

COURSE CODE	COURSE TITLE	L	T	P	C
1152CS306	Map Reduce and Hadoop Tools	0	0	4	2

Course Category: Program Core

A. Preamble:

This course focuses on installation of Hadoop Framework, to develop applications based on Map Reduce model and the students will be able to understand the Hadoop Analyzing and Storage Tool (Pig and Hive).

B. Pre-requisites:

Sl. No	Course Code	Course Name
1	1151CS302	Java Programming Lab
2	1151CS303	Database Management system Lab

C. Related Courses:

Sl. No	Course Code	Course Name
1	1152CS121	Big Data and Analytics
2	1152CS210	Big data analytics and Tools

D. Course Outcomes:

Students undergoing this course are able to

CO Nos.	Course Outcomes	Knowledge Level (Based on revised Bloom's Taxonomy)
1	Learn and install Hadoop to use its API.	K3,S3
2	Understand the concepts of MapReduce and implement applications.	K3,S3
3	Be familiar with the Pig and Hive tools.	K3,S3

E. Correlation with Programme Outcomes:

H- Strong; M-Medium; L-Low

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO 1	PSO 2	PSO 3
CO1.	M	M	H	M	H				L			L			L
CO2.	M	M	H	M	H				L		M	L		H	M
CO3	M	L	H	M	H				L			L		M	L

F. Course Content

LIST OF EXPERIMENTS:

S. No Experiment Name

1. Perform setting up and Installing Hadoop in standalone mode and use web based tools to monitor the setup.
2. Perform setting up and Installing Hadoop in distributed mode and monitor the setup.
3. Implement the following file management tasks in Hadoop:
 - Adding files and directories
 - Retrieving files
 - Deleting files
4. Program to use APIs of Hadoop to interact with it.
5. Mount the one node Hadoop cluster using FUSE.
6. Implement Word Count/Frequency Programs program to understand Map Reduce Paradigm.
7. Implement Hadoop Kmer counting.
8. Write Map Reduce program that mines weather data.
9. Implement Matrix Multiplication with Hadoop Map Reduce.
- 10&11. Install, Run Pig and write Pig Latin scripts to sort, group, join, project, and filter data.
- 12&13. Install, Run Hive and use Hive to create, alter, and drop databases, tables, views, functions, and indexes.

Model Examination

G. Learning Resources

i. References

- Tom White “Hadoop: The Definitive Guide” Third Edition, O’reilly Media, 2012.
- Dirk deRoos, Paul C. Zikopoulos, “Hadoop For Dummies”, A Wiley Brand, 2014
- Edward Capriolo, Dean Wampler, and Jason Rutherglen, “Programming Hive”, O’reilly Media, 2012.
- Michael Minelli, Michelle Chambers, and Ambiga Dhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013.
- Alan Gates and Daniel Dai, "Programming Pig –Dataflow scripting with Hadoop", O’Reilley, 2ndEdition, 2016.

iii. Online Resources

- <http://hadoop.apache.org/common/docs/r0.20.2/api/>
- <http://hbase.apache.org/docs/current/api/index.html>
- <https://hadoop.apache.org/>
- <http://pig.apache.org/docs/r0.7.0/tutorial.html>
- <https://cwiki.apache.org/confluence/display/Hive/Home>
- <http://schatz-lab.org/teaching/exercises/hadoop/>