



Курс: Apache AirFlow

Длительность: 16 ак. часов

Аудитория

Системные администраторы, системные архитекторы, разработчики Hadoop желающие получить практические навыки по управлению потоковыми данными с использованием **Apache AirFlow**.

Соотношение теории к практике 60/40

Необходимая предварительная подготовка

- Начальный опыт работы в **Unix**, опыт работы с текстовым редактором **vi** (желателен)
- Начальный опыт программирования **Python/bash**

Программа курса

1. Введение

- История появления, на чем написан (**python**)
- Основной объект (**DAG**)
- Операторы и задачи
- **Worker**
- **Scheduler**, **schedule interval** и **execution date**
- Пулы
- Приоритезация
- Метаданные
- **Airflow U/I** и **Webserver**
- Мониторинг (средства **Airflow** и кастомные варианты)
- Алерты
- Логирование

2. Разработка

- Создание и основные параметры **DAG**
- **Operators** и **plugins**

- **Hooks, connections, и variables**
 - Работающие из коробки и уже написанные community операторы, хуки и т.п.
 - Создание задач
 - Макросы (**Jinja**)
 - Управление зависимостями (внутри **DAG**, внешние зависимости, **timedelta**)
 - Визуализация в **Web UI**
 - Настройка расписания
 - **Контекст** (выполнения task)
 - Обмен сообщениями между tasks, **DAGS (xcom)**
 - Добавление настраиваемых операторов, сенсоров, хуков и т.п.
- 3. Развертывание и настройка Airflow**
- Установка **Airflow** в конфигурации по умолчанию (**SQLite, SequentialExecutor**)
 - Установка **Redis, Celery**
 - Настройка airflow.cfg (**PostgreSQL, Celery, Redis, parallel degree...**)
 - Запуск (**service, n-psystemctl, docker**)
 - Кластеризация (масштабируемость, безотказность)
- 4. Особенности и проблемы в Airflow**
- Версии **python**(2 или 3)
 - Debug
 - Тестирование
 - Логирование
- Практические занятия:**
- Настройка окружения (**Pycharm, python**, библиотеки для окружения в **Virtualenv**).
 - Создание **DAG** с задачами **BashOperator / PythonOperator** для получения данных из **Kafka**.
 - Создание кастомного оператора, осуществляющего по заданным параметрам выгрузку из **Kafka**. Создание **DAG**, использующего созданный оператор.