knóldus

Developer to Developer Training Series

Getting Productive with Spark - 2 Days

Course Outline

Apache Spark is the next-generation successor to MapReduce. Spark is a powerful, open-source 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 real-time actions, applied to a wide variety of use cases, architectures, and industries.

The course has a practical focus, mixing presentation with in-depth hands-on labs and exercises.

Proposed Structure

Day 1	Day 2
First brush • Big Data Why and What? • Introduction to Spark • Spark shell • Programming with Spark	**Spark Streaming** • Overview • Streaming operations • Sliding window operations • Streaming Applications

knóldus

Developer to Developer Training Series

RDDs

- Resilient Distributed Datasets
- RDD Operations
- Map Reduce
- Key value pair

Running Spark

- Stand alone
- Web UI
- Stand alone cluster
- Building and running

Parallel Programming

- Partitions and Data Locality
- Executing parallel operations

Caching and Persistence

- Caching Overview
- Distributed Persistence

Good to know

- Spark Context
- Spark Properties
- Logging
- Iterative Algorithms
- Graph Analysis
- Machine Learning
- Spark SQL

Spark and Hadoop

- HDFS
- Using HDFS with Spark
- Spark and MapReduce

Q&A

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