

# Apache Spark Application Performance Tuning

This hands-on training course teaches the key concepts and expertise developers need to optimize the performance of their Apache Spark applications. During the course, you will learn how to identify common causes of poor performance in Spark applications, techniques to avoid or resolve them, and best practices for monitoring Spark applications.

The course introduces the architecture and concepts behind Apache Spark and the underlying data platform, and then builds on this foundational understanding by teaching you how to optimize Spark application code. The course format focuses on instructor-led demonstrations that illustrate both performance issues and the techniques that address them, followed by hands-on exercises that give you the opportunity to practice what you've learned in an interactive notebook environment.

#### Course Contents

- Spark Architecture
- Data Sources and Formats
- Inferring Schemas
- Dealing With Skewed Data
- Catalyst and Tungsten Overview
- Mitigating Spark Shuffles
- Partitioned and Bucketed Tables
- Improving Join Performance
- Pyspark Overhead and UDFs
- Caching Data for Reuse
- Workload XM (WXM) Introduction
- What's New in Spark 3.0?

**E-Book** You will receive the original documentation as a E-Book (pdf).

#### Target Group

This course is aimed at software developers, engineers and data scientists who have experience developing Spark applications and want to learn how to improve the performance of their code. This is not an introduction to Spark.

#### Prerequisites

Spark examples and practical exercises will be presented in Python and the ability to program in this language is required. Basic familiarity with the Linux command line is required. Basic knowledge of SQL is helpful.

We also recommend our training courses in Programming languages and software development and Linux.

#### Course Target

Once you have successfully completed this course, you will be able to:

- Understand the architecture and job execution of Apache Spark and how techniques such as lazy execution and pipelining can improve runtime performance
- Evaluate the performance characteristics of core data structures such as RDD and DataFrames
- Select the file formats that offer the best performance for your application
- Identifying and resolving performance issues caused by data distortion
- Use partitioning, bucketing and join optimizations to improve SparkSQL performance
- Understanding the performance overhead of Python-based RDDs, DataFrames and user-defined functions
- Take advantage of caching for better application performance
- Understand how the Catalyst and Tungsten optimizers work
- Learn how Workload XM can help troubleshoot and proactively monitor Spark application performance
- Learn how the adaptive query execution engine improves performance

#### This Course in the Web



You can find the up-to-date information and options for ordering under the following link:

[www.expertech-training.com/go/CLTU](http://www.expertech-training.com/go/CLTU)

#### Reservation

On our Website, you can reserve a course seat for 7 days free of charge and in a non-committal manner. This can also be done by phone under +49 6074/4868-0.

#### Guaranteed Course Dates

To ensure reliable planning, we are continuously offering a wide range of guaranteed course dates.

#### Your Tailor-Made Course!

We can precisely customize this course to your project and the corresponding requirements.

| Training                  | Prices, excl. of V.A.T. |                |
|---------------------------|-------------------------|----------------|
| <b>Classes in Germany</b> | <b>3 Days</b>           | <b>€ 2,795</b> |
| <b>Online Training</b>    | <b>3 Days</b>           | <b>€ 2,795</b> |
| <b>Date/course venue</b>  | Course language English |                |
| 27/04-29/04/26            | ONOnline                |                |

Status 03/25/2026



## Our Training Offer for You:



### Classroom Training

The live training experience in our training centers or on your premises on-site.



### Online Training

Attend your course in online mode – without travel and accommodation expenses and effort.



### Hybrid Training

Classroom & online in one course – choose how you wish to participate in your training event.



### In-house Training

We create your customized training concepts for your project.



### Guaranteed Course Dates

ExperTeach guaranteed course dates allow for secure and reliable planning.

## Awards Won by ExperTeach



ExperTeach Benelux B.V.

Ceresstraat 1 · 4811 CA Breda · Phone: +49 6074 4868-0 · Fax: +49 6074 4868-109 · info@experteach.de · www.experteach-training.com