

Training

## Apache Spark



**Target audience:** Analysts, Software Developers,  
Software Architects

**Length:** 2 days

### Abstract

Whether for batch or stream processing, thanks to its performance as distributed in-memory technology, Apache Spark has firmly established itself among the Big Data tools ecosystem within a short space of time.

This Training course provides an introduction to Spark as a tool for analysing large volumes of data and covers both batch and streaming processes. The course emphasises the formulation of analytical queries and the use of Machine Learning processes. The participants are given specific business requirements and introduced to the architecture, techniques and tools they will need to fulfil these appropriately.

This course is heavily practice-focused. It centres on a complex database in which the participants practice methods, tools and techniques.

### Agenda:

- Spark basics and architecture
- Spark APIs and the RDD data structure
- Formulating queries with Spark SQL
- Transformations and actions in the Spark context
- Zeppelin as a Spark frontend

- Machine learning using the Spark MLlib
- Overview of the Apache Spark ecosystem
- Designing Spark architectures for implementing specific use cases

**Note:**

- The course fee includes Training materials, certificates of participation, lunches, drinks and snacks.
- Participants must bring their own laptop to the Training sessions.

If you have any questions about the Training, need a hotel recommendation or have any other concerns about our Training, please contact us at [trainings@inovex.de](mailto:trainings@inovex.de) or our Academy Director Collin Rogowski on tel. +49 (0) 172 5673497. Many thanks ! We look forward to you!