19BDH301 REG.NO.....

CHOICE-BASED CREDIT SYSTEM THIRD SEMESTER M.SC. DEGREE EXAMINATION
FEBRUARY 2021

M.Sc. BIG DATA ANALYTICS

BIG DATA ANALYTICS USING SPARK

Time: 3 Hrs Max. Marks: 70

PART - A

Answer any EIGHT of the following: **(2X8=16)**

1. How do you define Spark Context?
2. Define Actions in Spark.
3. What does a Spark Engine do?
4. Name the components of Spark Ecosystem.
5. What are broadcast variables?
6. What are the various data sources available in Spark SQL?
7. List any 4 transformation functions on two pair RDDS.
8. What is Worker Node in a Spark application?
9. Explain accumulators in Apache Spark.
10. Explain Caching in Spark Streaming.

PART – B

Answer any FOUR of the following: **(6X4=24)**

11. Explain the key features of Apache Spark in detail.
12. What are benefits of Spark over MapReduce?
13. Is there a module to implement SQL in Spark? How does it work?
14. What file systems does Apache Spark support and explain in detail.
15. List the functions of Spark SQL with example
16. How is Spark SQL different from HQL and SQL?

PART – C

Answer any THREE of the following: **(10X3=30)**

17. Explain the concept of Resilient Distributed Dataset (RDD) with an example and explain how DAG works.
18. Explain how Spark runs applications with the help of its architecture.
19. How would you compute the total count of unique words in Spark? Write a simple Scala / python program and explain each line of code.
    Assume the data resides inside this file path hdfs://Hadoop/user/test_file.txt
20. What do you understand by Executor Memory in a Spark application? Explain in detail with an example
21. What do you understand by Lazy Evaluation? Explain in detail with an example

****

19BDH302 REG.NO.....

CHOICE-BASED CREDIT SYSTEM THIRD SEMESTER M.SC. DEGREE EXAMINATION-FEBRUARY 2021

M.Sc. BIG DATA ANALYTICS

ARTIFICIAL INTELLIGENCE AND DEEP LEARNING

Time: 3 Hrs Max. Marks: 70

PART - A

Answer any EIGHT of the following: **(2X8=16)**

1. What are the four components of a well defined problem?
2. Justify that A* search is both complete and optimal.
3. With example demonstrate generalized modus ponens.
4. Write agent program for a simple reflex agent in the two state vacuum environment.
5. List the quantifiers used in first-order logic.
6. Draw a typical Bayesian network .
7. What is hybrid architecture?
8. What is meant by reinforcement learning?
9. Draw a Bayesian network structure corresponding to a first – order Markov process and second- order Markov Process.
10. Describe the core formula for Q-learning.

PART – B

Answer any FOUR questions: **(6X4=24)**

11. Illustrate the general model of learning agent.
12. With example explain the basic inference tasks that must be solved in temporal models.
13. With suitable example explain generalized events.
14. List the Backus-Naur form grammar of sentences in propositional logic.
15. Draw a decision tree to exemplify the problem for deciding "Whether to wait for a table?" Assume suitable attributes for the problem.
16. How backoff model is better than smoothing process? Discuss.

PART – C

Answer any THREE of the following: **(10X3=30)**

17. Design an alpha-beta search algorithm to illustrate the alpha-beta pruning by tracing its behaviour.
18. Explain the algorithm for smoothing with a fixed time lag of d steps.
19. Explain the concept of problem solving agent with pseudo code .
20. Suppose that the goal is to conclude the colour of a pet named fritz, given that he croaks and eats flies, and that the rule base contains the following four rules:
    If X croaks and X eats flies- then X is a frog.
    If X chirps and X sings – then X is a bird
    If X is a frog- then X is Green.
    If X is a bird - then X is yellow
    Solve using forward chaining assuming suitable facts.
21. Describe how to construct an active temporal deference learning agent through an exploratory Q- learning agent.

****

19BDS304 REG.NO.....

CHOICE-BASED CREDIT SYSTEM THIRD SEMESTER M.SC. DEGREE EXAMINATION FEBRUARY 2021

M.Sc. BIG DATA ANALYTICS

ADVANCED MACHINE LEARNING AND NLP

Time: 3 Hrs Max. Marks: 70

PART - A

Answer any EIGHT of the following: **(2X8=16)**

1. What is Regular Expression? Why it is important?
2. Define the term text/data wrangling?
3. Mention two approaches of parsing?
4. What is Chunking?
5. How do you do Named entity Recognition in Python?
6. What is Information Retrieval process?
7. Write the uses of Optical Character Recognition in NLP
8. Define Supervised learning technique.
9. What is a Web crawler and how does it work?
10. Why do you use Logistic regression model for classification?

PART – B

Answer any FOUR of the following: **(6X4=24)**

11. List out and explain the ways to achieve tagging task in NLTK.
12. Explain the variety of methods that are available for machine translation.
13. With suitable examples, explain the concept of speech recognition.
14. Define Machine learning and explain different types of machine learning techniques.
15. Explain Support Vector Machine (SVM) algorithm.
16. With an example explain the data flow process in scrappy.

PART – C

Answer any THREE of the following: **(10X3=30)**

17. Define text wrangling. Explain process of text wrangling with example.
18. With the help of a diagram explain in detail the dependency parsing mechanism.
19. Explain the concepts of Vector space model and probabilistic model.
20. With suitable illustrations explain how Naïve Bayes can be used to build text classifier.
21. Through an example highlight how decision trees can be used for predictive modelling techniques.

****

19BDE315 REG.NO.....

CHOICE-BASED CREDIT SYSTEM THIRD SEMESTER M.SC. DEGREE EXAMINATION
FEBRUARY 2021

M.Sc. BIG DATA ANALYTICS

FUNDAMENTALS OF DATA ANALYTICS

Time: 3 Hrs Max. Marks: 70

PART - A

Answer any EIGHT of the following **(2X8=16)**

1. What is a Distributed File System?
2. Why is Big Data Important?
3. What is Artificial Intelligence?
4. What are the components that make up the three parts of a cloud computing solution?
5. What are the traits of an "as a service" offering?
6. List out the benefits of using Cloud Computing.
7. What is virtualization?
8. What are the three major types of Server virtualization?
9. What is Big Data?
10. Define Predictive Analytics.

PART - B

Answer any FOUR of the following **(6X4=24)**

11. Explain about the drivers for Big Data.
12. Write short notes on AI and Machine Learning.
13. What is cloud computing and how does it work?
14. Write in brief about Server Virtualization.
15. Write briefly about how a business can use Prescriptive Analytics to its benefit.
16. List out the design principles to be followed for a successful service like Amazon.

PART – C

Answer any THREE questions **(10x3=30)**

17. Explain in Detail about the 4 V's of Big Data.
18. Write in detail about
    a. Cloud Security Services
    b. Issues in Cloud Security
19. Explain in detail about the Cloud Components, their purpose and role in delivering a cloud based application.
20. Explain in detail the tools and the products available for virtualization.
21. Write in Detail about the types of Data Analytics available in business.

****