

Программа курса «NIFI3: Эксплуатация Apache NIFI»

О курсе: Этот курс полностью построен на новейшей версии Apache NiFi 2, что гарантирует вам самые актуальные знания и навыки. Забудьте о рутине, ошибках в коде и проблемах с интеграцией. NiFi позволяет в графическом интерфейсе соединять любые системы, преобразовывать данные на лету и контролировать каждый шаг их перемещения. Вы научитесь не просто настраивать, а эффективно эксплуатировать эту мощную платформу, обеспечивая стабильную и надежную доставку данных для решения критически важных бизнес-задач.

Аудитория:

- Инженер данных
- Аналитик данных
- Администратор NiFi
- Разработчики

Уровень подготовки: Знание базовых команд Linux (опыт работы с командной строкой, файловой системой, POSIX, текстовыми редакторами vi, nano)

Начальный опыт работы с Hadoop или NiFi

Соотношение теории к практике 50/50

Продолжительность курса: 24 академических часа, 3 дня по 8 ак. часов дистанционно

Содержание программы

1. Введение в Apache NiFi

- Архитектура Hadoop Data Flow
- Архитектура решения и ключевые концепции Apache NiFi: Принципы потоковой обработки данных
- Data Lineage и Data Provenance (Data Provenance Events)
- Функционал NiFi по управлению потоковой обработкой данных: гарантированная доставка, управление каналом (буферизация, pressure), QoS, версияционность
 - **Лабораторная работа:** Установка Single Node Apache NiFi и построение потока данных.

2. Создание DataFlow

- Создание dataflow: работа с атрибутами FlowFile, маршрутизация, source и sink
- Репозитории FlowFile Repository, Content Repository, Provenance Repository. Анатомия FlowFile, процессоры и коннекторы
- Мониторинг dataflow
- Параметризация dataflow: переменные и параметры
 - **Лабораторная работа:** Применение политик доступа.

3. Программная обработка FlowFile

- Использование скриптов в NiFi: работа с атрибутами и данными FlowFile, программная маршрутизация FlowFile
- Сбор данных в месте их производства с помощью Apache minifi
 - **Лабораторная работа:** JSON процессинг в Apache Nifi

4. Работа с Records и Local State

- Controlling service и работа с Records и State: формат FlowFile и его

- преобразование, State Manager и его использование
- Reporting tasks и обработка результатов
- Происхождение данных: анализ provenance и lineage
- Apache NiFi Registry версионность и поддержка production deployment
 - **Лабораторная работа:** Отслеживание изменений и версионный контроль DataFlow

5. Интеграция и масштабирование

- Вертикальное и горизонтальное масштабирование и отказоустойчивость NiFi
- Доступные расширения и готовые «из коробки» процессоры и коннекторы
- Возможности интеграции кластера NiFi (consumer/provider) с Kafka
- Интеграция Apache Kafka, NiFi
- Типовые сценарии/схемы потоковой обработки данных
- Процессорные группы (Remote Processor Group)
 - **Лабораторная работа:** Интеграция **MySQL** и Apache NiFi: Работа с SQL.
 - **Лабораторная работа:** CDC для отслеживания изменений данных в NiFi

6. Задачи администрирование NiFi

- Установка и настройка кластера Apache NiFi
- Управление кластером NiFi
- Оптимизация DataFlow
- Инструменты мониторинга и уведомлений NiFi
- Безопасность для Apache NiFi (**SSL**, Аутентификация **LDAP**)
 - **Лабораторная работа:** Настройка распределенного кластера Apache Nifi. **Лабораторная работа:** Настройка аутентификации **Kerberos** для кластера Nifi.