MCA CT 204- Data Science & Big Data Analysis

Model questions

Module 1

Part A- 3 Marks Questions

1. Explain the steps of knowledge discovery from data

2. Define Knowledge base

3. What motivated Data Mining? Why is it important?

4. What are the major components in data mining architecture?

5. What is Data Mining?

6. Distinguish between the relational databases and transactional databases

7. What is a data warehouse ?

8. Explain the basic two categories of Data Mining task?

9. Illustrate data Discrimination with Example

10. Write down any three Data mining Task primitives

11. List out the applications of data mining

12. What are the major issues of data mining?

13. Illustrate the technologies used in data mining

14. What is decision Support system ?

15. Summarize Major Functionalities of Data Mining

16. What are Data objects?

17. What is an Attribute?

18. Explain Data Integration

19. Why data cleaning is needed in Data Mining ?

20. What do you mean by data Integration?

21. How do we handle missing values in data preprocessing?

22. Explain how we handle noisy Data

23. Discuss Data Reduction Methods

24. Basic Methods of Attribute subset selection

25. What is OLTP?

26. What kinds of data can be mined

27. Differentiate classification and clustering

28. Differentiate data warehouse and database

Part B- 9 marks Questions

1. Explain KDD (Knowledge Discovery from Data) process with suitable diagram

2. What kinds of patterns can be mined in data mining ? Explain

3. Explain the different kinds of Databases used in Data Mining

4. Explain the Architecture of a typical data mining system with diagram

5. What are the major issues of Data Mining?

6. Explain basic preprocessing Methods

7. What is Data Reduction ? Briefly explain different methods used in data Reduction

8. Difference between OLAP and OLTP

9. Illustrate the discretization and concept hierarchy generation methods

10. How do we handle data cleaning process in Data Mining

Module 2

Part A- 3 Marks Questions

1. Define itemset and frequent item sets?

2. Explain the strategy in Market Basket Analysis

3. Define Frequent itemset and closed item set.

4. Discuss how we can represent the frequent patterns.

5. Discuss association rule mining.

6. What is Apriori property?

7. How the Apriori property is used in Apriori Algorithm?

8. Illustrate how association rules are generated from frequent item set of a transactional database.

9. Explain Apriori algorithm.

10. What are strong association rules?

11. Explain support and confidence.

12. Explain hash-based technique.

13. Discuss the strategy adopted in FP growth.

14. How can you estimate the accuracy of the classifier built?

15. Compare Classification and Prediction.

16. Explain the two steps in classification.

17. Differentiate training data set and test data set.

18. Compare supervised learning and unsupervised learning process.

19. Discuss partitioning method of clustering.

20. Why did you say that clustering is an unsupervised learning process?

Part B- 9 marks Questions

1. Discuss Market Basket Analysis. Briefly explain how Association rules are generated.

2. Discuss FP Growth Algorithm with an example.

3. Explain Apriori Algorithm in detail

4. Illustrate FP Growth principle with an example

5. Explain the importance of Classification in Data mining.

6. Explain the general approach for building a Classification model in detail.

7. Briefly explain the various Clustering methods

8. Discuss the requirements of cluster analysis in Data mining

Module 3

Part A- 3 marks Questions

1. What does Data Science means?

2. Which are the facets of data science?

3. What are the benefits and uses of Data Science?

4. Explain the different facets of data in Data Science?

5. Describe the facets of data?

6. What are the different Data Science process?

7. Explain Data Science process?

8. Describe Data Exploration.

9. Explain data preparation in Data Science process?

10. Describe the steps involved in data preparation in Data Science process?

11. What is Big Data? Explain unstructured data?

12. Explain the three V's of Big Data?

13.What is the importance of unstructured data in Data Science?

14. What does Big Data Means?

15.What is Web Analytics? Explain credit risk management?

16. What is the purpose of Web Analytics?

17. Explain with diagram credit risk management?

18. Explain Big Data and algorithmic trading?

19.Briefly explain Big Data in healthcare

20. How is Big Data useful in healthcare?

21. Briefly explain Big Data in medicine?

22. Explain the different applications of Big Data?

23. How is Big Data useful in advertisement?

Part B- 9 marks Questions

1. Which are the facets of data science?

2. What are the different Data Science process?

3. Explain Web Analytics and Credit risk management?

4. Explain the different facets of data in Data Science

5. Describe the steps involved in data preparation in Data Science process

6. Explain Big Data in healthcare?

Module 4

Part A- 3 marks Questions

1. What do you know about the term “Big Data”? Explain its applications?

2. Differentiate between structured & unstructured data?

3. What are the four V’s of Big Data?

4. Explain the Types of Big data?

5. What is big data Analytics. Explain it types?

6. Explain parallel computing and its techniques?

7. Write a short note for distributed computing technique for processing large data?

8. Compare distributed and parallel system?

9. What is Hadoop?

10. Briefly explain In Memory Computing (IMC)?

11. Discuss the role of cloud services play in handling big data.

12. Discuss the features of cloud computing that can be used to handle big data.

13. Write about Hadoop Echo system ? Define HDFS?

14. Explain HDFS Architecture ?

15. Define Name node and Data Node ?

16. Define the Fetures of HDFS?

17. Explain about MapReduce ?

18. Write about YARN ?

19. Explain HDFS, list and describe HDFS Commands ?

20. Explain Regions in HBase ?

21. What are the features of HBase ?

22. Write about combining Hbase and HDFS?

23. What are the features of MapReduce ?

24. What are the techniques to optimize MapReduce jobs?

Part B- 9 marks Questions

1. Write a short note on different technologies for handling big data

2. Explain:- i) Structuring Big data
ii) Elements of Big data
iii) Big Data Analytics

3. Explain cloud computing and big data? Elaborate on cloud computing features and its types.

4. Briefly explain the technologies for handling Big Data.

i. Distributed and Parallel computing

ii. In-Memory Computing for Big data.(IMC)

5. Compare the distributed and parallel computing of big data with sufficient explanation

6. Briefly explain the technologies for handling Big Data.

7. Describe data models and computing models of Hadoop and distributed databases.

8. What is MapRdeuce Framework? Diagrammatically explain the working of Map Reduce Architecture

9. What is Hadoop echo system? Explain with the data processing elements YARN and MapRdeuce in detail

10. Describe elaborately on Hadoop architecture with HDFS and MapRdeuce

11. What is HDFS? Diagrammatically explain the architecture of HDFS along with Heartbeat Mechanism.

12. What is Hadoop? Explain the relevance of HBase in Big data.

Module 5

Part A

1. Define RDBMS

2. What are the characteristics of RDBMS?

3. What is non-relational database?

4. Give the characteristics of non-relational database technologies.

5. What is CAP theorem?

6. What are the issues with relational model and non-relational model?

7. Define ployglot persistence

8. What is the relationship between big data and a data warehouse?

9. What is MapReduce framework?

10. What are the functions of mapper and reducer?

11. What is YARN? What are the advantages of YARN?

12. Discuss about the components of YARN

13. What are containers and node managers?

14. Discuss fair scheduling in YARN.

15. What is the role of Hive?

16. What are the components of Hive?

17. Discuss about the data types in Hive

18. How can Hive be accessed? What are the different Hive services?

19. Write any 3 built-in functions and 3 aggregate functions of Hive

Part B

1. Compare the features of RDBMS and big data solutions.

2. Write notes on a) Polyglot persistence b) CAP theorem c)Advantages of YARN

3. What is YARN? Explain the working of YARN.

4. Explain the features of non-relational database.

5. Write notes on a)MapReduce framework b)Polyglot persistence

6. Explain YARN schedulers

7. Write notes on a) Any 9 built-in functions of Hive b) Any 9 aggregate functions of Hive

8. Explain a) Architecture of Hive b) Any 5 built-in functions in Hive

9. Write notes on a) Data types in Hive b) Variables, properties and queries in Hive

10. What are the data types in Hive ? Explain Hive services