STELLA MARIS COLLEGE (AUTONOMOUS) - CHENNAI 86
DEPARTMENT OF COMPUTER SCIENCE
END SEMESTER EXAMINATION - NOV 2019

SUB CODE: 15CS/UI/BD23 **MAX MARKS: 100**
SUB: BIG DATA ANALYTICS **DURATION : 3 HRS**

Section A **20 X 1 = 20**
Answer ALL the questions
CHOOSE THE BEST ANSWER

1. _____ is an appropriate example of a programming model for processing and generating large data sets with a parallel, distributed algorithm.
a) MapReduce b) C c) Java d) Windows
2. _____ is/are the examples of unstructured data.
a) image b) Video c) Audio d) All of the mentioned
3. An analytics engine provides a mechanism for accumulating all the customer profiles, insights, and matches as well as capabilities for analysing this data using _____ modeling and reporting tools.
a) descriptive b) prescriptive c) predictive d) diagnostic
4. _____ computing helps to identify and count, filter and associate events from a number of unrelated streams to score alternatives against previously specified predictive models.
a) Distributed b) Stream c) Hadoop d) Cassandra
5. _____ level encryption makes data unusable because many applications cannot analyse encrypted data.
a) File b) OS c) Internal file d) both a and c
6. Big Data has the potential to solve challengesin _____.
a) science and education b) environment and sustainability, medicine, commerce
c) cyber and national security d) All of the mentioned
7. _____ data processing eliminates the problem of slower disk access.
a) In-memory b) Distributed c) Sequential d) Random
8._____ learning provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
a) Artificial b) Machine c) Virtual d) Computer
9. _____ is the incremental copying of data to establish a point in time to which the system can be rolled back.
a) Snapshot b) Replication c) commit d) integrity
10. Hadoop framework works with _____ tools.
a) MapReduce, Hive b) MapReduce, MySQL c) MapReduce, foxpro. d) Heron and Trumpet

FILL IN THE BLANKS

11. _____ company has leveraged its Big Data well to create an extremely accurate representation of what products a customer should buy.
12. _____ analytics is a process that determines patterns of behavior from human-to-human and human-to-system interaction data.
13. IBM's _____ tool helps to visualise the unstructured data.
14. _____ algorithm removes or randomises PII but does not destroy statistical patterns required by a data scientist.
15. _____ refers to creations of the human mind, such as inventions, literary and artistic works, and symbols, names, images, and designs used in commerce.
16. _____ maps tasks across a cluster of machines, splitting them into smaller subtasks, before reducing the results into one master calculation.

17. IBM's _____, an artificial intelligence computer system can answerquestions posed in natural language.
18. _____ data contains dissimilar type of data.
19. The process of running multiple tasks simultaneously is called _____.
20. HDFS provides fault tolerance through _____.

Section B **5 X 2 = 10**

Answer ALL the questions

21. Comment on "Data and Data Analytics are getting more complex".
22. What is data obfuscation?
23. List the different data sources available.
24. Is there any value for anomalies while processing big data? If yes, explain.
25. List the four key computing resources.

Section C **8 X 5 = 40**

Answer any EIGHT of the following

26. List the big data characteristics and explain them in detail.
27. Write short notes on a) Cassandra b) Dynamo
28. Explain any three techniques of data analytics and Reporting done on an unstructured data.
24. Discuss on combining inferences drawn from structured and unstructured analysis and using them together for an overall ranking.
25. List and explain any three specific characteristics and features that a Big Data platform should offer to work effectively with Big Data analytics processes.
26. Explain the rules an organisation should embrace to protect IP.
27. Compare Revolutionary, Evolutionary and Hybrid approach in the implementation of big data analytics.
28. Explain the five levels of maturity in the analytics business maturity model.
29. Explain Map Reduce and YARN.
30. What are the key tasks that when deployed will result in a reliable massive file storage? Explain them.

Section D **3 X 10 = 30**

Answer any THREE of the following

31. Explain any five big data analytics applications.
32. Discuss in detail on Massively Parallel Processing Platforms.
33. Explain the challenges posed by Big Data with respect to storage.
34. Explain in detail the steps that businesses should take to help ensure a smooth deployment of big data analytics.
35. Write short notes on a) Zookeeper b) Hbase c) Hive d) Mahout