Updated 05/02/2025

Sign up

dbt training : Airflow workflow
2 days (14 hours)

Presentation

Over the past few years, dbt (data build tool) has become an indispensable tool in modern data architectures. Free, open source and designed for the cloud, it enables analytics teams to draw inspiration from software engineering practices to rapidly deploy data pipelines.

During this course, you'll set up a dbt-specific development environment, enabling you to develop and test your first models.

You'll also see how dbt fits into today's data architectures, and what issues this tool can address.

Following this training course, you'll learn all about dbt's main features, as well as the different ways of using it in production, in particular using Airflow and Kubernetes. As with all our training courses, this one will introduce you to the version and its new features: dbt Core v1.9.

Objectives

- Setting up a project from scratch
- How to use and configure the dbt IDE
- How to deploy a production with dbt
- Testing and documenting a model
- Internalize best practices

Target audience

- Data Scientists
- Data Analysts
- Project managers

- Data engineers
- Business Analysts

PREREQUISITES

- Knowledge of data pipelines
- How to use Docker
- Knowledge of python

technical prerequisites

- Have a Gmail account
- Have a Github account (linked to your Gmail account)
- Python 3 and Git installed
- Docker and Docker Compose installed
- Code-editing software (e.g. Visual Studio)
- Omyzsh

Our dbt training program

Introduction

- What is dbt?
- ETL and ELT: the differences
- dbt Core vs dbt Cloud
- dbt's role in the Modern Data Stack
- What is not dbt?
- Data Warehousing: Data Lakes, Data Warehouses, Data Lakehouse

Setting up the dbt environment

- Creating the Python virtual environment
- Installing dbt Core and the appropriate plugin
- Connecting dbt to the data warehouse
- YAML parameter files
- Best practices in project structure: raw data, staging, etc.

Models: Development and testing

- Basic model development
- Data warehouse organization
- Model configuration: schematics, materialization, etc.
- Introduction to the various dbt commands and possible flags
- Setting up tests and verifying results

Templates: Advanced configurations

- Use of sources and verification of freshness
- Ingesting flat files in dbt: seeds and their usefulness
- Incremental materialization: append or update
- Macros, hooks
- Snapshots

Documentation in dbt

- Presentation of documentation in dbt format
- The dbt docs command and its various uses
- Test types: format (e.g. generic, singular) + qualitative (e.g. structural, relational)
- Test examples
- Good documentation practices: using Jinja syntax
- DBT docs and lineage

Packages

- The Hub Package
  - dbt_utils
  - codegen
  - re_data

Dbt in production

- Continuous integration and development (CI/CD)
- dbt orchestration using Apache Airflow: BashOperator, KubernetesPodOperator
- Overview of different ways of executing projects
- Jobs scheduling
  - DBT cloud
  - Airflow

Companies concerned

This course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is confirmed, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives with regard to the training to come, within the limits imposed by the format selected. This

The questionnaire also enables us anticipate any connection or internal security (intra-company or virtual classroom) that could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.