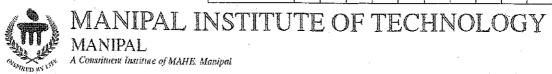
	1 1			
Reg. No.				
	1 1	1 1	 1 1	1 }



II SEMESTER M.TECH. END SEMESTER EXAMINATIONS, APRIL 2018 SUBJECT: OPEN ELECTIVE - BIGDATA ANALYTICS AND TECHNOLOGIES [ICT 5281] REVISED CREDIT SYSTEM (30/04/2018)

Time: 3 Hours

MAX. MARKS: 50

	Instructions to Candidates:]
	* Answer ALL the questions.	
	* Missing data if any, may be suitably assumed.	
1A.	Write a program for finding maximum temperature in a year using map-reduce technique. Also explain the logical data flow for the same.	5
1B.	What is sharding and replication in NOSQL data bases? Explain.	3
1C.	Compare GPFS-FPO and HDFS.	2
2A.	State and prove Brewers CAP theorem.	5
2B.	Distinguish between clustering and classification. What are the major approaches for clustering? Explain each approach in brief.	3
2C.	TABLE Q.2C below, shows the data on sales in a bookstore. The data has two fields	2
	namely year and sales in year. Using linear regression model, predict the sales in year 2020.	
	TABLE. Q.2C	
	Year Sales	
	2010 1200	

Year	Sales
2012	1200
2013	1100
2014	1300
2015	1250
2016	1150
-	

3A.	Write mongoDB commands for the following.			
	i. To create a Data base called TEST.			
	ii. To create a collection called EMPLOYEE			
	iii. To insert 5 documents (employee id, employee name, department, salary) into			
	the EMPLOYEE collection.			
	iv. To retrieve documents with salary greater than 50000 and employee id >10.	•		
	v. To Update schema by adding new attribute Hobbies for employee id =101.	•		
3B.	Explain Pig compilation architecture with a neat diagram. Also write a program in	3		
	Pig for finding wordcount.			
3℃.	Write an AQL module to create a view to fetch all email ids in a document.	2		

ICT 5281

4A.	Build a classification model using Perceptron learning algorithm for following data.						5
	Al	1	-1	1	2	B Galler	
	X2	2	1	-1	-2	in the first	
	Target class	0	0	1	1	- 100	
	initial weights:	w1=1, w2=0	.5. bias=0.	on: Sgn(s)=1 if s>(•		
4B.	List and explain the functions of namenode, secondary namenode and data node in HDFS.						
4C.	In a factory, machines A, B and C manufacture 15%, 25% and 60% of the total production of bolts respectively. The bolts manufactured by the machine A have 4% rate of defects, machine B have 2% rate of defects and machine C have 3% rate of defects. A bolt is drawn at random and is found to be defective. What is the probability that it was produced by B?					2	
5A.	oranch and mark	. Write a Stre ch branch an	eam Processing d maximum m	ng of the attribute g application which ark obtained by a	h displays total :	number	5
5B. 5C.		policy? Expla	ain various evi	ction policies used ble example	l in tumbling wi	ndow.	3 2