

Программа курса «HBASE: Администрирование кластера HBase»

О курсе: 4х-дневный практический курс по администрированию HBase разработан для специалистов, отвечающих за администрирование, настройку и сопровождение кластера Apache HBase.

Аудитория: Наши авторские курсы администрирования кластера HBase ориентированы на специалистов по администрированию и сопровождению кластеров, системных администраторов, архитекторов, разработчиков Big Data систем и DevOps-инженеров, которые хотят получить следующие знания и навыки:

- понять, что такое Apache HBase;
- научиться устанавливать, настраивать и поддерживать кластера HBase и Zookeeper;
- освоить методы администрирования кластера HBase;
- изучить настройки безопасности для Apache HBase;
- настраивать мониторинг и оптимизацию NoSQL-СУБД с использованием эффективных инструментов администратора Big Data NoSQL кластера;
- интегрировать HBase с другими Big Data компонентами (Apache Pig, Hive, Sqoop, Impala) и внешними системами.

Уровень подготовки:

- Знание базовых команд Linux (опыт работы с командной строкой, файловой системой, POSIX, текстовыми редакторами vi, nano)
- Начальный опыт работы с Hadoop
- Базовые навыки программирования Java

Продолжительность курса: 32 академических часа, 4 дня по 8 ак. часов дистанционно

Содержание программы

1. Введение в HBase

- Немного о Hadoop и NoSQL. Hive и HBase. NoSQL
- Архитектура HBase. Модель данных HBase: Column family, timestamp, версияность
- Сервисы и операторы HBase. Концепция CRUD: базовые операции с таблицами
- Сценарии использования HBase

2. Импорт данных и интеграция с HBase

- Импорт данных с использованием importTSV
- Взаимодействие компонент Hadoop и HBase: HDFS, PIG, Hive, sqoop, Impala
- Хранение файлов в HDFS. Форматы файлов для HDFS: параметры сжатия, ORC, Parquet
- Использование Java API для работы с HBase

3. Установка и настройка HBase

- Требования по установке HBase
- Дизайн схемы HBase для эффективного хранения данных и HA
- Установка Apache HBase
- Команды и интерфейс командной строки HBase: HBaseAdmin
- Ручная установка кластера HBase. Настройка кластера Zookeeper

4. Управление кластером HBase

- Высокая доступность (High Availability) кластера HBase. Настройка Snapshots. Настройка репликации кластера

- Безопасность HBase. Настройка аутентификации и авторизации при доступе к данным в таблицах HBase
- Резервное копирование и восстановление HBase

5. Advanced HBase

- Мониторинг Apache HBase. Анализ лог файлов HBase
- Операции обслуживания. Compactions и flushes
- Оптимизация параметров HBase: Hfile, WAL, Memstores, Bloom фильтр
- Диагностика и разрешение проблем (troubleshooting). Мониторинг и оптимизация JVM. Garbage Collection
- Region splitting