

Apache Spark Big Data Boot Camp

3 Days Classroom Session | 3 Days Live Online

Course Overview

Apache Spark: the second-generation big data platform that supplants and improves on Hadoop

The speed, expanded versatility, and new access to powerful APIs and libraries make Apache Spark the undisputed new toolset for powering big data solutions with distributed cluster computing. Also, for the first time ever Spark gives applications the ability to support serious data science capabilities with R-type dataframes and big data streaming to overcome time constraints.

Learn to use Spark for your own applications in three packed hands-on days

This fast-paced 3-day course is for data engineers, data analysts, data scientists, developers and operations teams and provides a thorough, hands-on overview of the Apache Spark Platform and various technologies and paradigms which are in Apache Spark.

- We will explore Apache Spark, how it came into existence, how it compares with Apache Hadoop – currently the de facto big data standard – and the new use cases that can be realized with Apache Spark as well as how your current use cases can be made more performant and powerful.
- We will also look at Apache Spark's Streaming Architecture which can help realize most of the real time-constrained needs of your business. We will also explore Apache Spark's SQL Architecture which provides very fast migration from traditional slower analytical tools like Hive to SparkSQL.
- We will spend some time on Apache Spark ML/ML Lib which provide a total integrated Architecture with both real-time and batch analytics.
- Finally, we will also look at Apache Spark GraphX which deals with Graph Algorithms.

All these workshops are delivered with guided hands-on labs allowing attendees to explore the data and the techniques and familiarize themselves with the various paradigms.