DHARMSINH DESAI UNIVERSITY, NADIAD

FACULTY OF TECHNOLOGY

B.TECH. SEMESTER VII [CE]

SUBJECT: (CE720) BIG DATA ANALYTICS

[

|  |  |  |  |
| --- | --- | --- | --- |
| **Examination** | **: First Sessional** | **Seat No** | **: \_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **Date** | **: 02/08/2019** | **Day** | **: Friday** |
| **Time** | **: 1:45 TO 3:00 PM** | **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) | i. How many data blocks will be created for a file that is 950 MB? The default block size if 64 MB and the replication factor is 3.  ii. Which file is used for updating HDFS settings? | | **[2]** |
| (b) | List at least four difficulties faced by MapReduce version 1. | | **[2]** |
| (c) | What is contained in editlog and fsImage files? | | **[2]** |
| (d) | Compare and contrast “Client-server architecture” to “Master-slave architecture”. | | **[2]** |
| (e) | Describe four categories of NoSQL as per industrial usage. | | **[2]** |
| (f) | Provide method signature and purpose of org.apache.hadoop.mapreduce.Mapper class. (any two) | | **[2]** |
|  | | | | |
| **Q.2** | Attempt ***Any Two*** from the following questions. | | |  |
| (a) | How output file is generated from input data by Hadoop MapReduce? Explain the entire process with figure and example. | | **[6]** |
| (b) | Explain HDFS write mechanism with figure. | | **[6]** |
| (c) | Compare Hadoop MapReduce version 1.0 and 2.0. Explain various components of YARN. | | **[6]** |
|  | | | | |
| **Q.3** | (a) | i. Provide names and functionalities of HDFS daemons.  ii. How to compile and run hadoop mapreduce job? Also, show ways to monitor the same. | | **[3]**  **[3]** |
|  | (b) | Provide overview of Hadoop Eco System. | | **[6]** |
| **OR** | | | | |
| **Q.3** | (a) | i. Describe HDFS with example command usage to deal with native and distributed file system.  ii. What is user equivalence? How does it help establishing hadoop cluster? Provide steps. | | **[3]**  **[3]** |
| (b) | Perform comparison of SQL, NoSQL and NewSQL. | | **[6]** |
|  | | |  | |