Yes-M Systems

Big Data Hadoop

From Yes-M Systems LLC

With Interview preparations, Resume preparations and Marketing Help

Student Location – To students from around the world

Delivery Method: Instructor-Led – Live online Training

A. Contact us for more details:

Company name: Yes-M Systems

Website: http://myyesm.com,

Phone numbers (USA): 678-643-7777, 678-248-0302
Phone number (India): 91-8220006968

Kudzu Reviews: http://www.kudzu.com/m/Yes-M-Systems-30363491/reviews/

Facebook: http://www.facebook.com/yesmsystems

Youtube: http://www.youtube.com/yesmsystems

B. Big Data Hadoop TOC Course Details

Module I. Introduction to Big Data and Hadoop

* What is Big Data?
* What are the challenges for processing big data?
* What technologies support big data?
*3V's of BigData and Growing.
* What is Hadoop?
* Why Hadoop and its Use cases
* History of Hadoop
* Different Ecosystems of Hadoop.
* Advantages and Disadvantages of Hadoop
* Real Life Use Cases

Module II. HDFS (Hadoop Distributed File System)

* HDFS architecture
* Features of HDFS
* Where does it fit and Where doesn't fit?
* HDFS daemons and its functionalities
* Name Node and its functionality
* Data Node and its functionality
* Secondary Name Node and its functionality
* Data Storage in HDFS
* Introduction about Blocks
* Data replication
*Accessing HDFS
* CLI(Command Line Interface) and admin commands
* Java Based Approach
*Hadoop Administration
*Hadoop Configuration Files
*Configuring Hadoop Domains
*Precedence of Hadoop Configuration
*Diving into Hadoop Configuration
*Scheduler
*RackAwareness
*Cluster Administration Utilities
*Rebalancing HDFS DATA
*Copy Large amount of data from HDFS
*FSImage and Edit.log file

Module III. MAPREDUCE

* Map Reduce architecture
* JobTracker , TaskTracker and its functionality
* Job execution flow
* Configuring development environment using Eclipse

* Map Reduce Programming Model
* How to write a basic Map Reduce jobs
* Running the Map Reduce jobs in local mode and distributed mode
* Different Data types in Map Reduce
* How to use Input Formatters and Output Formatters in Map Reduce Jobs
* Input formatters and its associated Record Readers with examples
* Text Input Formatter
* Key Value Text Input Formatter
* Sequence File Input Formatter
* How to write custom Input Formatters and its Record Readers
* Output formatters and its associated Record Writers with examples
* Text Output Formatter
* Sequence File Output Formatter
* How to write custom Output Formatters and its Record Writers
* How to write Combiners, Partitioners and use of these
* Importance of Distributed Cache
* Importance Counters and how to use Counters

Module IV. Advance MapReduce Programming

* Joins - Map Side and Reduce Side
* Use of Secondary Sorting
* Importance of Writable and Writable Comparable Api's
* How to write Map Reduce Keys and Values
* Use of Compression techniques
* Snappy, LZO and Zip
* How to debug Map Reduce Jobs in Local and Pseudo Mode.
* Introduction to Map Reduce Streaming and Pipes with examples
*Job Submission
*Job Initialization
*Task Assignment
*Task Execution
*Progress and status bar
*Job Completion
*Failures
*Task Failure
*Tasktracker failure
*JobTracker failure
*Job Scheduling
*Shuffle & Sort in depth
* Diving into Shuffle and Sort
* Dive into Input Splits
* Dive into Buffer Concepts
*Dive into Configuration Tuning
*Dive into Task Execution
*The Task assignment Environment

*Speculative Execution
*Output Committers
*Task JVM Reuse
*Multiple Inputs & Multiple Outputs
*Build In Counters
* Dive into Counters – Job Counters & User Defined Counters
* Sql operations using Java MapReduce
*** Introduction to YARN (Next Generation Map Reduce)**

Module V. Apache HIVE

* Hive Introduction
* Hive architecture
* Driver
* Compiler
* Semantic Analyzer
* Hive Integration with Hadoop
* Hive Query Language(Hive QL)
* SQL VS Hive QL
* Hive Installation and Configuration
* Hive, Map-Reduce and Local-Mode
* Hive DLL and DML Operations
* Hive Services
* CLI
*Schema Design
*Views
*Indexes
* Hiveserver

Metastore

* embedded metastore configuration
* external metastore configuration
* Transformations in Hive
* UDFs in Hive
* How to write a simple hive queries
* Usage
*Tuning
* Hive with HBASE Integration

Module VI. Apache PIG

* Introduction to Apache Pig
* Map Reduce Vs Apache Pig
* SQL Vs Apache Pig
* Different data types in Pig
* Modes of Execution in Pig
* Local Mode
* Map Reduce Mode

* Execution Mechanism
* Grunt Shell
* Script
* Embedded
* Transformations in Pig
* How to write a simple pig script
* UDFs in Pig

Module VII. Apache SQOOP

* Introduction to Sqoop
* MySQL client and Server Installation
* How to connect to Relational Database using Sqoop
* Sqoop Commands and Examples on Import and Export commands.
*Transferring an Entire Table
*Specifying a Target Directory
*Importing only a Subset of data
*Protecting your password
*Using a file format other than CSV
*Compressing Imported Data
*Speeding up Transfers
*Overriding Type Mapping
*Controlling Parallelism
*Encoding Null Values
*Importing all your tables
*Incremental Import
*Importing only new data
*Incrementing Importing Mutable data
*Preserving the last imported value
*Storing Password in the Metastore
*Overriding arguments to a saved job
*Sharing the MetaStore between sqoop client
*Importing data from two tables
*Using Custom Boundary Queries
*Renaming Sqoop Job instances
*Importing Queries with duplicate columns
*Transferring data from Hadoop
*Inserting Data in Batches
*Updating or Inserting at the same time
*Exporting Corrupted Data

Module VIII. Apache FLUME

* Introduction to flume
* Flume agent usage

Module IX. Apache OOZIE

* Introduction to Oozie
* Executing workflow jobs

Disclaimer: Yes-M Systems and/or their instructors reserve the right to make any changes to the syllabus as deemed necessary to best fulfil the course objectives. Students registered for this course will be made aware of any changes in a timely fashion using reasonable means.

C. About Yes-M Systems:

a. Established in 2005 (Atlanta, GA, USA), 10th year in business.

b. A+ accreditation from US Better Business Bureau (http://www.bbb.org/atlanta/business-reviews/internet-consultants/yes-m-systems-in-duluth-ga-27431372)

c. Received the "Best of 2012" and "Best of 2013" awards from US-based Kudzu (http://www.kudzu.com/m/Yes-M-Systems-30363491/reviews/)

d. Trained close to 3000+ students from all over the world.

e. Experienced, passionate and committed trainers

f. IT Training in various technologies including Java, Dot Net, SAP, Oracle, QA, BA etc (See Courses We offer section for more information)

g. Professional guidance/help with resumes and interview preparations.

h. Recruiter help with marketing/jobs

i. Certification at the end of the training.