| **Examination** | **: \_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **Seat No** | **: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| --- | --- | --- | --- |
| **Date** | **: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **Day** | **: \_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **Time** | **: \_\_\_\_\_\_\_\_ to \_\_\_\_\_\_\_\_\_** | **Max. Marks** | **: 60** |

| **INSTRUCTIONS:** | | | | |
| --- | --- | --- | --- | --- |
| 1. | Answer each section in a separate answer book. | | | |
| 2. | Figures to the right indicate maximum marks for that question. | | | |
| 3. | The symbols used carry their usual meanings. | | | |
| 4. | Assume suitable data, if required & mention them clearly. | | | |
| 5. | Draw neat sketches wherever necessary. | | | |
| **SECTION – I** | | | | |
| **Q.1** | **Do as directed.** | | **[10]** | |
|  |  | Fill in the blanks:  (i) Hadoop Distributed File System (HDFS) has master \_\_\_\_ and worker \_\_\_\_ named java processes.  (ii) YARN has master \_\_\_\_\_ and worker \_\_\_\_\_ named java processes. | [2] | |
|  | (b) | State true/false:  (i) Data science involves three disciplines: mathematics, business and technology.  (ii) hdfs-site.xml and core-site.xml are configuration files of Hadoop. | [2] | |
|  | (c) | Perform classification of analytics. | [2] | |
|  | (d) | Explain CAP theorem. | [2] | |
|  | (e) | Write a short note on SQL, NoSQL and NewSQL. | [2] | |
| **Q.2** | **Attempt the following questions.** | | **[10]** | |
|  |  | What is machine learning? How does Big Data Analytics relate to machine learning? | [5] | |
|  |  | Describe types of data: Structured, semi and unstructured, human generated, machine generated. | [5] | |
| **Q.3** | **Attempt the following questions:** | | **[10]** | |
|  | (a) | Provide mongoDB commands to achieve below requirements:  (I) Display all databases, start using a database named ‘ecommerce’, create a  collection named ‘survey’, display all collections created thus far in the chosen  database.  (II) Create documents in ‘survey’ having field and value as mentioned and verify:  doc#1 -> ‘name’:’James’ ‘age’:27 ‘education’:’BSc’ ‘scholarship’:’yes’  doc#2 -> ‘name’:’Harry’ ‘age’:25 ‘salary’: 25000  doc#3 -> ‘name’:’William’ ‘age’:25 ‘salary’: 15000  doc#4 -> ‘name’:’Michel’ ‘age’:20 ‘gender’:’female’ hobbies:’hockey’,  ‘painting’, ‘singing’  (III) Display all the documents having salary of ‘age’ <= 25  (IV) Display sorted all the documents in ascending order of age followed by  descending order of salary nicely.  (V) Display average salary as AvgSalary group by gender. | [5] | |
|  | (b) | Provide steps to setup hadoop cluster. | [5] | |
|  | | | | |
| **SECTION – II** | | | | |
| **Q.4** | **Do as directed.** | | **[10]** | |
|  | (a) | Fill in the blanks:  (I) Horizontal scaling refers to \_\_\_\_\_\_\_\_\_\_.  (II) Vertical scaling refers to \_\_\_\_\_\_\_\_\_. | [2] | |
|  | (b) | State true or false.  (I) Apriori algorithm is used to perform market basket analysis.  (II) Mean, mode and median are used for measuring central tendency of data. | [2] | |
|  | (c) | Give any example of concept hierarchy. | [2] | |
|  | (d) | Why mongoDB? Provide features of the same. | [2] | |
|  | (e) | What does it mean by noisy data? | [2] | |
| **Q.5** | **Attempt the following questions.** | | **[10]** | |
|  | (a) | What is Big Data Analytics? Describe Big Data characteristics using famous Vs. | [5] | |
|  | (b) | Describe the phases of Big Data Analytics with respect to knowledge discovery. | [5] | |
| **Q.6** | **Attempt the following questions.** | | **[10]** | |
|  | (a) | Describe Hadoop EcoSystem components and their purpose. | [5] | |
|  | (b) | What is map-reduce? Show working of map reduce using word count example. | [5] | |