knóldus

Getting Productive with Spark in 2 Days

Developer to Developer Training Series

www.knoldus.com

Course Outline

Apache Spark is the nextgeneration successor to MapReduce. Spark is a powerful, opensource processing engine for data in the Hadoop cluster, optimized for speed, ease of use,and sophisticated analytics. The Spark framework supports streaming data processing and complex, iterative algorithms, enabling applications to run up to 100x faster than traditional Hadoop MapReduce programs.

The 2 day Spark course is aimed at developers who are encountering Spark for the first time and want to understand how to build Big Data Products with Spark. The course would enable participants to build complete, unified Big Data applications combining batch, streaming, and interactive analytics on all their data.

Developers would be able to write sophisticated parallel applications to execute faster decisions, better decisions, and realtime actions, applied to a wide variety of use cases, architectures, and industries.

The course has a practical focus, mixing presentation with indepth handson labs and exercises.

Proposed Structure

Day 1	Day 2
First brush	**Spark Streaming**
◆ Big Data Why and What?	◆ Overview
◆ Introduction to Spark	◆ Streaming operations
◆ Spark shell	◆ Sliding window operations
◆ Programming with Spark	◆ Streaming Applications
RDDs	**Good to know**
◆ Resilient Distributed Datasets	◆ Spark Context
◆ RDD Operations	◆ Spark Properties
◆ Map Reduce	◆ Logging
◆ Key value pair	◆ Iterative Algorithms
	◆ Graph Analysis

	◆ Machine Learning
	◆ Spark SQL
Running Spark	**Spark and Hadoop**
◆ Stand alone	◆ HDFS
◆ Web UI	◆ Using HDFS with Spark
◆ Stand alone cluster	◆ Spark and MapReduce
◆ Building and running	
Parallel Programming	**Q&A**
◆ Partitions and Data Locality	
◆ Executing parallel operations	
Caching and Persistence	
◆ Caching Overview	
◆ Distributed Persistence	

Course Prerequisites

To benefit from this course you should have programming experience with Scala or with Python. The language of instruction is Scala. Basic Linux knowledge is expected.

- For more information on the course or a discussion on your custom need, send a mail to info@knoldus.com