



## Professional Hadoop Solutions

Course ISI-1415 4 Days Instructor-led, Hands-on

### Introduction

If you are ready to make the most of massively scalable analytics, you need to know how to take advantage of Hadoop technologies. Today's enterprise architects need to understand how the Hadoop frameworks and APIs fit together, and how they can be integrated to deliver real-world solutions.

With this course, you'll learn how! It focuses on the architecture and approaches for building advanced, Hadoop-based enterprise applications, taking a deeper, code-level dive into delivering real-world solutions. It also walks you through the data design and how it impacts implementation. It explains how MapReduce works, and shows you how to reformulate specific business problems in MapReduce. You'll find in-depth Java code examples that you can use, derived from applications that have been successfully built and deployed.

### At Course Completion

After completing this course, students will be able to:

- Explores the MapReduce architecture, its main components, and the MapReduce programming model
- Discusses how to create reliable MapReduce applications, including testing and debugging, as well as how to use built-in MapReduce facilities
- Explains how to extend Oozie and use it to integrate Hadoop-based implementations with other enterprise applications
- Describes how to build real-time Hadoop-based applications and how to leverage real-time Hadoop queries
- Shows how to extend Hadoop's security capabilities, including encryption, authentication, authorization, Single-Sign-On (SSO), and auditing
- Covers different approaches to running Hadoop in the Amazon Web Services (AWS) cloud

### Prerequisites

Prior knowledge of Hadoop is not required, but attendees should have basic experience with Linux.

Contact **ISI Inc** for more information at **916.920.1700** or by visiting our website at <http://www.isinc.com>



## Course Materials

The student kit includes a comprehensive workbook and other required materials for this class.

## Course Outline

### Module 1. Big Data and the Hadoop Ecosystem

- Big data meets Hadoop
- The Hadoop ecosystem
- Hadoop core components
- Developing enterprise applications with Hadoop

### Module 2: Storing Data in Hadoop

- HDFS
- Hbase
- Combining HDFS and HBase for effective data storage
- Using Apache Avro
- Managing metadata with HCatalog
- Choosing an appropriate Hadoop data organization for your applications

### Module 3: Processing Your Data with MapReduce

- Getting to know MapReduce
- Your first MapReduce application
- Designing MapReduce implementations

### Module 4: Customizing MapReduce Execution

- Controlling MapReduce execution with InputFormat
- Reading data your way with custom RecordReaders
- Organizing output data with custom output formats
- Writing data your way with custom RecordWriters
- Optimizing your MapReduce execution with the Combiner
- Controlling reducer execution with partitioners
- Using Non-Java Code with Hadoop

### Module 5: Building Reliable MapReduce Apps

- Unit testing MapReduce applications
- Local application testing with Eclipse
- Using logging for Hadoop testing
- Reporting metrics with job counters
- Defensive programming in MapReduce

### Module 6: Automating Data Processing with Oozie

- Getting to know Oozie
- Oozie workflow
- Oozie coordinator

Contact ISInc for more information at 916.920.1700 or by visiting our website at <http://www.isinc.com>



- Oozie bundle
- Oozie parameterization with Expression language
- Oozie job execution model
- Accessing Oozie
- Oozie SLA

### **Module 7: Using Oozie**

- Validating information about places using probes
- Designing place validation based on probes
- Designing Oozie workflows
- Implementing Oozie workflow applications
- Implementing workflow activities
- Implementing Oozie coordinator applications
- Implementing Oozie bundle applications
- Deploying, testing and executing Oozie applications
- Using the Oozie console to get information about Oozie applications

### **Module 8: Advanced Oozie Features**

- Building custom Oozie workflow actions
- Adding dynamic execution to Oozie workflows
- Using the Oozie Java API
- Using Uber Jars with Oozie applications
- Data ingesting conveyer

### **Module 9: Real-Time Hadoop**

- Real-time applications in the real world
- Using Hbase for implementing real-time applications
- Using specialized real-time Hadoop query systems
- Using Hadoop-based event-processing systems

### **Module 10: Hadoop Security**

- Understanding Hadoop security challenges
- Authentication
- Authorization
- Oozie authentication and authorization
- Network encryption
- Security enhancements with Project Rhino
- Best practices for securing Hadoop

### **Module 11: Running Hadoop Applications on AWS**

- Getting to know AWS
- Options for running Hadoop on AWS
- Understanding the EMR-Hadoop relationship
- Using AWS S3
- Automating EMR job flow create and job execution
- Orchestrating job execution in EMR

**Contact ISInc for more information at 916.920.1700 or by visiting our website at <http://www.isinc.com>**



## **Module 12: Building Enterprise Security Solutions for Hadoop Implementations**

- Security concerns for enterprise applications
- What Hadoop security doesn't natively provide for enterprise applications
- Approaches for securing enterprise applications using Hadoop

## **Module 13: Hadoop's Future**

- Simplifying MapReduce Programming with DSLs
- Faster, more scalable processes
- Security enhancements
- Emerging trends