cost-effectively. ClickHouse, an open-source columnar database management system, has emerged as a powerful solution for real-time analytics. Its ability to process billions of rows per second per server makes it ideal for handling big data workloads.

This advanced ClickHouse training is designed to deepen your understanding of ClickHouse, enabling you to harness its full potential for high-performance data analysis and real-time querying.

Learning outcomes

By the end of this training, participants will be able to:

- Understand the advanced architecture and internals of ClickHouse.
- Optimize ClickHouse configurations for enhanced performance.
- Implement complex queries and advanced data processing techniques.
- Use ClickHouse for real-time data analytics and streaming.
- Manage ClickHouse clusters and ensure high availability.
- Utilize the MergeTree family engines effectively for various use cases.

Prerequisites

To get the most out of this training, participants should have:

- Basic understanding of database concepts.

- Familiarity with SQL.
- Prior experience with ClickHouse basics.
- Knowledge of Linux command-line operations.

Training outline

1. Introduction to ClickHouse

Overview of ClickHouse

1. History and development of ClickHouse
2. Key features and benefits

2. Use cases of ClickHouse in various industries

2. ClickHouse Architecture

Columnar storage format

1. Differences between row-based and column-based storage
2. Benefits of columnar storage for analytics

Data partitioning and sharding

1. Understanding data partitioning
2. Sharding strategies and their impact on performance

ClickHouse storage engines

1. Overview of MergeTree family engines
2. Using Log, StripeLog, and other engines

3. Installation and Configuration

Installing ClickHouse

1. Installation on various platforms (Linux, Docker, Kubernetes)
2. Post-installation setup

Configuring ClickHouse for performance

1. Understanding configuration files
2. Tuning server settings (memory, CPU, disk I/O)
3. Optimizing network settings

4. Advanced Data Modeling

Creating and managing tables

1. Advanced table creation options
2. Partitioning and primary keys

Materialized views

1. Creating and using materialized views
2. Use cases and performance considerations

Nested data structures

1. Working with arrays and nested data types

5. ClickHouse and machine learning

1. Supervised Machine Learning
2. Using ClickHouse for ML model training and inference
3. Integrating with data science tools like Jupyter and Pandas

6. Query Optimization

Understanding the query execution process

1. Parsing, planning, and execution stages
2. Query profiling and analysis

Handling large datasets

1. Efficiently querying large tables
2. Using external storage and distributed queries

7. MergeTree Family Engines

Overview of MergeTree engines

1. MergeTree, ReplacingMergeTree, SummingMergeTree, etc.

Creating and managing MergeTree tables

1. Advanced table configurations

Performance optimization with MergeTree

1. Tuning settings specific to MergeTree

Practical use cases of MergeTree engines

1. Use cases and best practices

Maintenance and management

1. Ensuring data consistency and performance

8. Cluster Management and High Availability

Setting up a ClickHouse cluster

1. Cluster architecture and components
2. Configuring replication and distributed tables

Ensuring high availability

1. Strategies for fault tolerance and redundancy
2. Monitoring and managing cluster health

Scaling ClickHouse

1. Adding and removing nodes
2. Load balancing and resource management

By the end of this training, participants will have a comprehensive understanding of advanced ClickHouse functionalities, enabling them to efficiently manage and analyze large datasets in real-time.

© 2026 Faysal Aziz · Kuala Lumpur