**POORNIMA UNIVERSITY, JAIPUR**

**END SEMESTER EXAMINATION, APRIL 2023**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **2BC4116** | Roll No. | Total Printed Pages: 2 |
| **2BC4116** |  |
| BCA II Year IV Semester (Main/Back) End Semester Examination, April 2023  **(All Spl.)** | |
| **BCACCA4101 : Big Data Analysis** | | | |

# Time: **3** Hours. Total Marks: **60**

Min. Passing Marks: **21**

Attempt **five** questions selecting one question from each Unit. There is internal choice from Unit I to Unit V. Marks of each question or its parts are indicated against each question / parts. Draw neat sketches wherever necessary to illustrate the answer. Assume missing data suitably (if any) and clearly indicate the same in the answer.

Use of following supporting material is permitted during examination for this subject.

# **1.--------------------Calculator-----------------** **2.------------------Nil-----------------------**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **UNIT-I (CO1)** | **Marks** | **Bloom Level** |
| **Q.1** | **(a)** | Define Big Data and Explain the characteristics of Big Data with features of it? | **(6)** |  |
|  |  |  |  |  |
|  | **(b)** | Apply different 4 configuration files in Hadoop defined with the use of these? Summarize the important features of Hadoop? | **(6)** |  |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.2** | **(a)** | List out the Hadoop ecosystem components? Discuss the Hadoop input output streaming in detail. | **(6)** |  |
|  |  |  |  |  |
|  | **(b)** | List out the sources of big data? How can the Big Data affect our society? | **(6)** |  |
|  |  |  |  |  |
|  |  | **UNIT-II (CO2)** |  |  |
|  |  |  |  |  |
| **Q.3** | **(a)** | Explain in detail about HDFS Architecture. Explain the file management HDFS commands and describe with suitable example? | **(6)** |  |
|  |  |  |  |  |
|  | **(b)** | Discuss the import and export methods by Apache sqoop with neat diagram. Explain the Data Ingestion by Flume with neat diagram? | **(6)** |  |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.4** | **(a)** | How can big data analytics benefit business? What are some of the challenges that come with a big data project? | **(6)** |  |
|  |  |  |  |  |
|  | **(b)** | What is rack awareness in Hadoop clusters, Hadoop protect data against unauthorized access? | **(6)** |  |
|  |  |  |  |  |
|  |  | **UNIT-III (CO3)** |  |  |
|  |  |  |  |  |
| **Q.5** | **(a)** | How one can decide the number of map tasks? List all the steps of mapper job execution. | **(6)** |  |
|  |  |  |  |  |
|  | **(b)** | Explain the Stages of MapReduce? Discuss the Word Count problem with Map Reduce programming? | **(6)** |  |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.6** | **(a)** | What is HDFS and what are its main components? Define and describe the term FSCK? | **(6)** |  |
|  |  |  |  |  |
|  | **(b)** | What are the key steps in deploying a big data platform? Why is Hadoop so popular in big data analytics? | **(6)** |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  | **UNIT-IV (CO4)** |  |  |
|  |  |  |  |  |
| **Q.7** |  | Consider the table below. And find the answers of following queries using Hive.   * List all student list in Descending order. * List course wise total scholarship. * Show the name of the student who is having maximum scholarship. * List all the students who is having grade A or A+ or A++.  |  |  |  |  |  | | --- | --- | --- | --- | --- | | Student\_id | Student\_name | Course | Scholarship | Grade | | 1 | Ashwin | BCA | 20% | A | | 2 | Sneha | BCA | 10% | B | | 3 | Lavanya | BBA | 10% | B | | 4 | Karthik | BBA | 30% | A+ | | 5 | Shafeek | Btech | 40% | A++ | | 6 | James | Btech | 20% | A | | **(12)** |  |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.8** | **(a)** | How pig is different from SQL? List all the features of pig. | **(6)** |  |
|  |  |  |  |  |
|  | **(b)** | Hbase is a column-oriented database. Explain how data is stored in Hbase with example table. | **(6)** |  |
|  |  |  |  |  |
|  |  | **UNIT V (CO5)** |  |  |
|  |  |  |  |  |
| **Q.9** | **(a)** | What is Machine Learning? What are the different types of machine learning algorithms? Explain each. | **(6)** |  |
|  |  |  |  |  |
|  | **(b)** | Collaborative filtering is a very good technique for Recommendation System. How collaborative filtering can do the recommendation effectively explain with example. | **(6)** |  |
|  |  |  |  |  |
|  |  | **OR** |  |  |
|  |  |  |  |  |
| **Q.10** |  | A movie rating dataset is given below where first user have missed the last movie rating. Analyse the data and find what rating first user can give? Collaborative Filtering of Recommendation system.   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Movies  User | M1 | M2 | M3 | M4 | M5 | | Vivek | 5 | 3 | 4 | 4 | ?? | | Abhishek | 3 | 1 | 2 | 3 | 3 | | Uday | 1 | 5 | 5 | 2 | 1 | | **(12)** |  |