

**Hadoop Training Course - Big Data Training**

**Dates and locations:**

* 11- 12 December 2017, London
* 20-21 February 2018, London
* 11-12 April 2018, London

**Price:**

£1495+VAT

**Background:**

This course is designed to show Software Developers, DBAs, Business Intelligence Analysts, Software Architects and other vested stakeholders how to use key Open Source technologies in order to derive significant value from extremely large data sets.

We will show you how to overcome the challenges of managing and analysing Big Data with tools and techniques such as Apache Hadoop, NoSQL databases and Cloud Computing services.

Our Big Data with Hadoop course features extensive hands-on exercises reflecting real-world scenarios, and you are encouraged to take these away to kick-start your own Big Data efforts.

The course is delivered by an industry expert with extensive experience of implementing cutting-edge high-performance Data Analysis platforms and processes in large-scale retail, marketing and scientific projects.

**Summary in a nutshell:**

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

* Big Data Patterns and Anti-Patterns
* Hadoop, HDFS, MapReduce with examples
* NoSQL Databases with demonstrations in Cassandra, HBase and others
* Building Data Warehouses with Hive
* Integration with SQL Databases
* Parallel Programming with Pig
* Machine Learning & Pattern Matching with Apache Mahout
* Utilise Amazon Web Services

**Who should attend:**

The Hadoop Training Course is aimed at Data Scientists, Business Intelligence Analysts, Software Developers, Software Architects who are looking to employ the Hadoop stack to analyse large unwieldy databases - be it marketing / retail data, scientific data sets, banking & financial reports, document stores - the sky is the limit.

**Pre-requisites:**

Delegates should have an understanding of Enterprise application development, business systems integration and or Database Design / Querying / Reporting.

For the hands-on Hadoop exercises, delegates should sign up for an Amazon AWS account prior to the course: http://aws.amazon.com/ and bring their login details. Service usage is not likely to exceed $10 USD per person.

**Course syllabus:**

Hadoop Architecture

* History of Hadoop – Facebook, Dynamo, Yahoo, Google
* Hadoop Core
* Yarn architecture, Hadoop 2.0

Hadoop Distributed File System (HDFS)

* HDFS Clusters – NameNodes, DataNodes & Clients
* Metadata
* Web-based Administration

MapReduce

* Processing & Generating large data sets
* Map functions
* Programming MapReduce using SQL / Bash / Python
* Parallel Processing
* Failover

Data warehousing with Hive

* Data Summarisation
* Ad-hoc queries
* Analysing large datasets
* HiveQL (SQL-like Query Language)
* Integration with SQL databases
* n-grams analysis

Parallel Processing with Pig

* Parallel evaluation
* Query language interface
* Relational Algebra

Data Mining with Mahout

* Clustering
* Classification
* Batch-based collaborative filtering

Searching with Elastic Search

* Elastic search concepts
* Installation, import of the data
* Demonstration of API, sample queries

Structured Data Storage with HBase

* Big Data: How big is big?
* Optimised Real-time read/write access

Cassandra multi-master database

* The Cassandra Data Model
* Eventual Consistency
* When to use Cassandra

Redis

* Redis Data Model
* When to use Redis

MongoDB

* MongoDB data model
* Installation of MongoDB
* When to use MongoDB

Kafka

* Kafka architecture
* Installation
* Example usage
* When to use Kafka

Lambda Architecture

* Concept
* Hadoop + Stream processing integration
* Architecture examples

Big Data in the Cloud

* Amazon Web Services
* Concepts: Pay pay use model
* Amazon S3, EC2, EMR
* Google Cloud Platform
* Google Big Query

**Contact:**

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