K.S INSTITUTE OF TECHNOLOGY, BENGALURU-560109

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

18CS72 -BIG DATA ANALYTICS

EXHAUSTIVE QUESTION BANK

MODULE-I

1. How do you define data, web data and Big Data?
2. Using examples and diagrams classify data as structured, semi-structured, multi-structured and unstructured?
3. Explain the characteristics of Big Data with examples.
4. Consider the usage examples of Big Data for a car company. Assume that company manufactures five models of cars, and each model is available in five colours and five shades. The company collects inputs from customers and sales centres, and inputs of component malfunctions from service centres for different models. The company also uses social media inputs. Explain 3Vs characteristics in this company's data.
5. Explain various classification methods for data and Big Data.
6. Define Analytics, scalability, horizontal scalability, vertical scalability.'
7. Explain:
    a. Massive Parallel Processing
    b. Distributed Computing Model
    c. Cloud Computing
    d. Grid Computing
    e. Volunteer computing
8. Why do we use distributed computing for analytics of large datasets?
9. How-are data architecture layers used for analytics?
10. Explain the function of each of the five layers in Big Data architecture design.
11. List the functions of the ELT at data ingestion layer and at data storage layer.
12. List the functions in data management.
13. Why is data quality important in discovering new knowledge and decision making?
14. List the examples of cloud services for exporting data stores.
15. List the usages of three types of services that clouds offer. List Big Data cloud services, to data sources export from data store, and perform cloud during analytics, visualizations and intelligence

16. What are the traditional systems for data storage? How does in-memory columnar format help in OLAP? Give an example.
17. What is enterprise server? How does enterprise server data store differ from a web server?
18. What are the functions in SQL? List the differences between SQL data store and NoSQL data store.
19. With five applications for NoSQL Databases?
20. How does a Big Data stack help in analytics tasks?
21. Explain the phases in Big Data Analytics
22. How does a Berkeley Data analytics stack help in analytics tasks?
23. How do data inputs help in Big Data based Customer value analytics?
24. Give details of Louis Columbus ten ways using which Big Data analytics is revolutionizing marketing and sales.
25. Describe five data risks, described by Bernand Marr
26. How does Big Data help in credit risk management in financial institutions?
27. How does Big Data Analytics enable prevention of fraud, waste and abuse of healthcare system?
28. Describe ways of usages of Big Data analytics in marketing, sales and advertising
29. Why does Big Data offer the potential to transform the medicine and healthcare system?
30. Why are the Cloud services used for Big Data Analytics for customer acquisition, customer lifetime value analytics and other metrics?

Module 2

1. List Hadoop core components. Describe their usages
2. Explain the features of Hadoop
3. Explain using a diagram Hadoop Ecosystem Components/ Explain using a diagram the distributed storage, resource manager layer, processing framework and application APIs layers in Hadoop.
4. Explain the meanings of Hadoop distributed file system, clusters, Racks, DataNodes, Data Blocks, MasterNode, NameNode and metadata of files.
5. Explain Hadoop streaming and Pipes.
6. Explain Hadoop HDFS features.
7. Explain Hadoop Physical Organization
8. Differences between Hadoop 1 and Hadoop 2