

Reg. No.									
----------	--	--	--	--	--	--	--	--	--



**MANIPAL INSTITUTE OF TECHNOLOGY**  
**MANIPAL**  
*(A constituent unit of MAHE, Manipal)*

**VI SEMESTER B.TECH. (COMPUTER SCIENCE AND ENGINEERING)**  
**GRADE IMPROVMENT EXAMINATIONS, AUG 2021**  
**SUBJECT: Big Data Modelling and Management [CRA 4055]**  
**REVISED CREDIT SYSTEM**  
**(10/08/2021)**

Time: 2 Hours

MAX. MARKS: 40

**Instructions to Candidates:**

- ❖ Answer **ANY FOUR FULL** questions.
- ❖ Missing data may be suitably assumed.

- |            |  |           |
|------------|--|-----------|
| <b>1A.</b> | List and explain in brief the main characteristics of Big Data.  | <b>5M</b> |
| <b>1B.</b> | Explain in brief the benefits of Big Data processing with respect to real-world scenario.  | <b>5M</b> |
| <b>2A.</b> | Explain Hadoop Ecosystem in brief with a neat block diagram.   | <b>6M</b> |
| <b>2B.</b> | Explain in brief the requirements of programming models for Big Data   | <b>4M</b> |
| <b>3A.</b> | Explain 5P's of Data Science.  | <b>5M</b> |
| <b>3B.</b> | Explain how distributed file system works and mention its advantages.  | <b>5M</b> |
| <b>4A.</b> | Define serialized representation of data. With the help of sample data, explain the difference between serialized representation of data and data model.   | <b>5M</b> |
| <b>4B.</b> | Explain data streaming. List down the required characteristic of data streaming systems and challenges associated with it. Elaborate on why real time dynamic streaming data processing is an important part of today's Big Data applications. | <b>5M</b> |
| <b>5A.</b> | Explain how volume and variety are addressed in Big Data management systems in the following cases:<br>i. Usage of database system that can perform parallel operations<br>ii. When to use Hadoop or Yarn style system?                        | <b>6M</b> |
| <b>5B.</b> | What are the problems associated with file system. Explain the advantages of using a DBMS over a file system.  | <b>4M</b> |
| <b>6A.</b> | Explain how data lakes enable batch processing of streaming data. How data are stored in data lakes?   | <b>5M</b> |
| <b>6B.</b> | Write short note on (i) Vertica system (ii) Distributed R.   | <b>5M</b> |

\*\*\*\*\*