

## Курс: Hadoop Hive

**Длительность: 8 ак. часов**

### О курсе

Базовый курс для администраторов и разработчиков для **Apache Hive**. Рассматриваются методы доступа к данным с использованием HiveQL, загрузка и выгрузка данных в Hive, HDFS, реляционные БД и **NoSQL**, организация и оптимизация хранения данных с использованием **Hive** на **HDFS**, вопросы интеграции с другими инструментами экосистемы Hadoop. Практические занятия выполняются в **AWS** и локальной кластерной системе с использованием дистрибутивов **Cloudera Hadoop** и **HortonWorks Data Platform**.

### Аудитория

Специалисты SQL имеющие и желающие применить опыт работы с реляционными базами данных для организации хранения, настройке импорта и экспорта данных с использованием **Apache Hive** в кластере **Hadoop**.

### Предварительная подготовка

- Начальный опыт работы в **Unix, SQL**
- Знание основ **Hadoop**

## Программа курса

### 1. Основные концепции Hive

Архитектура **Hive**. Инструменты администрирования и исполнения запросов: **beeline/Hive shell & HUE Query Editor**. Схема чтения и записи в Hive. Структура и формат хранения файлов HIVE. Синтаксис **HiveQL**. **Hive DDL** и **DML** операции.

### 2. Импорт и экспорт в Hive

Загрузка данных в таблицы **Hive**. Внешние и управляемые таблицы. Оптимизации производительности с использованием Partition, Buckets. Импорт/экспорт данных БД. Форматы хранения данных и применение компрессии.

### 3. JOIN в Hive

Map JOIN. Comon JOIN. Skewed JOIN. Collocated JOIN. Оптимизация запросов.

### 4. Advanced Hive

Hive с **Pig, Impala, Hbase, Spark, Sqoop**. Thrift сервер. User Defined Functions.

### 5. Сценарии применения Hive

<mailto:sales@bigdataschool.ru>

+7 (985) 162-29-63