

## Курс: Администрирование кластера Apache Kafka

**Длительность: 24 ак. часов**

### О курсе

3х-дневный курс для специалистов желающих получить и систематизировать знания по использованию **Apache Kafka** для организации распределенной потоковой обработки событий (**Event Streaming Processing**) на практических примерах с потоковыми данными в реальном времени. Курс содержит расширенные сведения по установке распределенного кластера **Apache Kafka**, настройке безопасности **Kerberos** и интеграции с другими продуктами, мониторинга и оптимизации **Kafka** в соответствии с "**best practices**".

### Аудитория

Специалисты желающие изучить методы распределенной обработки очереди сообщений **Kafka** с использованием практических примеров реализации.

**Соотношение теории к практике 40/60**

### Предварительная подготовка

- Начальный опыт работы в **Unix**, текстовый редактор **vi**
- Начальный опыт работы в **Hadoop** (желателен)
- Начальный опыт программирования (**Java**) (желателен)

## Программа курса

### 1. Введение в Apache Kafka

Что такое **Apache Kafka**?

Сценарии использования **Kafka**

Компоненты **Kafka** и **Kafka** экосистемы

Архитектура **Kafka**

Кластер **Zookeeper**

Потоковая Обработка Событий(**Event Stream Processing**): Терминология и основные концепции

### 2. Основные концепции Apache Kafka

**Как работает Kafka**: Топики (**Topics**) и партиций (**Partitions**). Репликация данных.

**Брокеры (Brokers)**: Развертывание. Конфигурация **Multibroker**.

**Поставщики данных (Producers)**: Типы партиций. Синхронные и асинхронные **Producers**.

Параметры конфигурации. **Java API**.

**Потребители данных (Consumers)**: Очереди **Consumers**. Широковещание (**Broadcasting**).

### 3. Установка кластера Apache Kafka

Требования по установке кластера **Kafka**

Установка и настройка кластера **Zookeeper**: Настройка кворума. Конфигурация сервисов и файловой системы. Оптимизация и отказоустойчивость **Zookeeper**.

Установка и настройка кластера **Kafka**: Планирование установки и **sizing** кластера **Kafka**.

Типовые проблемы при установке. Запуск и первоначальное конфигурирование **Kafka**.

Конфигурация **Listeners**. Отказоустойчивость кластера **Kafka**. **Kafka** менеджер.

Установка кластера **Kafka** на **AWS**

#### 4. Базовые операции в Apache Kafka

Операции с Топиками (**Topic**)

Публикация данных с консольного **Producer**

Получение данных с консольного **Consumer**

Пользовательские интерфейсы топика **Kafka** (**Topics UI**)

Написание кода **producer**

Написание кода **consumer**

#### 5. Оптимизация кластера Apache Kafka

Оптимизация производительности компонент кластера

Масштабирование кластера

Балансировка разделов(partition) и использование Partition Reassignment Tools

**Best Practices**.

Мониторинг кластера **Kafka**

#### 6. Интеграция с Apache Kafka

Интеграция **Spark** и **Spark Streaming**

Интеграция потоковой обработки **Spark Streaming** и **Kafka**

Использование **Apache NiFi**

Интеграция **Apache Flume** с **Kafka**

#### 7. Расширенная конфигурация кластера Apache Kafka

**Расширенная конфигурация топиков (Topic)**: Изменение конфигурации топиков.

**Sizing**(количество партиций и фактор репликации). Сегменты и индексы. Политика очистки логов (**Log Cleanup Policy**). Сжатие и консолидация данных в логах.

Пользовательский интерфейс **Landoop Kafka Topics**

**Confluent REST Proxy**

Настройка **Confluent Schema Registry**

Архитектура **Kafka Connect**

#### 8. Безопасность в Kafka

**Шифрование SSL**: Установка центра сертификатов (**CA**), настройка **SSL** для **Kafka** кластера и клиентов.

Аутентификация **SSL/SASL**

**Аутентификация Kerberos**: Настройка **Kerberos**. **Keytabs** и **principals**. Настройка кластера **Kafka** и клиентов для поддержки **Kerberos**.

Авторизация в **Kafka**. Управление списками управления доступом (**ACL**)