GTC Transforming Emerging Economies

GTC Training Consulting Group Ltd,
22 Kumasi Crescent,
Off Aminu Kano Crescent,
Wuse 2, Abuja.
Tel: +234(0) 9056761232
Email: enquiries@thegtcgroup.com
Web: www.thegtcgroup.com

Scalable AI with Apache Spark and Mllib

COURSE OVERVIEW

This course provides a comprehensive introduction to building scalable machine learning solutions using Apache Spark and its MLlib library. Participants will learn how to process large datasets, develop distributed machine learning models, and optimize workflows for performance across clusters. The course is a combination of theory with practical training to help participants apply Spark's APIs in Python or Scala, enabling scalable and efficient AI workflows for big data environments.

WHO SHOULD ATTEND?

This course is designed for data engineers, machine learning engineers, data scientists, and technical professionals working with large-scale data. It is also important for software developers and architects looking to build distributed AI systems using Apache Spark. Familiarity with Python or Scala and basic machine learning concepts is recommended.

COURSE OUTCOMES

Delegates will gain the knowledge and skills to:

- Know the fundamentals of distributed computing with Apache Spark.
- Build and train scalable machine learning models using Mllib.
- Preprocess and transform large datasets for model training.
- Apply classification, regression, clustering, and recommendation techniques.
- Optimize ML workflows and manage model persistence.
- Use Spark ML pipelines for automation and reproducibility.
- Integrate Spark with cloud platforms and data lakes.
- Troubleshoot performance issues and tune distributed systems.

KEY COURSE HIGHLIGHTS

At the end of the course, you will understand;

- Spark architecture and distributed data processing.
- Hands-on implementation with Spark MLlib algorithms.
- Data wrangling and feature engineering at scale.
- Building ML pipelines for automation and deployment.
- Distributed model evaluation and parameter tuning.
- Integration with Hadoop, HDFS, and cloud services (e.g., AWS, Azure).
- Real-time data processing with Spark Streaming.
- Case studies in fraud detection, predictive analytics, and recommendations.
- Best practices for scaling AI workflows in production.
- Labs using Databricks or standalone Spark environments.

All our courses are dual-certificate courses. At the end of the training, the delegates will receive two certificates.
1. A GTC end-of-course certificate
2. Continuing Professional Development (CPD) Certificate of completion with earned credits awarded

BAC BRITISH ACCREDITATION COUNCIL
Dual Accreditation DUAL PROVIDER: PDC/D1013 2018 - 2020
Project Management Institute Registered Education Provider
ilm A City & Guilds Group Business
ISO 9001:2015
CIPR CHARTERED INSTITUTE OF PUBLIC RELATIONS