



Data Sheet

Apache Hadoop Masterclass

This one-day course is designed to help both IT professionals and decision-makers understand the concepts and benefits of Apache Hadoop and how it can help them meet business goals.

You will gain an understanding of the Hadoop technology stack, including MapReduce, HDFS, Hive, Pig, HBase, and provides an initial introduction to Mahout and other common utilities.

Course Objectives

This informative class will help you better understand:

- The essential components of a Hadoop-based data management solution
- Pros and cons of implementing Hadoop
- How does Hadoop fit into our existing environment and architecture?
- The differences between various Hadoop distributions

Course Outline

Why Hadoop?

- History & background
- Real-world use cases and case studies

The Hadoop Platform

- Introduction to MapReduce and Hadoop File System (HDFS)
- Data warehousing with Hive
- Parallel processing with Pig
- Data mining with Mahout
- Data storage with HBase
- Common utilities - Sqoop, Flume, Hue, Scribe, Zookeeper, HCatalog
- Hadoop distributions - Apache Foundation, Cloudera, Hortonworks, MapR, IBM & others

The future of Hadoop

- YARN - Next generation MapReduce
- Other programming paradigms on Hadoop

Additional Information

The course content can be customised to cover any specialised material you may require for your specific training needs.

This course is offered as private on-site training to be hosted at your offices. For more information, please contact us at training@bigdatapartnership.com.



Big Data Partnership is the leading EMEA-based big data service provider in data science, data engineering and certified training, helping unlock value from complex data. With a team of industry experts and thought leaders in big data technologies including the Apache Hadoop ecosystem, Apache Cassandra and NoSQL databases, Big Data Partnership has the knowledge and experience needed to Discover, Develop & Deliver scalable and reliable big data solutions, driving exciting results for its customers.

Big Data Partnership is a Systems Integrator and Certified Training Partner of Hortonworks, MapR Technologies & DataStax.

© 2013 Big Data Partnership Ltd. All rights reserved. Apache Hadoop and Hadoop are trademarks of the Apache Software Foundation and not affiliated with Big Data Partnership.

Tel: +44 (0)20 7205 2550
Email: training@bigdatapartnership.com
www.bigdatapartnership.com/training

1 Bath Street
London EC1V 9LB
United Kingdom