

## Программа курса «TRINO: TRINO для инженеров данных»

**О курсе** Наш 4х дневный курс Trino — это комплексная программа для инженеров и аналитиков данных, которая научит вас строить быструю и гибкую аналитическую платформу с единой точкой доступа ко всем вашим данным.

### Аудитория:

- Инженер данных
- Аналитик данных
- Data Architect
- Backend-разработчик

### Уровень подготовки:

- Уверенное знание SQL.
- Опыт работы с командной строкой Linux.
- Базовое понимание экосистемы Big Data (что такое HDFS, S3, Hive).

**Продолжительность курса:** 16 академических часов, 4 дня по 4 ак. часа дистанционно

## Содержание программы

### 1. Знакомство с Trino



- **Введение:** история, место в экосистеме больших данных
- **Архитектура:** координатор, воркеры, основные компоненты
- **Быстрый старт:** установка, первые запросы через CLI
- **Системные таблицы** и базовый мониторинг
  - **Практическое задание.** Знакомство с интерфейсом Trino CLI, Web UI. Выполнение первых запросов к встроеным системным таблицам. Запускаем TPC-H запросы.

### 2. Коннекторы и федеративные запросы

- **Концепция коннекторов:** каталоги, схемы, таблицы
- **Популярные коннекторы:**
  - PostgreSQL/MySQL (реляционные БД)
  - Hive/Iceberg/Parquet (озера данных)
  - ClickHouse (аналитические БД)
  - Apache Kafka
  - Memory/TPC-DS (тестовые)
- **Федеративные запросы** между разными источниками особенности работы с ними
- **Настройка каталогов** и troubleshooting подключений
  - **Практическое задание.** Построение витрин данных из разных источников федеративными запросами. Настройка коннекторов к Hive, S3, PostgreSQL и Kafka. Выполнение federated-запроса, объединяющего данные из всех четырех источников в одном SQL-выражении.

### 3. Выполнение запросов и оптимизация в Trino

- **Жизненный цикл запроса:** stages → tasks → splits → drivers
- **Чтение планов выполнения** и поиск узких мест. Чтение и анализ вывода EXPLAIN ANALYZE
- **Оптимизатор Trino:** правила, статистика, hints. Cost-Based Optimizer и важность сбора статистики по данным.
- **Производительность:** параллелизм, память, сетевое взаимодействие.
- **Техники оптимизации:** правильный JOIN, использование approx\_distinct(), партиционирование и бакетирование данных в Data Lake.

 ШКОЛА БОЛЬШИХ ДАННЫХ	ООО «Учебный центр «Коммерсант» © <b>«Школа Больших Данных»</b> <a href="http://www.bigdataschool.ru">www.bigdataschool.ru</a> 2026	
---	--	--

- **Практическое задание.** Оптимизация и поиск узких мест в запросах бенчмарка TPC-H. Работа с данными из Apache Kafka.

#### 4. Продуктивная работа с Trino

- **Мониторинг:** Web UI, системные таблицы, внешние инструменты
- **Управление ресурсами:** resource groups, session properties
- **Многопользовательская работа:** роли, права доступа
- **Типовые проблемы** и их решение
  - **Практическое задание.** Решаем бизнес-задачу на реальных данных с полным циклом — от настройки коннекторов до оптимизации запросов.

#### Бонусные задачи и время

Для завершения работы над финальным проектом «Школа больших данных» предоставляет дополнительное время работы на кластерах в Облаке.

Для тех, кто выполнил все задачи на кластере, мы также предоставляем дополнительные бонусные лабораторные работы, для закрепления практических навыков использования и настройки SQL engine Trino.

12 часов рабочего времени на кластере