

STELLA MARIS COLLEGE (AUTONOMOUS) – CHENNAI 86
DEPARTMENT OF COMPUTER SCIENCE (Self-financing)
BACHELOR OF COMPUTER APPLICATIONS
END SEMESTER EXAMINATION – APRIL 2018

Subject : BIG DATA ANALYTICS

Duration: 3 Hrs

Subject Code: 15CS/UI/BD23

Marks : 100

SECTION – A

Answer ALL the questions:

(20 X 1 = 20)

1. The different formats of data are referred as _____.
a. Unstructured b. Variety c. Multiple d. Structured
2. _____ is the open source frame work to handle big data.
a. SQL b. Hadoop c. Database d. Data Warehouse
3. The method by which companies analyse customer data in an effort to identify patterns and discover relationships between different data elements is often referred to as _____.
a. Data Mining b. Data Digging c. Analyse d. Data management
4. _____ is the name of the programming framework originally developed by Google that supports distributed computing environment.
a. Mapreduce b. Hive c. ZooKeeper d. Mahout
5. The storage challenge of big data include _____.
a. Persistence b. Velocity c. Change d. Value
6. _____ is an analytics which require interactive mode of processing.
a. Real time b. Batch c. Both a & b d. None of these
7. _____ are the gateway to the encrypted data.
a. Cryptographic Keys b. Xml c. Security d. RDF
8. The advantage of in-memory processing is _____.
a. Speed b. Reliability c. Cost effective d. All of these
9. _____ is the core component of Hadoop framework.
a. HDFS b. Hive c. Pig d. Mahout
10. According to analysts, _____ systems can provide a foundation when they are integrated with hadoop.
a. Big data management and data mining b. Data warehousing and BI
c. Management of Hadoop clusters d. Collecting and storing unstructured data

Choose the correct answer:

Fill in the blanks:

11. _____ is normally associated with an employee or a business intelligence consultant.
12. _____ is described as extremely large datasets that have grown beyond the ability to manage and analyze.
13. _____ will fall between structured and unstructured data.

14. _____ is the leading database technology that can store 2 million columns in a single row.
15. Big Data involves _____ processing of data.
16. _____ is copying the contents of a computer's hard drive.
17. Patents are filled and backed with _____ rights.
18. _____ can best be described as a programming model used to develop Hadoop-based applications that can process massive amounts of data.
19. _____ has the world's largest Hadoop cluster.
20. _____ is a platform for constructing data flows for extract, transform and load (ETL) processing and analysis of large datasets.

SECTION – B

Answer ALL the questions:

(5 X 2 =10)

21. What is the significance of evolution of big data technology?
22. What is meant by data privacy protection?
23. What is IPR?
24. What is meant by in- memory processing?
25. What is YARN?

SECTION – C

Answer any EIGHT questions:

(8 X 5 = 40)

26. What are the characteristics of big data?
27. Differentiate parallel and distributed computing.
28. Explain about data analytics.
29. Compare and contrast structured and unstructured data.
30. Write short notes on map reduce with a neat diagram.
31. What are the open source tools available for big data implementation?
32. Explain the storage methods adopted for big data?
33. Write notes on integrating big data with MDM.
34. How business intelligence is arrived by analysis the big data? Explain with real time example.
35. What are steps involved in big data knowledge discovery?

SECTION D

Answer any THREE questions:

(3 X 10 =30)

36. Explain in detail about big data visualization techniques.
37. Discuss on the security issues related with big data implementation.
38. Elaborate on the applications of the big data with an example.
39. Explain in detail about big data governance.
40. Explain in detail HDFS file system.