

Курс: HBase администратор

Длительность: 32 ак. часа

О курсе

Данный курс разработан для специалистов отвечающих за администрирование, настройку и сопровождение кластера Apache HBase. Apache HBase представляет NoSQL колоночную (column oriented) базу данных использующую HDFS, обеспечивающая динамическое партиционирование данных, распределение нагрузки, поддержку кэширования и возможности доступа в режиме реального времени. Большое внимание уделяется вопросу проектирования схемы HBase и обеспечения высокой доступности кластера HBase.

Аудитория

Специалисты по администрированию и сопровождению HBase, архитекторы и разработчики

Соотношение теории к практике 40/60

Предварительная подготовка

- Начальный опыт работы в **Unix/SQL**, текстовый редактор **vi**
- Начальный опыт работы с **Hadoop**
- Базовые навыки программирования **Java**

Программа курса

1. Введение в HBase

Немного **Hadoop** и **NoSQL**. **Hive** и **HBase**. **NoSQL**.

Архитектура HBase. **Модель данных HBase**: Column family, timestamp, версионность. Сервисы и операторы HBase. Концепция **CRUD**: базовые операции с таблицами. Сценарии использования **HBase**.

2. Импорт данных в HBase

Ипорт 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.

Настройка аутентификации и авторизации при доступе к данным в таблицах HBase.

Резервное копирование и восстановление HBase.

Безопасность HBase.

5. Advanced HBase

Мониторинг **Apache HBase**. Анализ лог файлов HBase.

Операции обслуживания. Compactions.

Оптимизация параметров **HBase**: Hfile, WAL, Memstores, Bloom фильтры.

Диагностика и разрешение проблем (troubleshooting). Мониторинг и оптимизация JVM.

Garbage Collection. Region splitting.