

Updated 07/27/2023

Sign up

Dagster training: Orchestrating data pipelines in the cloud

3 days (21 hours)

Presentation

The [Dagster](#) training course will teach you how to use this data orchestrator. You'll learn how to use this framework to create data pipelines and dispatch them at extraordinary speed.

During this apprenticeship, your team will learn to ingest data from external sources or other data resources. You'll also be able to combine and transform data in meaningful ways.

Dagster will be useful to you because it has three key points that make it unique. Your company will be able to develop locally and run automated tests. In particular, you'll be able to calculate the value of an asset using a pure declarative Python function. Thanks to its decoupling of pipelines from the environment, you'll have access to a wide range of functionalities.

Like all our training courses, this one will introduce you to the latest [version](#) and its new features: Dagster 1.1.

Objectives

- Integrate various tools into Dagster
- Knowing where to use a stored asset
- Master all concepts
- Understanding and implementing deployment architecture

Target audience

- Data Scientists
- Data Analysts
- Analysis engineers
- Data engineers
- DevOps engineers

Pre-rigged

- General data knowledge
- Knowledge of the Python language

Dagster training program

Introduction

- Introducing Dagster
- Installation
- First step with Dagster
- Telemetry

Creating a new project

- Using the project skeleton
- Starting the Today Web server
- Adding new Python dependencies
- Unit testing
- Environment variables and secrets

Assets

- Define an asset
- Resource graph construction
- Define resources without return values or arguments
- Asset testing
- Automated asset materialization
- Asset allocation
- Customize asset storage
- Non-asset work

Concepts

- Operations
 - Hooks
 - Events
 - Attempts
- Graphics
 - Dynamics
 - Interlocking
- Work
 - Metadata and tags
 - Carrying out the work
- Programs
 - Sensors
 - Partitions
 - Start filling
- Launching the Dagster user interface

Deployment

- Deployment architecture
- Long-term services
- Configuration
- Task execution flow
- Dagster Cloud

Integrations

- Using Dagster with Airflow
- Spark
 - Executing PySpark code in operations
 - Submitting PySark operations on EMR
- Validating Pandas DataFrames with Pandas types
- Snowflake and Dagster
 - Use Python to analyze data stored in Snowflake without SQL queries
 - Integrate your Snowflake tables with other tools

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.

Teaching methods

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

Organization

The course alternates theoretical inputs from the trainer supported by examples and

brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire is used to check that skills have been correctly acquired.

Sanction

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