+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated 05/28/2024

Sign up

Azure Databricks training

2 days (14 hours)

Presentation

Azure Databricks is the solution for handling massive numbers of database clusters. Azure Databricks leverages the power of the Microsoft Cloud to simplify your Big Data process.

Azure Databricks is an enhanced version of Apache Spark managed entirely by Azure. This makes the system easily scalable and controllable, reducing your costs during the development phase.

The system features various connectors such as ADSL, Azure SQL DW and PowerBI. It's also a highly secure tool, integrating the Azure Active Directory security framework.

Our Azure Databricks training course will teach you how to use this tool in your ETL process. We'll show you how to extract, transform and load data with Azure Databricks, as well as how to create dashboards and pipelines.

Our Azure Databricks training course will introduce the latest version of the tool, Azure Databricks.

Objectives

- Getting to know the specifics of Azure Databricks
- Extracting data with Azure Databricks
- How to transform and load data
- Use dashboards and deploy your process

Target audience

- Developers
- Data Engineer
- Architects
- System administrators
- Data miners
- Data scientists
- Data analysts
- Business intelligence analysts
- Market inteligence analysts

Prerequisites

- Ideally, you should have taken our Spark ML or Spark Tuning Advanced training courses.
- Knowledge of SQL and Python

Software requirements

- Have an Azure account
- Have a Databricks account

Azure Databricks training program

INTRODUCTION

- Introducing Databricks
- The benefits of the tool
- Databricks vs Apache Spark
- Interface presentation
- Create your workspace
- Notebook presentation
- Create Azure Databricks components (cluster, job, pool)

DATA EXTRACTION

- The various tables (schemas, databases, tables)
- Sources and destinations
- Browsing system files
- The Filestore folder
- Import data from your computer
- Import data from the web
- Importing data with SQL and Python
- Connecting Excel, PowerBI and PyCharm data

DATA TRANSFORMATION

- Find and modify null values
- Delete duplicate data
- Delete extreme values
- Caching
- Data compression

DATA LOADING

- Backup methods
- Managed vs unmanaged tables
- Manage partitions
- Connection to Azure SQL, Oracle and MongoDB

DASHBOARD AND PROCESS DEPLOYMENT

- Dashboard notebooks
- Create a scheduled job to refresh the dashboard
- Overview of an Azure Databricks CI/CD pipeline
- Developing and validating your code
- Define your build pipeline
- Define your release pipeline

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 finalized, the learner receives a self-assessment questionnaire enabling 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 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.