

--	--	--	--	--	--	--	--	--	--

Second Semester M.Tech. Degree Examination, June/July 2015
Managing Big Data

Time: 3 hrs.

Max. Marks: 100

Note: Answer any FIVE full questions.

- 1 a. What are the three types of big data that are a big deal for market? Discuss briefly each of them. (10 Marks)
- b. Analyse the operation of Distributed file system and RDBMS. (10 Marks)
- 2 a. What is credit analytics? How does it help in the real world? (08 Marks)
- b. How is it possible to detect brand with Big Data in social network analysis? (06 Marks)
- c. Consider a computer system which has four I/O channels and each channel can record data at 16MB/Sec. What is the time taken to read 2TB of data? What is the number of system required to read all the data if the time given is 5Sec. (06 Marks)
- 3 a. How is NOSQL polyglot persistent? (03 Marks)
- b. How is crowd sourcing a great way to capitalize on the resources that can build algorithms and predictive models? (05 Marks)
- c. What is data aggregation? What are different data models used in NOSQL? Give example and explain. (12 Marks)
- 4 a. What is the main process of operation is master slave replication? What are the pros and cons of maser slave replication. (10 Marks)
- b. What does CAP theory convey? Give an example to support CAP theorem. (10 Marks)
- 5 a. With a diagram explain the sequence of events that takes place when reading a file between the client and HDFS. (10 Marks)
- b. Write two racks with a at least 5 nodes in each rack. Show two data counter and process of pipeline replication placement when replication factor is 3. Show this distance for
 - i) Distance $(d^1/r^1/n^1, /d^1/r^1/n^1)$
 - ii) Distance $(/d^1/r^1/n^1, /d^1/r^1/n^2)$
 - iii) Distance $(/d^1/r^1/n^1, /d^1/r^2/n^3)$
 - iv) Distance $(/d^1/r^1/n^1, /d^2/r^3/n^4)$
 When n^1, n^2, n^3, n^4 nodes $r^1, r^2,$ and r^3 are racks and d^1, d^2 are data centre. (10 Marks)
- 6 a. What is the function of a combiner in Map reduce? How does it differ from Reduce function? (04 Marks)
- b. What is the procedure to recover from a failed name node? (05 Marks)
- c. What is HDFS Federation? Explain with a diagram how Namenode , Namespace, Block pools and common storage configuration system work. (11 Marks)
- 7 a. List the different failures that need to be considered for MapReduce programs running on YARN. What is the effect of each failure. (10 Marks)
- b. What are different job schedulers available in later version of Hadoop? Explain the merit and demerit of each. (10 Marks)
- 8 a. What is the difference between HBASE and HDFS in Hadoop? (05 Marks)
- b. How is Cassandra considered as distributed decentralized, elastic, fault tolerant and column-oriented? (10 Marks)
- c. What are the main features of Big Latin script. (05 Marks)