


REG. NO										
---------	--	--	--	--	--	--	--	--	--	--

	<p style="text-align: center;"><b>MANIPAL INSTITUTE OF TECHNOLOGY</b>          (A constituent Institute of MANIPAL ACADEMY OF HIGHER EDUCATION)  <b>MANIPAL - 576 104, KARNATAKA, INDIA</b></p>
---	---

VII SEMESTER B.TECH. (COMPUTER SCIENCE AND ENGINEERING)  
 END SEMESTER MAKEUP EXAM - JAN 2018  
 SUBJECT: BIG DATA ANALYTICS (CSE 4006)  
 DATE: 02-JAN-2018

TIME: 3 HOURS

MAX.MARKS: 50

**Instructions to Candidates**

- **Note:** Answer all full questions.

- 1A. Which ‘V’ in big data is associated with the speed of data that is placed in and taken out from a database? Explain the predictive and prescriptive analysis along with an example? **4M**
- 1B. List five different dimensions of big data. Explain the concept of binary data in *MongoDB* and discuss how the scalability is achieved. **5M**
- 1C. What is the output of the following syntax? **1M**  
`db.Students.find().skip(db.Students.count() – 2).pretty();`
- 2A. Discuss four common characteristics of *NoSQL* **4M**
- 2B. List the features of *MongoDB* and explain the feature associated with horizontal scaling. What is the technical term used to represent the primary key of *RDBMS* in *NoSQL*? Write its structure and discuss how the value is assigned to this field. **5M**
- 2C. Consider a collection with `_id`, `StudName` and `Grade`. Display only the `StudName` and `Grade` from all the documents of the `Students` collection. The identifier `_id` should be suppressed and NOT displayed. Write the syntax in *MongoDB*. **1M**
- 3A. Write the *Key/Value* pair declaration for the requirement “Store the list of subjects for each student of each semester”. **2M**
- 3B. Which data type is associated with the expiration of a variable? What are the different layers of *Hadoop* Ecosystem? Which feature of *Cassandra* can be compared to the Daemon associated with execution of the task in *Hadoop*? How many resource managers a node manager runs in a cluster having 4 data nodes? Is this statement true? Justify your answer. **4M**

- 3C.** Along with a diagram explain the feature of Cassandra associated with optimizing the Cluster consistency **4M**
- 4A.** Write the block diagram along with all the necessary entities involved when a client submits an application in *YARN* framework. **4M**
- 4B.** What are synchronization markers and how they are used in serialization format? Explain briefly. **2M**
- 4C.** Along with a neat diagram explain different phases of *MapReduce* for processing the following text and write the output of each phase (finally, the words and number of their instances has to be calculated) **4M**
- Input data:**  
Hadoop Definitive guide,  
Hadoop in action,  
Map Reduce Design patterns
- 5A.** Write the block diagram of data analytics project life cycle in *RHadoop* **2M**
- 5B.** Explain the case study of stock change data analytics in share market. **5M**
- 5C.** Explain the *RHadoop* architecture with a diagram **3M**