



Академия АйТи
a Softline Company



Проектирование и разработка в ClickHouse

Код курса: CLH-DEV

Проектирование и разработка в ClickHouse

Код курса: CLH-DEV

Длительность	32 ак. часа
Формат	
Разработчик курса	Академия АйТи
Тип	Учебный курс
Способ обучения	Под руководством тренера

О курсе

Курс предназначен для технических специалистов, работающих с базами данных и заинтересованных в освоении основ администрирования и настройки ClickHouse — высокопроизводительной аналитической базы данных. Курс дает слушателям необходимые навыки для работы с ClickHouse, включая установку, конфигурирование, управление данными и мониторинг производительности.

Подробная информация

Профиль аудитории:

- разработчики и администраторы ClickHouse
- IT-специалисты, работающие с высоконагруженными аналитическими системами
- архитекторы данных, проектирующие инфраструктуру ClickHouse
- все, кто хочет изучить расширенные возможности ClickHouse

Цели:

- освоить архитектуру и основные принципы работы ClickHouse для разработки и администрирования высоконагруженных систем
- изучить методы проектирования эффективных схем данных для аналитики, учитывая особенности работы с большими объемами данных
- научиться оптимизировать производительность запросов в ClickHouse, включая использование индексов и оптимизацию агрегации
- овладеть навыками настройки и управления репликацией, шардированием и кластеризацией для обеспечения отказоустойчивости и масштабируемости
- разработать и внедрить стратегии мониторинга и диагностики для повышения стабильности и быстродействия системы
- освоить интеграцию ClickHouse с внешними системами, такими как Kafka, для обеспечения потоковой обработки и интеграции с другими СУБД
- изучить лучшие практики работы с материализованными представлениями, проекциями и методами их оптимизации для аналитических задач
- научиться применять методы резервного копирования и восстановления данных в ClickHouse,

чтобы обеспечить безопасность и доступность данных

Предварительные требования:

- базовые знания SQL: опыт работы с запросами, понимание основных операций (SELECT, INSERT, UPDATE, DELETE)
- опыт работы с базами данных: понимание принципов работы с реляционными и NoSQL базами данных
- основы Linux: опыт работы с командной строкой и базовыми административными задачами в Linux-средах
- знания в области аналитики данных: базовые знания методов обработки и анализа больших объемов данных
- знание принципов работы с распределенными системами (желательно): знакомство с концепциями масштабируемости и отказоустойчивости будет полезно, но не обязательно

По завершении курса слушатели смогут:

- развертывать и настраивать ClickHouse-кластеры, опираясь на понимание его архитектуры и ключевых принципов работы
- проектировать эффективные схемы данных для аналитических задач, адаптируя их под большие объёмы информации
- оптимизировать запросы и повышать производительность с помощью индексов, грамотной агрегации и профилирования
- настраивать репликацию, шардирование и кластеризацию, обеспечивая отказоустойчивость и масштабируемость системы
- внедрять стратегии мониторинга и диагностики, чтобы своевременно выявлять узкие места и поддерживать стабильность работы

Программа курса

Модуль 1: Введение в архитектуру и возможности ClickHouse

- Преимущества ClickHouse для аналитики больших данных.
- Архитектура и особенности работы ClickHouse.
- Обзор терминологии и ключевых понятий.
- Сравнение с другими СУБД (PostgreSQL, MySQL и т.д.).

Модуль 2: Проектирование и работа с данными в ClickHouse

- Типы данных: простые и сложные (Array, Tuple, Map, Nested, Nullable).
- Специфика работы с данными в ClickHouse.
- Преимущества и ограничения типов данных для работы с большими объемами.
- Использование движков таблиц: MergeTree, ReplacingMergeTree, SummingMergeTree, AggregatingMergeTree.

Модуль 3: Язык запросов в ClickHouse

- Основы языка запросов ClickHouse.
- Применение фильтрации, агрегации и группировки данных.

- Работа с джойнами, включая ассоциативные джойны и ASOF джойны.
- Оптимизация запросов с использованием WHERE, PREWHERE, индексов и кластерных запросов.

Модуль 4: Оптимизация запросов и производительности

- Профилирование запросов с помощью EXPLAIN и таблиц system.query_log.
- Методы оптимизации: индексы, использование примитивов для фильтрации, оптимизация агрегации.
- Оптимизация работы с большими объемами данных (управление памятью, многозадачность).
- Подходы к быстрому анализу и обработке данных в ClickHouse.

Модуль 5: Репликация и кластеризация в ClickHouse

- Основы настройки и создания кластеров ClickHouse.
- Репликация данных, настройка отказоустойчивости.
- Управление шардированием и распределение нагрузки.
- Использование ядра ReplicatedMergeTree для репликации.

Модуль 6: Проекция и материализованные представления

- Что такое проекции и как они ускоряют запросы.
- Создание и использование материализованных представлений.
- Управление обновляемыми материализованными представлениями для преобразования данных.
- Применение оконных функций и их влияние на производительность.

Модуль 7: Интеграция ClickHouse с внешними системами

- Интеграция с Kafka для потоковой обработки данных.
- Взаимодействие ClickHouse с другими СУБД (PostgreSQL, MySQL, MongoDB).
- Использование внешних таблиц для интеграции и выгрузки данных.
- Настройка интеграции с системами мониторинга и аналитики.

Модуль 8: Работа с большими данными

- Вставка данных: механика загрузки данных в ClickHouse.
- Использование различных типов вставки и опций слияния.
- Применение TTL для управления временем жизни данных.
- Обработка непрерывной загрузки данных и использование интеграционных таблиц.

Модуль 9: Мониторинг и диагностика производительности

- Использование системных таблиц для мониторинга работы ClickHouse.
- Как диагностировать проблемы с производительностью: анализ запросов, времени ответа, загрузки процессора и памяти.
- Применение инструментов мониторинга, таких как Prometheus и Grafana, для отслеживания состояния ClickHouse.
- Разбор типичных проблем и узких мест в производительности.

Модуль 10: Бэкапы и восстановление данных

- Стратегии резервного копирования и восстановления данных в ClickHouse.
- Настройка автоматического бэкапа данных.
- Управление версиями и восстановление после сбоев.

Модуль 11: Лучшие практики и кейс-стади

- Реальные примеры использования ClickHouse для аналитики больших данных.
- Кейсы из практики по проектированию архитектуры ClickHouse.
- Лучшие практики по оптимизации и масштабированию ClickHouse.

[Посмотреть расписание курса и записаться на обучение](#)

Обращайтесь по любым вопросам

к менеджерам Академии АйТи

+7 (495) 150 96 00 | academy@academyit.ru