22BDAH301 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. THIRD SEMESTER DEGREE EXAMINATION DECEMBER 2023

BIG DATA ANALYTICS

Data Analytics using SPARK

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

PART A

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

a) Discuss the architecture of the Spark stack with an illustratiion.

b) Explain with syntax the general format for spark-submit command.

c) Explain in detail the three methods required to implement custom partitioners in Apache Spark.

d) Discuss the two deploy modes supported by the standalone cluster manager in Apache spark.

e) With an illustration, describe how a DStream is processed as a continuous series of RDDs.

PART B

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

UNIT-I

2. Explain as to why distinct() is considered an expensive transformation.

3. Consider a text file data.txt. Use transformations and actions to demonstrate creation of an RDD to count the occurences of each word. Also write a function to check if a given string is present in the RDD.

UNIT-II

4. Create an RDD with the words ["Big", "Data", "Analytics", "Hadoop", "Hive", "Spark", "SQL"]. Create a pair RDD with the length of each string as the key. Group all the rows with the same key into a single row. Identify the output of reduceByKey(), groupByKey(), sortByKey, values(), keys(), on this pair RDD. Filter out the lines which are longer than 6 characters.

5. Describe the need for shared variables when executing in a cluster environment. Develop Python code using Accumulator for computing blank line count in a text file containing log data.

UNIT-III

6. Discuss in detail the different purposes of memory within executors.

7. Develop a python program to count the number of warnings in a log file consisting of log messages of textual information. Demonstrate Sparks phases of execution on this example application using both the RDD graph and the Physical plan of execution.

8. Explain the fault tolerance mechanisms available for driver, workers and receivers in Spark streaming.

9. Explain any five mechanisms to be considered for improving the performance of MLlib in Spark applications.

****

22BDAH302 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. THIRD SEMESTER DEGREE EXAMINATION DECEMBER 2023

BIG DATA ANALYTICS

Artificial Intelligence

Duration:3 Hours Max Marks:70

PART A

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

a) Describe the various dimensions / characteristics of task environments for
1. Refinery Controller and 2. Automated Taxi driving.

b) Differentiate between Informed and Uninformed search strategies using suitable examples.

c) 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)).

d) Explain the term Sample Space and Belief State with an example each.

e) Explain the concept of Reinforcement Learning with an example. Discuss any two applications of Reinforcement Learning.

PART B

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

UNIT-I

2. With a detailed illustrative example, explain the complete workflow of a Genetic Algorithm.

3. Develop the problem formulation for the following problems: 1. A discrete Vacuum cleaner world. 2. Airline travel planning website.

UNIT-II

4. Express the following sentences in First Order Logic.
1. One's mother is one's female parent. 2. Male and female are disjoint categories.
3. Parent and Child are inverse relations. 4. A grandparent is a parent of one's parent.
5. A sibling is another child of one's parent. 6. One's husband is one's male spouse.
7. Sonal is either an accountant or Professor. 8. Only one student knows French.
9. All professors are teachers. 10. Someone at School is Smart.

5. a. With examples enumerate the various syntatic components used to form sentences in propositional logic.

b. Consider the following five sentences of the Wumpus World example labelled R1 to R5. P stands for Pit and B stands for Breeze in the 4x4 grid of the wumpus world environment. The wumpus is located in 1x3 and pits are located at 3x1 and 3x3.

R1: ⌐P1,1. R2: B1,1 <=> (P1,2 v P2,1). R3 : B2,1 <=> (P1,1 v P2,2 v P3,1).

R4 : ⌐B1,1. R5 : B2,1. Using Inference rules and Equivalences prove that there is no pit in 1,2 nor in 2,1.

UNIT-III

6. With pseudocode explain the Variable Elimination algorithm for inference in Bayesian networks.

7. Describe in detail the four basic inference tasks that must be solved in temporal models.

UNIT-IV

8. Explain the working of Candidate Elimination algorithm with the help of a pseudocode.

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

****

22BDAS304 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. THIRD SEMESTER DEGREE EXAMINATION DECEMBER 2023

BIG DATA ANALYTICS

Cloud Computing

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

PART A

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

a) Discuss on the following computing platforms and technologies:
a) Microsoft Azure b) Google AppEngine

b) Explain VMware as desktop virtualization.

c) Explain the following components that perform a variety of tasks in IaaS:
1. Pricing and billing component 2. Monitoring component 3. Reservation component
4. VM repository component 5. Provisioning component

d) Explain HPC and MTC.

e) Explain Access control and service bus provided by AppFabric.

PART B

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

UNIT-I

2. Briefly Summarize the following :
1. Hardware-assisted virtualization. 2. Operating system-level virtualization.

3. Discuss the concept of mainframe computing, cluster computing and grid computing that acts as major milestones that have led to cloud computing.

UNIT-II

4. Elaborate on the concept of public clouds.

5. a) Explain scheduling service provided by application service.
b) Explain execution service provided by application service.

UNIT-III

6. Elaborate on file management with respect to task based applications.

7. Discuss on the concept of thread API's with an illustration.

UNIT-IV

8. Explain the concept of data grids.

9. Elaborate on the concept of MapReduce framework.

****

22BDAE315 Reg No : .....

CHOICE BASED CREDIT SYSTEM

M.Sc. THIRD SEMESTER DEGREE EXAMINATION DECEMBER 2023

BIG DATA ANALYTICS

Fundamentals of Big Data Analytics

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

PART A

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

a) Elaborate on supervised machine learning.

b) Elaborate on Real-Time Benefits of Big Data.

c) Explain the following Components of Cloud Computing Architecture:
1. Management 2. Application 3. Runtime Cloud 4. Storage 5. Infrastructure

d) Explain Type 1 hypervisor with an illustration.

e) Elaborate on any 05 tools used for data analytics.

PART B

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

UNIT-I

2. Explain the following features of distributed file system:
1. User mobility 2. Performance 3. Data integrity 4. High reliability 5. Scalability

3. Elaborate on the application and benefits of Big Data Analytics.

UNIT-II

4. a) Elaborate on any 05 Cloud Security Risks. (5)
b) Elaborate on any 05 benefits of cloud security. (5)

5. Explain on SAAS. Discuss on the advantages and disadvantages of SAAS.

UNIT-III

6. Explain the following virtualization tools for Developers:
1. Microsoft Hyper-V 2. Red Hat Virtualization

7. Explain virtualization in cloud computing.

UNIT-IV

8. Elaborate on the real-world implementation and the uses of Prescriptive Analysis.

9. Discuss on the concept of Quartile Deviation.

****