DHARMSINH DESAI UNIVERSITY, NADIAD

FACULTY OF TECHNOLOGY

B.TECH. SEMESTER VII [CE]

SUBJECT: (CE720) BIG DATA ANALYTICS

[

|  |  |  |  |
| --- | --- | --- | --- |
| **Examination** | **: First Sessional** | **Seat No** | **: \_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **Date** | **: 31/08/2019** | **Day** | **: Saturday** |
| **Time** | **: 1:30 p.m. to 2:45 p.m.** | **Max. Marks** | **: 36** |

|  |  |
| --- | --- |
| **INSTRUCTIONS:** | |
| 1. | Figures to the right indicate maximum marks for that question. |
| 2. | The symbols used carry their usual meanings. |
| 3. | Assume suitable data, if required & mention them clearly. |
| 4. | Draw neat sketches wherever necessary. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Q.1** | **Do as directed.** | | |
| (a) | List advantages of HDFS. | **[2]** |
| (b) | State true or false and justify your answer: Hadoop is fault tolerant. | **[2]** |
| (c) | Differentiate InputSplit and block in mapreduce with example. | **[2]** |
| (d) | Justify how hadoop attains master-slave architecture. | **[2]** |
| (e) | Why is NoSQL so popular nowadays? | **[2]** |
| (f) | Write class prototype and purpose of any two from Package “org.apache.hadoop.mapreduce” | **[2]** |
| **Q.2** | Attempt ***Any Two*** from the following questions. | |  |
| (a) | Explain YARN architecture with figure and also explain application submission in YARN. | **[6]** |
| (b) | How input file is converted to key-value pairs which are given as an input to map phase in mapreduce? Explain the entire process of conversion. | **[6]** |
| (c) | Explain functions of secondary namenode and checkpointing in detail. | **[6]** |
| **Q.3** | (a) | i. What output ‘jps’ command will show after successful execution of start-dfs.sh script of hadoop sbin? What are the purposes of each?  ii. Match the following:   |  |  | | --- | --- | | ***Column A*** | ***Column B*** | | NLP | Document Data Store | | MongoDB | Machine generated structure data | | Point-of-sale | Noisy unstructured data | | HBase | Comprehend human or natural language input | | Chats | UIMA | | IBM | Column-oriented Data Store | | **[3]**  **[3]** |
|  | (b) | Discuss any use case of Hadoop. i.e. Click Stream Data. | **[6]** |
| OR | | | |
| **Q.3** | (a) | i. Write definition of BigData as described by Gartner.  ii. Match the following:   |  |  | | --- | --- | | ***Column A*** | ***Column B*** | | PostgreSql | Key-Value Data Store | | Membase | ACID | | Relations/Tables | Human generated unstructured data | | Neo4j | Relational data model with scalable performance | | Social Media data | Graph Data Store | | NewSQL | Open source relational database | | **[3]**  **[3]** |
| (b) | Draw a diagram of Hadoop Eco System. Is mongoDB part of Hadoop Ecosystem? | **[6]** |