

NIIT UNIVERSITY
NEEMRANA 301705, Dist. Alwar (Rajasthan)
DS 402 – Big Data Concepts
COURSE HANDOUT
January-May 2018

Course No.: DS 402

Course Title: Big Data Concepts

LTPC: 2-0-4-4

Course-in-Charge: Ram Narayan Yadav (Email: ramn.yadav@niituniversity.in)

Course Objectives:

1. Study of fundamentals of big data storage and processing.
2. Learning use of Hadoop for storage and processing of big data.
3. Learning different framework working with Hadoop.
4. Solve real world problems.

Brief coverage of the course

An introduction to Hadoop, Map Reduce framework, Review of Pig, a scripting language for controlling MapReduce processes on Hadoop, HBase and Hive, Overview and integration of Impala with Flume and Solr. A introduction and working of Spark will be added.

Text Books:

1. Hadoop-The Definitive Guide, Tom White, O'Reilly, 1st edition, 2009, ISBN-10: 0596521979.
2. Hadoop in Practice, Alex Holmes, Manning Publications, 2nd edition, 2014, ISBN-10: 1617292222.

(Additional reading material will be announced from time to time during class)

Reference Books:

1. Hadoop Operations, Eric Sammer, O'Reilly, 2nd edition, 2015, ISBN-10: 1491923830.
2. Professional Hadoop Solutions, Boris Lublinsky, Kevin T. Smith, and Alexey Yakubovich, O'Reilly Media, 2014, ISBN-10: 8126551070.
3. MapReduce Design Patterns-Building Effective Algorithms and Analytics for Hadoop and Other Systems, Donald Miner, Adam Shook, O'Reilly Media, 2012, ISBN-10: 1449327176.
4. HBase--The Definitive Guide, Lars George, Shroff, 1st edition, 2011, ISBN-10: 935023503X.
5. Learning Spark: Lightning-Fast Big Data Analysis, Matei Zaharia, Patrick Wendell, Andy Konwinski, Holden Karau, O'Reilly Media, 1st Edition, 2015, ISBN-10: 1449358624
6. Java - The Complete Reference, Herbert Schildt, Mcgraw Hill Education, 9th Edition, 2014, ISBN-10: 0071808558.

(Additional reading material will be announced from time to time during class)

Evaluation Scheme

Evaluation Component	Date	Weightage	Time	Mode
Mid Term Test I	TBA	15%	1 hour	Online
Mid Term Test – II	TBA	15%	1 hour	Online
Lab Assignment	Continuous Evaluation	15%	1 hour	Java programming and Hadoop books are allowed for reference in the labs.
Course Project	Continuous Evaluation	15%	-	
Attendance and Class Participation	Continuous Evaluation	10%	—	

Comprehensive Exam.	TBA	30%	2 Hours	Online
---------------------	-----	-----	---------	--------

Lecture Plan (Tentative)

Sl No	Topic	# Lectures
1	Introduction to Big Data, Hadoop and its components	1
2	MapReduce Framework: Design Patterns, MapReduce in various environments	3
3	HDFS, YARN and other components of Hadoop	1
4	Introduction to Pig, Grunt, Pig's Data Model, Pig Latin, Developing and testing scripts with Pig, Pig and other members of Hadoop community	5
5	HBase: Introduction, Client API, MapReduce Integration	4
6	Hive: Introduction, Basics, Database Processing with Hive	4
7	Impala and Flume: Introduction and Integration with Flume and Solr	3
8	Stream Computing	3
9	Spark	6

Lab Plan (Tentative)

Sl No	Topic	# Labs
1	A demo of installation of Hadoop on single node cluster and multinode cluster	2
2	Development of environment for writing and testing MapReduce programs, Writing MapReduce programs and deployment to Hadoop cluster	7
3	Installing and running Pig, Using grunt for writing scripts and writing Pig Latin scripts	8
4	HBase: Introduction, Client API, MapReduce Integration	9
5	Hive: Introduction, Basics, Database Processing with Hive	8
6	Impala and Flume: Introduction and Integration with Flume and Solr	9
7	Stream Computing	7
8	Spark: Installation of Scala and Spark, Spark Core Programming, Deployment and Spark Advanced Programming	10

Make up Policy

Students who are likely to miss a component of evaluation on a **genuine reason** may be given a make-up of that component by the Course Instructor-in-charge. The students are required to approach the Course Instructor-in-charge immediately for the same within 24 hours of the commencement of component of evaluation. The decision of the Course Instructor-in-charge in this matter shall be final.

Grading Policy

Scores obtained in all the components of evaluation shall be totaled and the final score will be converted into letter grades (A, B, C, D, E, or NC) as per NIIT University policy.

Consultation Hour.

Students are welcome to contact the course in charge for clarification of their doubts. Please take appointment on mail before.

Plagiarism Policy

Any form of plagiarism will immediately result in zero marks in that particular assignment or exam.

Attendance Policy and class participation policy

Attendance and class participation will be graded according to the NIIT University attendance policy