| U.S.N. | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|

# B.M.S. College of Engineering, Bengaluru-560019

**Autonomous Institute Affiliated to VTU**

## July 2023 Semester End Main Examinations

**Programme: B.E.** **Semester: VI**
**Branch: Information Science and Engineering** **Duration: 3 hrs.**
**Course Code: 20IS6PEBDA** **Max Marks: 100**
**Course: Big Data Analytics** **Date: 17.07.2023**

**Instructions**: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

**Important Note:** Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

| | | UNIT - I | CO | PO | Marks |
|---|---|---|---|---|---|
| 1 | a) | How do the characteristics of data, such as volume, velocity, variety, and veracity impact data management and analysis in the era of big data? | *CO2* | *PO1* | **06** |
| | b) | Explain the high level architecture of Hadoop. | *CO1* | | **08** |
| | c) | Identify the type of data in the following: Table, MS Access, videos Images, Social media, excel sheet. | *CO3* | *PO2* | **06** |
| | | **UNIT - II** | | | |
| 2 | a) | What is the significance of DFS? How does client perform the read and write data in HDFS, explain the process with a suitable block diagram. | *CO1* | | **06** |
| | b) | Considering the HDFS Architecture, answer the following questions<br>i. How are issues w.r.t Availability and Fault tolerance addressed in HDFS.<br>ii. How a secondary name node differs from the name node in HDFS<br>iii. Specify the role of job tracker and task tracker in HDFS. | *CO3* | *PO2* | **08** |
| | c) | Write a map reduce program to count number of words in a three text files. | *CO2* | *PO1* | **06** |
| | | **OR** | | | |
| 3 | a) | What is the significance of block size in HDFS? How does the block size impact the performance and storage efficiency of HDFS? | *CO2* | *PO1* | **06** |
| | b) | Describe the role of NameNode and DataNode in HDFS architecture. How do they work together to ensure fault tolerance and high availability? | *CO2* | *PO1* | **08** |
| | c) | Describe the process of writing data to HDFS. What happens when a client writes data to HDFS, and how is the data split into blocks and distributed across the cluster? | *CO2* | *PO1* | **06** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | **UNIT - III** | | | |
| | 4 | a) | Using CQL create a data base for college. Create a table Bangalore College(ID, College Name, Branch, Area)<br>i. Insert required details to the table.<br>ii. Display College name and College ID whose area = Basavanagudi<br>iii. Add a new filed NRIF Rank with value 50 for college ID=3 | *CO2* | *PO1* | **08** |
| | | b) | Create a hive table for faculty with appropriate attributes and store it in text file, sequence file, RC file, AVRO, ORC, PARQUET file formats. | *CO2* | *PO1* | **06** |
| | | c) | Write a simple UDF to find minimum element among three elements. | *CO2* | *PO1* | **06** |
| | | | **OR** | | | |
| | 5 | a) | Apply query options available in Apache Cassandra for data retrieval and manipulation on a bank table. Perform CURD operations on the bank table. | *CO2* | *PO1* | **08** |
| | | b) | Discuss the capabilities and commands used for importing and exporting data to and from Cassandra using CQLSH. | *CO1* | | **04** |
| | | c) | What is a SerDe in Hive, and what is its role in data serialization and deserialization? Explain how SerDe facilitates the processing of different data formats in Hive. | *CO1* | | **08** |
| | | | **UNIT - IV** | | | |
| | 6 | a) | Consider RDD1 = {1,2,3,4,5}<br>RDD2={3,4,5,6,7,8}<br>Perform union intersection, map on RDD1 (x=x+3), filter on even numbers on RDD2. | *CO2* | *PO1* | **08** |
| | | b) | What are the key components and stages involved in the execution of a Spark job? Please explain the anatomy of a Spark job run. | *CO1* | | **06** |
| | | c) | What are the options available in Sqoop for exporting data from Hadoop to relational databases? Discuss the process of exporting data using Sqoop. | *CO1* | | **06** |
| | | | **UNIT - V** | | | |
| | 7 | a) | Explain reads and writes on Zoo Keeper. | *CO1* | | **08** |
| | | b) | What are the core components and architecture of Apache Flume's data flow model? Explain how data is ingested, transformed, and delivered to different destinations in the Flume ecosystem. | *CO1* | | **08** |
| | | c) | What is Apache ZooKeeper, and what role does it play in Big Data systems? | *CO1* | | **04** |

**********