Updated 09/08/2024

Sign up

AWS Glue training

2 days (14 hours)

Presentation

Learn more about AWS Glue, amazon's ETL solution. AWS Glue provides a scalable, serverless ETL solution where scripts can be written in Python. It's a very practical cloud tool at a time when more and more companies are turning to this kind of solution.

During our training course, you'll discover this tool and its key features. You'll learn how to configure the environment for your specific needs. You'll also learn how to use the tool to transform your data and integrate it into your pipelines.

Our training includes practical workshops that will enable you to validate the skills you have acquired.

Like all our training courses, it will be run on the latest version of the tool: AWS glue 4.

Objectives

- Understanding the role of AWS glue in ETL
- Master architecture and fundamental concepts
- Configure connectors and manage data destinations

Target audience

- Data engineer
- Data scientists
- Data analysts

Prerequisites

Basic knowledge of Python

OUR AWS GLUE TRAINING PROGRAM

INTRODUCTION TO AWS GLUE

- Overview of AWS Glue and its components
- Benefits and typical use cases
- AWS Glue solution architecture
- AWS Glue console demonstration
- Discussion of real-life scenarios

CONFIGURING THE AWS GLUE ENVIRONMENT

- Setting up and configuring the AWS Glue environment
- Exploring and configuring the AWS Glue data catalog
- Security and IAM for AWS Glue
- Practical lab: Initial configuration and environment verification

DATA TRANSFORMATION WITH AWS GLUE

- Basics of data transformation and ETL jobs
- Creating and managing ETL jobs with AWS Glue
- Using DynamicFrames and DataFrames
- Practical lab: Creating and running a simple ETL job

USING AWS GLUE CRAWLERS

- Introduction to crawlers and how they work
- Configuring and running crawlers for data exploration
- Using custom classifiers
- Practical lab: Setting up and using a crawler to analyze a dataset

ADVANCED ETL TRANSFORMATIONS AND ERROR HANDLING

- Advanced data transformation techniques
- Error and exception handling in ETL jobs
- Optimizing ETL job performance
- Practical lab: Implementing complex transformations and error handling

MONITORING, TROUBLESHOOTING AND BEST PRACTICES

- Job monitoring and logging with AWS CloudWatch
- Troubleshooting common problems in AWS Glue
- Best practices for safety, performance and cost reduction
- Case study and Q&A session
- Course summary and discussion of next steps

Companies concerned

This training 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 finalized, 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 for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which 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, with 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.