



Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous Institute Affiliated to University of Mumbai)

Course Code	Course Name	Teaching Scheme (Hrs/week)			Credits Assigned				
		L	T	P	L	T	P	Total	
CEL912	PG Laboratory –II (Big Data Analytics and Managements Laboratory)	--	--	2	--	--	1	1	
		Examination Scheme							Total
		ISE		ESE			20		
		40		Practical		Oral		60	

Pre-requisite Course Codes	Data Structures, Analysis of Algorithms, CE912(Big Data Analytics and Managements)
At the end of successful completion of the course, students will be able to	
Course Outcomes	CO1 Describe big data and use cases from selected business domains
	CO2 Install, configure, and run Hadoop and HDFS
	CO3 Perform map-reduce analytics using Hadoop
	CO4 Clarify NoSQL big data management
	CO5 Use Hadoop related tools such as HBase, Cassandra, Pig, and Hive for big data Analytics

Exp. No.	Experiment Details	Ref.	Marks
1	HDFS: Start by reviewing HDFS. You will find that its composition is similar to your local Linux file system. You will use the hadoop fs command when interacting with HDFS. 1. Review the commands available for the Hadoop Distributed File System 2. Copy file foo.txt from local disk to the user's directory in HDFS 3. Get a directory listing of the user's home directory in HDFS 4. Get a directory listing of the HDFS root directory 5. Display the contents of the HDFS file user/fred/bar.txt 6. Move that file to the local disk, named as baz.txt 7. Create a directory called input under the user's home directory 8. Delete the directory input old and all its contents 9. Verify the copy by listing the directory contents in HDFS	1,2	10
2	MapReduce 1. Create a JOB and submit to cluster 2. Track the job information 3. Terminate the job 4. Counters in MR Jobs with example 5. Map only Jobs and generic map examples 6. Distributed cache example 7. Combiners, Secondary sorting and Job chain examples	1,2,3	10



Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous Institute Affiliated to University of Mumbai)

3	MapReduce (Programs) Using movie lens data 1. List all the movies and the number of ratings 2. List all the users and the number of ratings they have done for a movie 3. List all the Movie IDs which have been rated (Movie Id with at least one user rating it) 4. List all the Users who have rated the movies (Users who have rated at least one movie) 5. List of all the User with the max, min, average ratings they have given against any movie 6. List all the Movies with the max, min, average ratings given by any user	1,2,3	10
4	Extract facts using Hive OR Extract sessions using Pig Hive allows for the manipulation of data in HDFS using a variant of SQL. This makes it excellent for transforming and consolidating data for load into a relational database. In this exercise you will use HiveQL to filter and aggregate click data to build facts about user's movie preferences. The query results will be saved in a staging table used to populate the Oracle Database.	4,3	10
Total Marks			40

References:

- [1] Tom White, "Hadoop: The Definitive Guide", Third Edition, O' Reilly, 2012.
- [2] Eric Sammer, "Hadoop Operations", O'Reilly, 2012.
- [3] Vignesh Prajapati, Big data analytics with R and Hadoop, SPD 2013.
- [4] E. Capriolo, D. Wampler, and J. Rutherglen, "Programming Hive", O'Reilly, 2012.
- [5] Lars George, "HBase: The Definitive Guide", O'Reilly, 2011.
- [6] Alan Gates, "Programming Pig", O'Reilly, 2011.