



Spark Developer Course:

Course Overview

This 32 hours Instructor lead training course is aimed at aspiring SPARK developer to teach Spark programming fundamentals and advance programming. This course introduces the Apache Spark distributed computing engine, and is suitable for

developers, data analysts, architects, technical managers, and anyone who needs to use Spark in a hands-on manner. The course provides a solid technical introduction to the Spark architecture and how Spark works. It covers the basic building blocks of Spark (e.g. RDDs and the distributed compute engine), as well as higher-level constructs that provide a simpler and more capable interface. It includes in-depth coverage of Spark SQL, DataFrames, and DataSets, which are now the preferred programming API. This includes exploring possible performance issues and strategies for optimization. The course also covers more advanced capabilities such as the use of Spark Streaming to process streaming

data, and integrating with the Kafka server

Prerequisites

Students should be familiar with any programming language like Scala, Python, Java or SQL and basic Unix knowledge is helpful. It is very nice to have previous Hadoop experience but not mandatory.

Course outline

Developers will learn to build simple Spark applications for Apache Spark version 2.1. You will use Spark's interactive shell to load and inspect data, then learn about the various modes for launching a Spark application. Also covered are working with DataFrames, datasets, and User-Defined Functions (UDFs).

Big Data overview

Big Data Use case

Hadoop Overview

HDFS Overview

HDFS commands

Yarn Architecture/ Overview

Core Spark:

SPARK Overview

Spark RDD:



The purpose & function of RDD

Spark programming basics

Spark transformation

Spark actions

Multiple RDD's

Pair RDD

RDD Partitioning and Transformation

Spark Streaming

Describe Spark Streaming

Create and view basic data streams

Perform basic transformations on streaming data

Utilize window transformations on streaming data

Spark SQL

Spark SQL components

An Overview of SPARK Data frame

DataFrames & tables

Creating Data frames

Manipulating Dataframes

Spark Data frame Programming

Data frame Transformation and Action

Data frame SQL based query's

Spark dataset Overview

Spark data set Programming

Spark dataset transformation & Action

Spark programming model



Lab demonstration Spark Dataframe & Dataset

Spark Job monitoring

- Spark Job structure

- Spark Application UI

- Spark performance Tuning

- Broadcast Variables

- Joining strategies

Spark programming for Grouping, Reducing & Joining

- Using Spark Variables (Broadcast & Accumulator)

- Spark program caching, Storage

- Spark programming shuffling

Spark Application submission

- YARN client mode

- Yarn cluster mode

- Spark configuration

Spark programming tuning

- Spark programming Optimization

- Spark API's

- Building & Running Spark application

- Spark cluster mode

- Spark YARN mode

Spark Machine learning Overview

Course Duration

The total duration for this course is 32 hours



About AiQuest

The mission of Ai Quest (AiQ) is to help organizations and knowledge workers to explore and realize their true potential in the Artificial Intelligence (AI) landscape. The true potential of Big Data in the AI realm goes beyond implementing new technologies and having appropriate data analytics. The strategy must include well trained resources, right performance measures that affect the corporate performance, exploiting existing technological resources to maximize the value and continuous investment in corporate training.

Course design philosophy

We want to bring corporate quality and industry standard training to individuals seeking a career in Big Data. Our courses are modelled based on extensive industry experience and cater to current Industry needs to provide relevant practical experience and real-time working knowledge. Our elite courses cover core concepts in Big data as offered by corporate solution partners-Horton Works, Cloudera and Pivotal.

For those looking to certify, the course has been designed specifically to help take the certification examination with ease. Also, the courses are designed with an ideal theory to practical ratio of 50:50, ensuring learning conceptual knowledge backed by practical applicable skills relevant to the work force. The courses are delivered by professional trainers who offer corporate trainings to companies and are working as consultants and architects on Big Data projects.