

Полная программа

ClickHouse для инженеров и архитекторов БД

ClickHouse для инженеров и архитекторов БД

Длительность курса: 154 часа

Модуль 1. Знакомство с ClickHouse

Тема 1

Аналитические движки

Цель занятия

понять, как эффективно обучаться на курсе; разбираться в принципах устройства аналитических движков и их практического использования.

Краткое содержание

обсуждение программы курса; аналитические СУБД как класс; ключевые принципы устройства аналитических СУБД; знакомство с ClickHouse; подготовка инфраструктуры.

Тема 2

Область применения и первое представление // ДЗ

Цель занятия

узнать, что такое ClickHouse и область его применения; понять, как начать работу с ClickHouse и, где искать ответы на свои вопросы.

Краткое содержание

знакомство с ClickHouse; сильные и слабые стороны БД; комьюнити и Knowledge Base.

Домашние задания

Адаптация ClickHouse

Цель: привнести при успешной адаптации ClickHouse, описать закрытую бизнес-потребность.

Тема 3

Развертывание и базовая конфигурация, интерфейсы и инструменты // ДЗ

Цель занятия

рассмотреть различные варианты установки ClickHouse; изучить файл конфигурации ClickHouse; понять принципы конфигурирования и настройки данной СУБД; познакомиться с популярными вариантами доступа к БД ClickHouse.

Краткое содержание

cloud-based / on-premise варианты установок; базовая настройка для начала работы работы с параметрами; консольный клиент; клиентские библиотеки; библиотеки для интеграции; визуальные интерфейсы.

Домашние задания

Установка ClickHouse

Цель: настроить ClickHouse с базовой конфигурацией, оценить производительность до и после оптимизации параметров системы.

Тема 4

MergeTree и типы данных

Цель занятия

познакомиться с основными принципами работы MergeTree в ClickHouse; разработать различные типы данных; показать примеры использования различных типов данных и их влияние на производительность запросов.

Краткое содержание

введение в MergeTree и его принципы работы; разбор различных типов данных, их особенностей; оптимизация запросов с учетом выбора правильных типов данных.

Модуль 2. Работа с ClickHouse

Тема 1

Язык запросов SQL // ДЗ

Цель занятия

познакомиться с языком запросов ClickHouse; изучить именовы и спецификации; рассмотреть "подкапотную" часть ClickHouse.

Краткое содержание

create, insert, select, update*, delete*; загрузка datasets; тестирование производительности ClickHouse.

Домашние задания

Работа с SQL в ClickHouse

Цель: освоить создание и модификацию таблиц, выполнение базовых операций с данными и управление партициями в ClickHouse.

Тема 2

Функции для работы с типами данных, агрегатные функции и UDF // ДЗ

Цель занятия

изучить функции для работы с типами данных и агрегатные функции в ClickHouse; рассмотреть специфика создания UDF в данной БД.

Краткое содержание

изучение функций для работы с различными типами данных в ClickHouse, такими как строки, числа, даты и т.д.; практическое применение агрегатных функций в ClickHouse для обработки данных в базе данных; создание и использование пользовательских функций (UDF) в ClickHouse для решения конкретных задач.

Домашние задания

Обработка данных

Цель: освоить агрегатные функции, функции работы с типами данных и UDF в ClickHouse.

Тема 3

Движки MergeTree Family // ДЗ

Цель занятия

изучить и попрактиковаться в применении класса MergeTree Family; разработать принципы работы таких движков, как с логической, так и с физической точек зрения.

Краткое содержание

MergeTree; ReplacingMergeTree; SummingMergeTree; AggregatingMergeTree; CollapsingMergeTree; VersionedCollapsingMergeTree; GraphLikeMergeTree; практика применения и типовые задачи.

Домашние задания

Движки MergeTree

Цель: идентифицировать движки таблиц ClickHouse по их поведению при вставках и выборке данных.

Тема 4

Индексы в ClickHouse

Цель занятия

ответить на вопрос что такое индекс; корректно задавать индексы при разработке и сопровождении схем БД; выявлять потребность во вторичных индексах.

Краткое содержание

что такое индекс; первичный индекс в ClickHouse; вторичные индексы; использование индексов; рефлексия.

Тема 5

Другие движки

Цель занятия

изучить работу с другими движками БД; рассмотреть их практическое применение.

Краткое содержание

работа с движками; варианты базовой интеграции.

Тема 6

Джоины и агрегации // ДЗ

Цель занятия

рассмотреть базовые и дополнительные типы соединений, доступных в ClickHouse; изучить параметры, которые влияют на рассматриваемые понятия.

Краткое содержание

базовые и дополнительные типы соединения ClickHouse; тюнинг параметров БД для агрегаций и джоиннов; рекомендации по оптимизации производительности.

Домашние задания

Джоины и агрегации

Цель: применить различные типы соединений и агрегирующие операции для анализа данных на датасете IMDB.

Тема 7

Словари, оконные и табличные функции // ДЗ

Цель занятия

разработать использование словарей, оконных и табличных функций в ClickHouse; научиться применять словари для эффективного хранения и обработки справочной информации; изучить принципы работы оконных, табличных функций и их применение для аналитических задач.

Краткое содержание

введение в словари и их использование для оптимизации запросов. обзор оконных функций, их синтаксис и примеры использования. практические примеры работы с оконными функциями для аналитических задач. ознакомление с табличными функциями в ClickHouse и их применение.

Домашние задания

Словари

Цель: закрепить навыки работы со словарями и оконными функциями для выборки данных и вычисления агрегационной суммы.

Тема 8

Сессия Q&A

Цель занятия

получить ответы на вопросы по темам занятий и домашних заданий.

Краткое содержание

типичные ошибки при выполнении ДЗ; преподаватели ответят на ваши вопросы.

Модуль 3. Масштабирование и манипуляции с данными

Тема 1

Проекции и материализованные представления // ДЗ

Цель занятия

ознакомиться с концепцией проекций и материализованных представлений в ClickHouse; изучить основные типы проекций и их применение для оптимизации запросов; научиться создавать и использовать материализованные представления в ClickHouse.

Краткое содержание

введение в проекции данных: понятие, назначение, виды проекций; оптимизация запросов с помощью проекций: выбор типа проекции, настройка параметров, создание и использование материализованных представлений: преимущества, особенности работы, примеры использования.

Домашние задания

Проекции и материализованные представления

Цель: создать и сравнить проекцию и материализованное представление для автоматической агрегации данных о продажах.

Тема 2

Репликация и другие фоновые процессы // ДЗ

Цель занятия

разобрать основные принципы репликации данных в ClickHouse; изучить процессы фоновой обработки данных, такие как компакция, мерджинг и удаление данных; научиться настраивать и мониторить репликацию и другие фоновые процессы в ClickHouse.

Краткое содержание

введение в репликацию данных: основные понятия, модели и сценарии применения; основные процессы фоновой обработки данных в ClickHouse: мерджинг, удаление данных; настройка и мониторинг репликации данных в ClickHouse.

Домашние задания

Репликация и удаление

Цель: развернуть реплицированную таблицу и настроить политику хранения данных с автоматическим удалением устаревших записей.

Тема 3

Шардирование и распределенные запросы // ДЗ

Цель занятия

познакомиться с концепцией шардирования данных в ClickHouse; изучить принципы шардирования запросов и их применение для улучшения производительности; научиться настраивать и оптимизировать распределенные запросы в ClickHouse.

Краткое содержание

введение в шардирование: определение, принцип работы, виды шардирования; распределенные запросы в ClickHouse: основные принципы, механизмы распределения запросов; настройка и оптимизация распределенных запросов: выбор стратегии шардирования, оптимизация запросов для распределенной обработки.

Домашние задания

Шардирование

Цель: шардировать свой инстанс.

Тема 4

Сессия Q&A

Цель занятия

получить ответы на вопросы по темам занятий и домашних заданий.

Краткое содержание

типичные ошибки при выполнении ДЗ; преподаватели ответят на ваши вопросы.

Тема 5

Мутация данных и манипуляции с партициями // ДЗ

Цель занятия

познакомиться с понятием мутации данных в ClickHouse и методами их выполнения; изучить основные операции манипуляции с партициями данных в ClickHouse; использовать методы манипуляции с партициями для эффективного управления данными в ClickHouse.

Краткое содержание

введение в мутации данных: определение, типы мутаций (вставка, обновление, удаление), особенности и применение; манипуляции с партициями данных: создание, удаление, перемещение, слияние партиций в ClickHouse; практические примеры использования мутаций и манипуляций с партициями для оптимизации работы с данными.

Домашние задания

Мутации данных и манипуляции с партициями

Цель: применить мутации для изменения данных и отработать операции удаления партиций по дате.

Модуль 4. Управление ресурсами

Тема 1

RBAC контроль доступа, квоты и ограничения // ДЗ

Цель занятия

ознакомиться с концепцией RBAC (Role-Based Access Control) для управления доступом к данным в ClickHouse; изучить возможности установки квот и ограничений для контроля использования ресурсов в ClickHouse; нарабатывать практические навыки настройки RBAC, установки квот и ограничений в ClickHouse.

Краткое содержание

введение в RBAC: определение, принципы работы, роли и привилегии в контексте ClickHouse; управление доступом: создание пользователей, назначение ролей, настройка прав доступа к данным; квоты и ограничения: установка лимитов по использованию ресурсов (CPU, память, дисковое пространство); ограничение количества запросов и параллельных соединений; практические примеры настройки RBAC, установки квот и ограничений в ClickHouse.

Домашние задания

Контроль доступа

Цель: освоить базовые операции управления пользователями и ролями;

Тема 2

Storage Policy и резервное копирование // ДЗ

Цель занятия

познакомиться с понятием Storage Policy в ClickHouse и его роль в организации хранения данных; изучить методы резервного копирования данных в ClickHouse для обеспечения их сохранности и восстановления; получить практические навыки настройки Storage Policy и резервного копирования в ClickHouse.

Краткое содержание

введение в Storage Policy: определение, назначение, принципы работы, типы политик хранения данных в ClickHouse; организация резервного копирования: методы создания резервных копий данных, стратегии резервного копирования, хранение и восстановление резервных копий; настройка Storage Policy: определение параметров хранения данных, оптимизация процесса хранения и доступа к данным. 4. Практические примеры настройки Storage Policy и резервного копирования в ClickHouse.

Домашние задания

Storage Policy и резервное копирование

Цель: изучить возможности резервного копирования в ClickHouse.

Тема 3

Метрики и мониторинг. Логирирование // ДЗ

Цель занятия

собирать метрики, логи; изучить доступные инструменты мониторинга; работать с БД и ресурсами BM.

Краткое содержание

оптимизация производительности; методики мониторинга; Grafana; Prometheus.

Домашние задания

Мониторинг

Цель: настроить мониторинг и оптимизировать производительность ClickHouse.

Тема 4

Профилирование запросов // ДЗ

Цель занятия

познакомиться с понятием профилирования запросов в ClickHouse для оптимизации производительности; изучить инструменты и методы профилирования запросов в ClickHouse для анализа времени выполнения, использования ресурсов и оптимизации запросов.

Краткое содержание

введение в профилирование запросов: определение, назначение, принципы работы профилирования запросов в ClickHouse; инструменты профилирования запросов: описание инструментов и методов для сбора и анализа профилей выполнения запросов в ClickHouse; анализ результатов профилирования: интерпретация данных профилей выполнения запросов, выявление узких мест и оптимизация запросов; практические примеры профилирования запросов: использование инструментов профилирования для оптимизации производительности запросов в ClickHouse.

Домашние задания

Профилирование запросов

Цель: сравнить выполнение запросов с использованием первичного ключа и без него, проанализировав логи и план запросов.

Тема 5

Сессия Q&A

Цель занятия

получить ответы на вопросы по темам занятий и домашних заданий.

Краткое содержание

типичные ошибки при выполнении ДЗ; преподаватели ответят на ваши вопросы.

Модуль 5. Популярные интеграции

Тема 1

Оркестраторы/DI Tools // ДЗ

Цель занятия

научиться взаимодействовать с оркестраторами и DI Tools; получить практику и полезные советы от экспертов.

Краткое содержание

Airflow; Alibute.

Домашние задания

Загрузка данных в ClickHouse

Цель: настроить ETL-процесс для регулярной загрузки данных из внешнего источника в ClickHouse.

Тема 2

Интеграция с Kafka и подводные камни // ДЗ

Цель занятия

научиться взаимодействовать с Kafka; получить практику и полезные советы от экспертов.

Краткое содержание

Apache Kafka; Kafka-templ; MV.

Домашние задания

Интеграция с Apache Kafka

Цель: изучить интеграцию ClickHouse с Kafka и настроить пайплайн для записи и чтения данных.

Тема 3

Интеграции с BI-инструментами // ДЗ

Цель занятия

научиться взаимодействовать с BI; получить практику и полезные советы от экспертов.

Краткое содержание

Superset; Redash; DataLens.

Домашние задания

Визуализация данных

Цель: подключить Superset к ClickHouse и построить дашборд с набором визуализаций для анализа данных.

Тема 4

PG/Greenplum // ДЗ

Цель занятия

научиться взаимодействовать с PG/Greenplum; получить практику и полезные советы от экспертов.

Краткое содержание

Postgres Engine; JDBC; ODBC.

Домашние задания

Интеграция ClickHouse с PostgreSQL

Цель: настроить базовую интеграцию ClickHouse с PostgreSQL, используя функции и движок Postgres.

Тема 5

ClickHouse и dbt

Цель занятия

познакомиться с Data Build Tool – мультитул для работы с DWH; рассмотреть основные возможности и принципы dbt; понять, как инструменты подобные dbt могут помочь инженерам и аналитикам при работе с ClickHouse.

Краткое содержание

Dbt building blocks and principles; Connecting to DWH: profiles.yaml; Configuration: dbt_project.yaml; Macros + Jinja templates; SQL best practices: Complex SQL transformations + CTE.

Модуль 6. Проектная работа

Тема 1

Выбор темы и организация проектной работы