

## Программа курса «DBT: Data Build Tool для инженеров данных»

**О курсе:** Быстрое освоение CLI-инструмента для эффективных ETL-операций.

Как создать, настроить и использовать dbt-проект: подключение к базе данных, написание моделей, использование макросов и материализации, создание пользовательских макросов, запуск и тестирование dbt-моделей

**Аудитория:**

- Начинающие инженеры данных
- Опытные дата-инженеры, которые хотят изучить dbt
- Аналитики данных

**Уровень подготовки:**

- Базовые знания SQL
- Базовые знания Linux

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

## Содержание программы

### 1. Общая информация

- Введение в dbt
- Основные концепции и принципы работы инструмента
- Роль dbt в процессах обработки данных (ETL/ELT)

### 2. Инициализация проекта

- Установка и настройка dbt
- Создание нового проекта в dbt

### 3. Настройка проекта

- Использование конфигурационных файлов
- Управление проектом и зависимостями
- Работа с переменными и параметрами проекта

### 4. Оркестрация Spark-приложений в Airflow

- Создание и организация моделей данных
- Использование SQL для трансформации и очистки данных
- Обработка отношений между моделями

### 5. Конфиги и макросы

- Использование конфигураций для моделей
- Создание собственных макросов
- Использование готовых макросов dbt для рутинных операций

### 6. Запуск моделей

- Автоматизация процессов загрузки данных
- Выполнение моделей и обновление результатов
- Граф зависимостей моделей и их выполнение в правильной последовательности

## 7. Тесты

- Проведение тестирования данных с помощью dbt
- Создание и применение тестов для обеспечения корректности данных
- Интеграция тестов в процесс разработки и обслуживания