NIRMA UNIVERSITY

Institute of Technology

Semester End Examination (IR), December 2023
BTech in Computer Science and Engineering, Semester VII
2CS702 – Big Data Analytics

Roll / Exam No. [   ]

Supervisor's Initial with Date [   ]

Time: 3 Hours

Max Marks: 100

Instructions:
1. Attempt all questions
2. Figures to the right indicate full marks
3. Assume necessary data.
4. Use a section-wise separate answer book.
5. Draw neat sketches wherever necessary.

SECTION-I

Q:1 Answer the following questions. (4 X 4)   **[16]**
CO1, BL1,2

1. Define data science. Explain the role of big data analytics in data science applications. **[4]**
2. Explain the role of the ETL process in the big data pipeline. **[4]**
3. Explain the role of Ambari architecture in the Hadoop ecosystem. **[4]**
4. Is IoT data considered big data? Justify your answer with suitable examples. **[4]**

Q:2 Answer the following questions (8 X 2)   **[16]**
CO2, BL4

1. Explain the advantages of conventional database management systems in comparison with HDFS architecture. **[8]**
2. How metadata is managed in HDFS architecture. Explain in detail each component involved in metadata management with a proper diagram. **[8]**

OR

2. Discuss the CAP theorem. Give an example where you can apply the CAP theorem with proper justification. **[8]**

Q:3 Answer the following questions.   **[18]**
CO1, BL3

1. Consider the following dataset in text format, **[10]**
Sales_order (OrderNo, ClientNo, OrderDate, SalesmanNo, DeliveryType, Billyn, DeliveryDate, OrderStatus)

   1. Write pseudo code for mapper tasks and reducer tasks to find the number of Cancelled orders.
   2. Illustrate the execution of the Map and Reduce phase with a proper diagram for the following input data.

O19001	C01	12-01-16	S01	F	N	20-01-16	In Process
O19002	C02	25-01-17	S02	P	N	27-01-17	Cancelled
O46865	C03	18-02-17	S03	F	Y	20-02-17	Fulfilled
O19003	C01	03-04-16	S01	F	Y	07-04-16	Fulfilled
O46866	C04	20-05-16	S02	P	N	22-05-16	Cancelled
O19008	C05	24-05-16	S04	F	N	26-05-16	In Process

2   Explain in detail the role of the record reader and record writer in the **[8]**
    MapReduce programming model.

SECTION-II

Q:4   **Answer the following questions [6 X 3]**   **[18]**
CO3, BL3

   1. Explain schema-on-read and schema-on-write with suitable example.   **[6]**
   2. 'MapReduce architecture is more suitable for iterative query processing   **[6]**
      than RDBMS architecture'. State whether the statement is correct or
      not with proper justification.
   3. Discuss the limitations of Hadoop V1.0. Explain the technological   **[6]**
      solutions to overcome the same.

Q:5   **Answer the following questions [8 X 2]**   **[16]**

1   Explain the significance of PIG architecture in the Hadoop ecosystem.   **[8]**
CO3, BL3   Discuss the features of PIG architecture.

OR

1   Explain RDD in spark architecture. Discuss the properties of RDD.   **[8]**
CO3, BL3

2   Explain collection and document in MongoDB. Does MongoDB support join   **[8]**
CO3, BL5   operations? Justify your answer.

OR

2   Discuss how CRUD operations are performed using Cassandra with   **[8]**
CO3, BL5   suitable examples.

Q:6   **Answer the following questions.**   **[16]**

1   Kiyara is working on MongoDB and referring to the below Employee Table.
CO3, BL5   Kindly help her to find answers to all queries mentioned below.

   Employee (Emp_id, Emp_name, department, hire_date, salary)

   Write the below queries,
   1) Find all employees in the IT department.
   2) Find employees hired before April 1, 2021, with a salary greater than
      70000.
   3) Update the salary of an employee "Bob" to 80000.
   4) Find the number of employees in each department.
   5) Delete the employee named "Charlie".
   6) Calculate the average salary of employees in the Finance
      department.
   7) Find the employees hired in 2021.
   8) Find an employee with highest salary.

page 2 of 2