24BDAH301 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. BIG DATA ANALYTICS THIRD SEMESTER DEGREE EXAMINATION

NOVEMBER 2025

Data Analytics using SPARK

**Duration:3 Hours** **Max Marks:70**

---

PART A

**I. Answer any FOUR of the following** **(4×5= 20)**

a) Highlight the significance of SparkContext object. Develop code to demonstrate initializing SparkContext in Python.

b) Discuss the three ways to reduce the processing time of batches when using Spark for streaming.

c) Discuss any six Numeric RDD operations used in descriptive statistics on RDDs.

d) Describe in detail the purpose of memory inside each executor.

e) Consider a dataframe containing movie details: Movie_ID, Title, Category, User_ID, Rating. Write queries for the following in Spark SQL:

1. List details of movies that user_ID 5 has reviewed with a rating of 4.
2. List Movie title and average rating for the movies with the word "Desh" in the title.
3. List details of movies that user_ID 5 has not reviewed yet.

PART B

**II. Answer any FIVE questions selecting at least one question from each unit:**

**(5×10= 50)**

UNIT-I

2. Analyse how Apache Spark and Mlib assist in 1. Finance 2. Healthcare sectors and 3. Product Recommendation.

3. Differentiate between the reduce(), fold() and aggregate() transformations with suitable examples.

UNIT-II

4. Consider two dataframes df1 containing: Student_ID, Student Name, Open elective chosen, Marks scored and df2 containing: Student_ID, Student Name, Department, Semester. Using python code with the two given dataframes, demonstarte the working of inner join, outer join, left outer join and right outer join.

5. Describe in detail the working of Accumulators and Broadcast variables in Apache Spark. Write Python code to demonstrate the use of Accumulators for computing the line count in a file containing user data.

## UNIT-III

6. Highlight the features of Apache Mesos as a Cluster Manager. Discuss the recommendation guidelines to be followed when choosing a cluster manager.

7. Discuss in detail the following properties supported by the standalone cluster manager in Apache spark: 1. The two deployment modes . 2. The two settings used to control resource allocation.

## UNIT-IV

8. Consider a JSON file with student details: Register_No., Department, Gender, Age, Hometown, Elective chosen, CGPA. Import the file into Spark SQL and write queries for the following:
   1. List the number of male students who have scored a CGPA of 8.2.
   2. Display the details of 10 students who belong to the Physics department.
   3. List the details of students who belong to the Chemistry department with hometown Ahmedabad.
   4. Display the count of Male and female students.
   5. List the number of students who belong to the Computer Science department with a CGPA of 7.0 and above.

9. Explain the Spark Streaming process with a detailed workflow diagram.

**********

24BDAH302 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. BIG DATA ANALYTICS THIRD SEMESTER DEGREE EXAMINATION

NOVEMBER 2025

Artificial Intelligence

Duration:3 Hours Max Marks:70

---

PART A

I. Answer any FOUR of the following (4×5= 20)

a) With an illustrative example, analyse the following Search strategies: Depth-First search and Iterative deepening search.

b) Comment on the significance of each of the three quantities used to express the complexity of an algorithm used to solve searching based problems in AI.

c) Explain the concept of Stochastic Game. Justify why game playing is considered a significant application of Artificial Intelligence.

d) With examples enumerate the various syntatic components used to form sentences in propositional logic.

e) With a suitable example, describe the working of a typical Multi-Agent system.

PART B

II. Answer any FIVE questions selecting at least one question from each unit: (5×10= 50)

UNIT-I

2. For each of the following activities, give a PEAS description of the task environment:
   1. Playing football. 2. Satellite Image Analysis System.
   3. Performing a high jump. 4. Shopping for used Cars on the Internet.

3. Discuss the concept of Heuristics with a suitable example. With an illustrative example explain the working of Greedy best-first search and the A* search algorithms.

UNIT-II

4. Differentiate between Stochastic Hill Climbing, First-choice hill climbing and Simulated Annealing search algorithms.

5. 1. Describe the concept of Constraint Satisfaction Problems with suitable examples.
   2. Explain briefly the three common techniques followed in solving Constraint Satisfaction Problems of AI.

## UNIT-III

6. 1. Discuss the two Inference rules for quantifiers: Universal Instantiation and Existential Instantiation.

   2. Consider the following four sentences in the knowledge base.

   1. Knows(John, Jane). 2. Knows(y, Bill). 3. Knows(y, Mother(y)).

   4. Knows(x, Elizabeth).

   Describe the concept of Unification for the query AskVars(Knows(John, x))

7. Consider the following problem: The law says that it is a crime for an American to sell weapons to hostile nations. The country Nono, an enemy of America, has some missiles, and all of its missiles were sold to it by Colonel West, who is American.

   Write first-order definite clauses and construct the proof tree using Backward-Chaining to prove that West is a criminal.

## UNIT-IV

8. Describe the working principle of Expectation-Maximization algorithm with a suitable example.

9. Explain the significance of applying robotics in various applications. Describe the key hardware components in a Robot.

**********

24BDAS304 Reg No : .....

**CHOICE BASED CREDIT SYSTEM**

**M.Sc. BIG DATA ANALYTICS THIRD SEMESTER DEGREE EXAMINATION**

**NOVEMBER 2025**

**Cloud Computing**

**Duration:3 Hours** **Max Marks:70**

---

**PART A**

**I. Answer any FOUR of the following** **(4×5= 20)**

a) Discuss the advantage and disadvanatage of binary translation.

b) Elaborate on the cloud definitions proposed by:
1. Armbrust 2. RajKumar Buyya.

c) Explain how 'cloud definition' has become a challenge in cloud computing.

d) Elaborate on Thread priorities and Type serialization in Aneka threads.

e) Explain the additional computing services offered by Google AppEngine.

**PART B**

**I. Answer any FIVE questions selecting at least one question from each unit:** **(5×10= 50)**

**UNIT-I**

2. Explain the hierarchy of privileges in the form of ring-based security with an illustration.

3. Discuss any five platforms and technologies that uses cloud computing.

**UNIT-II**

4. Explain briefly Aneka SDK.

5. a) Explain profiling and monitoring services provided by fabric service.
b) Explain resource management services provided by fabric service.

**UNIT-III**

6. Explain the concept of workflow with respect to task dependencies.

7. Elaborate on variety of computing categories for task computing.

**UNIT-IV**

8. Elaborate on S3 key concepts.

9. Explain the concept of NoSQL system.