inovex

Trainings

Developing Apache Spark Applications (Spark v 2.1)

Training for developers who want to build Big Data Applications using Apache Spark 2.1.

Dauer: 3 days

Zielgruppe: Software Architects, Software Developers

The Training sessions are usually held in German. Please contact us if you are interested in Training sessions in English.

This Training course teaches all the knowledge required to develop Big Data Applications based on Apache Spark 2.1. Participants first learn how to use the Spark Shell to load and interactively analyse records from various sources and formats. Building on this, participants develop a stand-alone Spark application to process data in the form of datasets and data frames locally or on a computing cluster. The Training concludes with an introduction to Spark streaming to process data streams, GraphFrame to analyse graphs and the MLlib Machine Learning library.

Prerequisites:

- Basic Hadoop skills

- Basic Linux skills (including command line options such as ls, cd, cp and su)

- Good Java or Scala skills

- Good SQL skills

Agenda:
1. Apache Spark Basics (DEV 360)

- Apache Spark features

- Spark framework components

- Case studies

2. Creating datasets

- Defining data sources, structures and schemas

- Working with datasets and data frames

- Converting data frames to datasets

Practice:

- Loading data and creating datasets using Reflection

- Simple case study: word count with datasets (optional)

3. Operations for datasets

- Basic operations on datasets

- Caching datasets

- User-defined functions (UDFs)

- Partitioning datasets

Practice:

- Analysing SFPD data

- Creating and applying UDFs

- Analysing data with the help of UDF and queries

4. Developing a simple Apache Spark Application (DEV 361)

- Spark Application lifecycle

- Using SparkSession

- Starting Spark Applications

Practice:

- Importing and configuring Application files

- Building, deploying and starting Applications

5. Monitoring Apache Spark Applications

- Logical and physical Spark schedules

- Spark Web UI for monitoring Spark Applications

- Debugging and tuning Spark Applications

Practice:

- Using Spark UI

- Interpreting Spark system properties

6. Creating Apache Spark streaming Applications (DEV 362)

- Introduction to the Spark streaming architecture

- Developing Spark structured streaming Applications

- Applying operations to streaming data frames

- Developing your own Windows functions

Practice:

- Loading and analysing data using the Spark Shell

- Spark streaming in the Spark Shell

- Building and running a streaming application with SQL

- Building and running a streaming application with Windows function and SQL

7. Using Apache Spark GraphFrames
- Introduction to GraphFrame

- Defining regular, directed and property graphs

- Creating property graphs

- Perform operations on charts

Practice:

- Graph analysis with GraphFrames

8. Using Apache Spark MLlib
- Introduction to Apache Spark MLlib (Machine Learning Library)

- Collaborative filtering for user selection prediction

Practice:

- Data analysis using the Spark Shell

- Developing a Spark application for film recommendations

- Analysing a simple flight system with decision trees

Note:

- The course fee includes Training materials, lunches, drinks and snacks

- Participants have to bring their own notebook for Training.