

**Apache Spark Training Course**

**Dates and locations:**

* 13- 14 December 2017, London
* 22-23 February 2018, London
* 9-10 April 2018, London

**Price:**

£1495+VAT

**Background:**

*Learn to create real-time data analysis & ML solutions with Spark!*

The Apache Spark Training Course is aimed at Data Scientists, Analysts, software developers, architects, who need to gain hands-on experience using Apache Spark to create real-time Data Stream analysis and large-scale Machine Learning solutions.

The Spark training course has recently been expanded to include even more hands-on exercises. You are encouraged to bring along your own laptop so you can learn in a familiar environment and take away everything you have worked on during the class, to implement in your own projects or to display in your portfolio of work.

You'll be guided by an industry expert who has first-hand experience of designing and implementing commercial-scale Big Data analysis solutions.

**Summary in a nutshell:**

By the end of this course, you will have learnt:

* Apache Spark architecture
* How to use Spark with Scala
* How to integrate Spark with NoSQL and other Big Data technologies
* How to scale calculations to a cluster of servers
* How to deploy Spark projects to the Cloud
* Machine Learning with Spark

**Who should attend:**

This course is aimed at Data Scientists, Analysts, Software Developers, Database Developers, Data Warehouse Managers & Business Intelligence Specialists, Software Architects.

**Pre-requisites:**

During the Spark course we will give you some fairly simple Scala code examples to run and edit.

It would be ideal if you have some experience of software development / scripting, or database development using a SQL-based RDBMS (e.g. SQL Server, Oracle, MySQL, DB2...).

If you are from a more dashboard-oriented Business Intelligence background (or have good knowledge of a platform such as Excel, SAS etc) you should also benefit from this course - please let us know if you have any questions or concerns.

**Preparing for the course:**

If you wish to participate in the hands-on exercises you should sign up for an Amazon AWS account at least 48 hours prior to the course: http://aws.amazon.com/ - and don't forget to save your login details! You may incur around $10 - $20 of cloud storage and computation usage during the course.

**Course syllabus:**

Big Data Fundamentals

* Overview of Hadoop/HDFS and Amazon AWS/S3

Spark and TDD

* Creating a Spark Project in Intellij
* Running and debugging a Spark project
* Building and deploying a Spark Project with SBT on AWS
* Spark Core (RDD)

Spark SQL

Spark Streaming with Kafka

* Kafka and Soak Streaming Examples

Spark Machine Learning

Real-time Analytics with Spark

* Spark architecture
* Installation and running Spark in the Cloud
* Programming with Spark
* Streaming data with Spark
* Integrating Spark with NoSQL and other Big Data technologies
* Spark demo + avro + pig + hive
* Spark and Kafka integration

Streaming algorithms

* Dynamic sampling
* Distinct count, cardinality estimation
* HyperLogLog
* Moving average

Integration with third-party applications and languages

* Python
* R – examples for beta distribution
* Hadoop
* Lambda architecture

**Contact:**

[www.unicom.co.uk](http://www.unicom.co.uk) | info@unicom.co.uk | +44 (0) 1895 256 484