

Apache Airflow – orkiestracja zadań

Opis

Szkolenie Apache Airflow lub po prostu szkolenia Airflow zapoznaje uczestników z technologią Apache Airflow (w skrócie Airflow) w sposób praktyczny. Na kursie zostaje przekazana praktyczna wiedza z zakresu planowania i sterowania przepływami pracy zadań powiązanych z ETL. Podczas zadań zostaną omówione sposoby implementacji poszczególnych zadań, sterowanie całym planem wykonania oraz tworzenie połączeń do źródeł zewnętrznych (np. bazy danych). Nie zabraknie również podstaw administracji oraz wykonywania poleceń z poziomu API narzędzia

Grupa docelowa: DevOps, Programiści, Inżynierowie Danych

Środowisko pracy: 1. Konteneryzacja za pomocą Dockera, 2. Maszyna wirtualna z systemem operacyjnym Linux (preferowany Debian/Ubuntu)

Wymagana wiedza: Bardzo dobra znajomość języka Python, Podstawy programowania w Shell Script (Bash)

Czego Cię nauczymy

Architektura i porównanie Wersji

Nauczmy Cię podstawowej architektury Apache Airflow oraz przeprowadzimy porównanie między wersją 1.0 a 2.0. Dowiesz się, jakie zmiany zaszyły między tymi dwiema wersjami, co pomoże Ci lepiej zrozumieć, jak efektywnie korzystać z najnowszych funkcji. Po ukończeniu tego modułu będziesz gotów zainstalować środowisko pracy i przystąpić do praktycznego wykorzystania tej wiedzy.

Techniki związane z operatorami i Sensorami

Omówimy bardziej zaawansowane aspekty korzystania z operatorów, takich jak Python czy Http. Nauczysz się tworzyć połączenia do różnych źródeł zewnętrznych, a także korzystać z sensorów, takich jak File, Http czy Python, aby skutecznie kontrolować przepływ pracy zadań.

Tworzenie pierwszego planu i praca z operatorami

W tym module skoncentrujemy się na praktycznych aspektach korzystania z Apache Airflow. Nauczysz się korzystać z interfejsu, tworzyć podstawowe plany oraz operować na operatorach, takich jak Empty czy Bash. Dzięki temu zrozumiesz, jakie możliwości oferuje narzędzie w zakresie planowania i wykonywania zadań.

Zarządzanie zadaniami i optymalizacja wykonań

Ten moduł koncentruje się na skutecznym zarządzaniu zadaniami w Apache Airflow. Nauczysz się tworzyć harmonogramy zadań, radzić sobie z catchup & backfill, wykorzystywać tagi oraz resetować wykonania. Poznasz także zaawansowane koncepcje, takie jak operator PythonBranch i trigger rules, które umożliwią Ci jeszcze lepszą kontrolę nad przepływem pracy.

Program szkolenia

1. Wprowadzenie

- Architektura
- Porównanie wersji 1.0 i 2.0
- Instalacja środowiska pracy

2. Pierwszy plan

- Omówienie interfejsu
- Tworzenie pierwszego planu
- Operatory: Empty, Bash

3. Operatory i sensory

- Operatory: Python, Http
- Tworzenie połączeń
- Sensory: File, Http, Python

4. Zarządzanie zadaniami

- Harmonogram zadań
- Catchup & Backfil
- Tagi
- Resetowanie wykonania

- Operator: PythonBranch
- Trigger rules

5. Zmienne DAGa

- Zmienne Airflow
- Obiekt x-com
- Konfiguracja DAGa
- Jinja Template

6. Praca z zbiorami danych

- Połączenie z bazami danych SQL
- Połączenie z MongoDB
- Połączenie z S3

7. Zarządzanie środowiskiem Airflow

- Instalacja pakietów
- Tworzenie połączeń (overview)
- Aktualizacja narzędzia
- Uruchamianie klastra / wersji HA

Czas trwania

3 dni | 24 godziny zajęć

Certyfikat

Uczestnicy szkolenia otrzymują imienne certyfikaty sygnowane przez Expose Sp. z o.o.

Cena szkolenia

2 990 PLN netto (VAT 23%) za osobę (szkolenie grupowe)

Cena szkolenia zawiera

- ✔ zapewnienie autorskich materiałów szkoleniowych dla uczestników szkolenia
- ✔ wystawienie certyfikatów po zakończonym szkoleniu
- ✔ rekomendacje dla uczestników szkolenia w zakresie dalszej pracy w obszarze szkolenia
- ✔ pakiet konsultacji z wykładawcą po zakończonym szkoleniu w razie jakichkolwiek niejasności przez okres 3 miesięcy
- ✔ całodzienny serwis kawowy oraz lunch