

**End Term Assessment – January 2022**

**Semester – V**

**B.Tech. CSE (Devops) – 2019**

**Subject Code: CDV3021**

**Duration: 3 hours (including time for uploading)**

**Subject Name: Big Data Overview**

**Max. Marks: 50**

**Instructions**

- **Write name and registration number, page number, on all the pages, convert into one PDF, tag it with your registration number\_Name\_subjectcode\_subject title**
- **The Assessment consists of 2 sections**
  - **Part A contains 10 questions of 2 marks each and all questions are compulsory.**
  - **Part B consists of 4 questions of 10 marks each, out of which 3 questions to be attempted.**
- **Hand written responses to be submitted/uploaded as scanned pages of answer sheets (max. 5 pages) within the mentioned duration. 6<sup>th</sup> page and onwards won't be evaluated**

**PART – A**

***2 \* 10 = 20 Marks (Each answer- Word Limit- 50 Words)***

1. What are the limitations of a structured data? (2)
2. What are relational databases? Give examples also. (1+1)
3. What do you understand by Big data processing? (2)
4. Write any 2 advantage associated with horizontal scaling. (2)
5. Write any 2 reason for choosing data lake over data warehouse with the help of an example. (2)
6. Give any 4 examples of unstructured data. (4\*0.5=2)
7. What is the purpose of using Big data analytics? (2)
8. What is the function of a Job tracker in Hadoop? (2)
9. Is Load balancing required in cloud environment? Give reason also. (1+1)
10. List the components of a data lake. (2)

**PART – B**

***10 \* 3 = 30 Mark (Each answer- Word limit- 250 words)***

11. What are the problems associated with data explosion? Enumerate 4 sources of data generation with atleast one example? (6+4)
12. What do you understand by the term Big Data? Elaborate different dimensions of big data. (2+8)
13. What is the need of horizontal scaling? Is horizontal scaling good for big data? (6+4)
14. For what use cases data lake is suitable? Explain the components of a data lake in brief. (3+7)