

## Программа курса «CLICH: Построение DWH на ClickHouse»

**О курсе:** Освойте ClickHouse — самую быструю колоночную СУБД для онлайн-аналитики и обработки больших данных (big data). Наш интенсивный 5 дневный курс ClickHouse разработан для аналитиков данных и инженеров данных, которые хотят научиться строить высоконагруженные корпоративные хранилища (DWH), обрабатывать миллиарды строк за секунды и принимать решения на основе анализа данных в режиме реального времени.

**Аудитория:**

- Аналитик данных
- Дата-инженер
- Архитектор ClickHouse
- Администратор кластера

**Уровень подготовки:** Обязательны базовые навыки работы с SQL.

**Продолжительность курса:** 24 академических часа, 5 дней

**Тестирование:** 45 минутный тест на знание теории и практики по написанию SQL запросов в рамках пройденного курса.

## Содержание программы

### 1. Основы ClickHouse

- Назначение и архитектура ClickHouse
- Логическая структура базы данных
- Преимущества и ограничения колоночных СУБД
- Сценарии использования и развертывания кластера (облако, on-premises, Docker)
- Настройка клиентских подключений и сессий
  - **Лабораторная работа:** Настройка подключения (ClickHouse client, python driver, Dbeaver) к кластеру ClickHouse (Варианты Managed Service/Self-managed), управление сессиями (ограничения на ресурсы и права) и написание запросов к системным таблицам.

### 2. Структура БД в ClickHouse

- Движки баз данных и таблиц (семейство MergeTree, Log, Memory, Buffer)
- Констрейнты, индексы и проекции для ускорения запросов
- Простые и материализованные представления
- Внешние и внутренние словари
  - **Лабораторная работа:** Проектирование и создание баз данных, таблиц, словарей, индексов и представлений в ClickHouse.

### 3. Особенности ClickHouse SQL

- Специфические типы данных (включая вложенные структуры)
- Команды DDL и DML
- Эффективное написание SQL-запросов и best practices
- Пользовательские функции (UDF)
- Изменение таблиц с ALTER TABLE и TTL.
- Управление партициями и таблицами
  - **Лабораторная работа:** Изменение параметров таблиц и управление партициями.

### 4. Оптимизация и производительность

- Профилирование запросов с помощью EXPLAIN и system.query\_log
- Партиционирование и кластеризация данных
- Стратегии предварительной агрегации и кэширования
  - **Лабораторная работа:** Находим и ускоряем «тяжёлые» запросы, проводим профилирование и оптимизацию.

## 5. Распределённые системы

- Шардирование и репликация данных для масштабирования и надёжности
- Настройка отказоустойчивого кластера
- Балансировка нагрузки при **обработке запросов**
- Распределённые таблицы
  - **Лабораторная работа:** Создание и работа с реплицируемыми и распределёнными таблицами. Обслуживание репликации и мониторинг Zookeeper/Keeper.

## 6. Интеграция и расширенные возможности

- Работа с JSON, массивами и геоданными
- Поточная **обработка больших данных** с помощью ClickPipes
- Интеграция с Kafka, PostgreSQL
- Инструменты для **визуализации данных**: Datalens, SuperSet, Grafana, Metabase
  - **Лабораторная работа:** Создаём конвейер потоковой обработки данных на ClickHouse с Kafka/RabbitMQ.

## 7. Администрирование и безопасность ClickHouse (факультатив в видео формате)

*Материалы данного модуля предоставляются в формате видеоматериалов, с комплектом лабораторных работ, которые вы можете выполнить дополнительно, в свободное время, и при необходимости задать вопросы преподавателю.*

- Управление пользователями, ролями и доступами
- Методы аутентификации
- Настройка **Named Collections** для упрощения защиты интеграции с внешними источниками
- Резервное копирование и восстановление кластера
- **Мониторинг** и кастомные метрики
- Настройка алертинга (Prometheus, Grafana)
  - **Лабораторная работа:** Настройка бэкапов, реализация кастомных метрик и построение дашбордов мониторинга.

## Бонус: +8 часов материалов для глубокого погружения!

Для администраторов кластера ClickHouse и дата инженеров которые выполняют задачи обслуживания и администрирования кластеров ClickHouse мы записали дополнительные модули, которые могут быть полезны. После просмотра видеоуроков, вы сможете задать вопросы преподавателю и выполнить дополнительные лабораторные работы для закрепления материала на практике.

- **3 часа дополнительных лекций** по продвинутым темам администрирования и обслуживания.
- **5 часов дополнительных лабораторных работ** для оттачивания навыков сопровождения кластера ClickHouse.