

# Introduction to Hadoop and Map Reduce Training

C123-102

Learn and experience Hadoop HDFS, Map Reduce and Apache Spark

This training introduces the HADOOP ecosystem with emphasis on Map Reduce and Hadoop architecture. and provides basic hands on experience on each of the solutions themselves.

## AUDIENCE

System Administrators

Operations

Database Administrators

Support

DevOps

Developers

## KNOWLEDGE REQUIREMENTS

Relational Database SQL experience

## LENGTH

3 Days

## BONUS

Hands-on lab sessions

## SYLLABUS

Course Introduction  
Introduction to Big Data  
Introduction to Hadoop and HDFS  
Hadoop Map Reduce  
Apache Spark  
Introduction to Hive  
Other Hadoop Components is short  
Data Architect/Analyst for Big Data  
Summary

## HARDWARE AND SOFTWARE REQUIREMENTS

### Computer Requirements

- RAM: minimum 8 GB of RAM required for exercises and platform to operate, 16 GB and up recommended.
- Disk Space: At least 30 GB of free disk space
- Internet connection
- All machines connected to same Network

### Supported Operating Systems

- 123Completed provides a VM ready for training session. A Linux Ubuntu

### Additional Software Requirements

- PDF Reader
- VMPlayer 14.0 and up
- Zip software

### Class HW required

- Projector 1024\*768 minimum resolution
- White Board
- Erasable Markers
- Desktops or Laptops (see HW Requirements)
- 12-24 ports Switch
- Internet connectivity
- Electricity outlets for all computers/monitors and other equipment.
- At least 3 electricity outlets next to instructor location.

AGENDA	
<b>Day 1: Lesson 1: Course Introduction</b>	Duration: 30 minutes
<ul style="list-style-type: none"> <li>• Course Introduction</li> <li>• Courseware walkthrough</li> <li>• Documentation</li> <li>• Lab</li> </ul>	
<b>Day 1: Lesson 2: Introduction to Big Data</b>	Duration: 0.5 Day
<ul style="list-style-type: none"> <li>• What is Big Data?</li> <li>• Big Data challenges and complexity</li> <li>• General concepts</li> <li>• Architecture considerations</li> <li>• Presenting use cases of internet companies (e.g. Facebook)</li> <li>• The Data Scientist</li> <li>• RDBMS: Advantages and disadvantages / Impedance Mismatch</li> <li>• CAP theorem vs. ACID / Dynamic schema, sharding, replications and caching / Performance</li> <li>• Scaling vs consistency</li> </ul>	
<b>Day 1+2: Lesson 3: Introduction to Hadoop and HDFS</b>	Duration: 0.5 Day
<ul style="list-style-type: none"> <li>• Hadoop Pseudo-Distributed Mode installation</li> <li>• Lab – Hadoop Installation</li> <li>• HDFS Assumptions and Goals</li> <li>• Scale and Feature requirements</li> <li>• Architecture</li> <li>• Data Replication</li> <li>• Robustness</li> <li>• Data Organization</li> <li>• Accessibility</li> <li>• Space Reclamation</li> <li>• Lab (Install Configure and Experience)</li> </ul>	
<b>Day 2: Lesson 4: Hadoop Map Reduce</b>	Duration: 0.5 Day
<ul style="list-style-type: none"> <li>• Map Reduce Basics</li> <li>• Inputs and Outputs</li> <li>• Thinking Map Reduce</li> <li>• Map Reduce Code Example</li> <li>• Map Reduce Additional Details</li> <li>• Map Reduce V1 vs. V2</li> <li>• Hadoop 2.0 - YARN – Yet Another Resource Negotiator.</li> <li>• Lab (Install Configure and Experience)</li> </ul>	
<b>Day 3: Lesson 5: Introduction to Apache SPARK</b>	Duration: 0.5 Day
<ul style="list-style-type: none"> <li>• Apache SPARK Introduction</li> <li>• Getting Started</li> <li>• SPARK architecture</li> <li>• SPARK processing</li> <li>• Map Reduce</li> <li>• Example</li> <li>• Lab (Install Configure and Experience)</li> </ul>	

<b>Day 4: Lesson 6: Other Hadoop Components is short</b>	<b>Duration: 0.5 Day</b>
<ul style="list-style-type: none"> <li>• HIVE</li> <li>• HBase</li> <li>• PIG</li> <li>• HUE – UI</li> <li>• Zookeeper</li> <li>• Connectors</li> </ul>	
<b>Day 3: Lesson 7: The Data Architect/Analyst</b>	<b>Duration: 0.5 Day</b>
<ul style="list-style-type: none"> <li>• The Data Architect</li> <li>• Every Data element should be analyzed</li> <li>• Data Characteristics (Read rarely, Read once, Read many, Write Once, Write Many (Updated after inserted)</li> <li>• Read if Exist</li> </ul>	
<b>Day 3: Lesson 8: Summary</b>	<b>Duration: 30 minutes</b>
<ul style="list-style-type: none"> <li>• Summary</li> <li>• Wrap Up</li> </ul>	