Data Bricks Using PySpark

Apache Spark Training Overview

Apache Spark is a powerful framework for big data analytics that provides a unified API for developers, data scientists and analysts to perform various tasks efficiently. It supports multiple popular programming languages, including Python, R, SQL, Java, and Scale. The primary goal of Apache Spark is to offer a hands-on experience in batch processing and real-time data stream analysis and large-scale machine learning solutions, benefiting data scientists, data analysts and software developers.

Apache Spark Training Objectives

- Apache Spark Architecture How to use Spark with Python(PySpark) to build data pipelines.

Pre-requisites of the Course

- Basic knowledge of object-oriented programming is enough Knowledge of Python will be an added advantage
- Learners who have basic knowledge on Database, SQL Query will be an added advantage for Learning this Course

Who should do the course

- Developers, Architects, IT Professionals
- Aspiring Software Engineers, Data scientists and Analysts

Python Basics (1–2 weeks)

1. **Introduction to Python**
   - Python syntax and semantics
   - Variables, data types (strings, integers, floats, Booleans)

- Conditional statements(if, elif, else)
- Loops(for, while)

2. **Functions**
    - Defining functions
    - Arguments and return values
    - Lambda functions
    - Recursion
3. **Data Structures**
    - Lists, tuples, sets and dictionaries
    - List comprehensions
    - Iterators and Generators
4. **Error Handling and Exceptions**
    - Try/except blocks
    - Raising exceptions
    - Debugging techniques
5. **Object-Oriented Programming**
    - Classes and Objects
    - Instance and class variables
    - Methods and magic methods(init, str, etc.)

Introduction to Big Data& PySpark(2 weeks)

1. **What is Big Data?**
    - Introduction to Big Data technologies
    - Hadoop ecosystem vs. Spark ecosystem

- Distributed computing fundamentals

2. **Apache Spark Basics**
    - What is Apache Spark?
    - Spark components(Core, SQL, MLlib, GraphX)
    - Setting up Spark locally and in Data bricks
3. **Spark RDD(Resilient Distributed Dataset)**
    - Understanding RDDs
    - Transformations(map, filter, flat Map)
    - Actions(collect, count, save)
    - Persisting RDDs
4. **PySpark Data Frames**
    - Introduction to Data Frames in PySpark
    - Creating Data Frames
    - Data Frame transformations and actions
5. Operations on Data Frame (select, filter, groupBy, join)SparkSQL
    - SQL queries in Spark
    - Registering Data Frames as SQL tables
    - Usingspark.sql()to run SQL queries

Working with Data bricks(1 week)

1. **Introduction to Data bricks**
    - Setting up a Data bricks workspace

- Data bricks note books
- Cluster setup and management
- Running jobs on Data bricks

2. **Data bricks Notebooks for Data Analysis**
   - Basic note book operations(cells, markdown, etc.)
   - Visualizations in Data bricks
   - Using Data bricks notebooks for collaboration
3. **Integrating PySpark with Data bricks-**
   - Writing PySpark code in Data bricks note books
   - Running PySpark jobs on Data bricks clusters
   - Accessing data from Data bricks file system
4. **Using Data bricks with Cloud Storage**
   - Working with AWSS3, Azure Blob Storage and GCP Storage in Data bricks
   - Data loading and saving(CSV, Parquet, JSON)

Advanced PySpark and Data bricks(1week)

1. **Advanced PySpark Topics**
   - Window functions
   - Advanced aggregations
   - Handlings kewed data
   - PySpark performance tuning
2. **Streaming with PySpark**
   - Introduction to Spark Streaming
   - Real-time data processing using PySpark Streaming
   - Kafka integration with PySpark

3. **Optimizing Spark Jobs**
   - Caching and persisting RDDs/Data Frames
   - Partitioning and shuffling
   - Tuning Spark jobs for performance
4. **Advanced Data bricks Features**
   - Data bricks jobs and work flows
   - Scheduling note books and jobs
   - Collaborating with team members using Data bricks

================== Connect with US ===================

- **Connect with me in LinkedIn :** https://www.linkedin.com/in/bollepalli-ashok/
- **Subscribe To Our YouTube Channel :** https://www.youtube.com/c/AshokIT?sub_confirmation=1
- **Visit Our Website :** https://www.ashokit.in/
- **Follow us in whatsapp channel For More Job Alerts :** https://www.whatsapp.com/channel/0029Va9NnSdCHDyqwAoeIB1G