



BIG DATA TRAINING

Course Background



This 3-day Hadoop & Storm Big Data training course combines Big Data with an Apache Hadoop Training Course and Real-Time Big Data Processing with Spark & Storm training courses.

During this workshop you will gain valuable hands-on experience using platforms including Apache Hadoop, Spark, and Elastic Search to process, analyse and manipulate large data sets, using real-world scenarios.

This Big Data with Hadoop training 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.

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

Key Learning Points from the course:



- ✓ 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
- ✓ Spark & Storm architecture
- ✓ How to use Spark with Java
- ✓ 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

Organised by:



BIG DATA TRAINING



Who Should Attend?



Data Warehouse Managers & Business Intelligence Specialists, Software Developers, Database Developers, Software Architects, IT Managers & IT Directors

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 TRAINING



Big Data in the Cloud

- ✓ Amazon Web Services
- ✓ Concepts: Pay per use model
- ✓ Amazon S3, EC2, EMR
- ✓ Google Cloud Platform
- ✓ Google Big Query

Streaming algorithms

- ✓ Dynamic sampling
- ✓ Distinct count, cardinality estimation
- ✓ HyperLogLog
- ✓ Moving average

Real-time Analytics with Spark

- ✓ Spark architecture
- ✓ Installation and running Spark in theCloud
- ✓ Programming with Spark
- ✓ Streaming data with Spark
- ✓ Integrating Storm with NoSQL and other Big Data technologies
- ✓ Spark demo + avro + pig + hive
- ✓ Spark and Kafka integration

Integration with third-party applications and languages

- ✓ Python
- ✓ R – examples for beta distribution
- ✓ Hadoop
- ✓ Lambda architecture

Real-time Analytics with Spark



No prior experience of working with Big Data tools or platforms is required. Delegates should ideally have an understanding of Enterprise application development, Business Systems Integration and / or database development. This course incorporates hands-on exercises using cloud-based environments. If you wish to participate in the hands-on exercises you should sign up for an Amazon AWS account prior to the course: <http://aws.amazon.com/> and bring your login details

Bring your own device



For this course it is necessary to bring your own device. Laptops can be provided if requested beforehand.

Ticket Prices

13 Jun 2016, London	3 Days Course	Price:	£ 1895.00 +VAT
---------------------	---------------	--------	----------------

12 Sep 2016, London	3 Days Course	Price:	£ 1895.00 +VAT
---------------------	---------------	--------	----------------

Contact



UNICOM Seminars Ltd
OptiRisk R&D House
One Oxford Road
Uxbridge UB9 4DA, UNITED KINGDOM



+44 (0) 1895 256484



+44 (0) 1895 813095



www.unicom.co.uk



Info@unicom.co.uk