

**Subject:BIG DATA**  
**Sem: 7th**

**SECTION A**

- Ques1. Describe the best practices for big data analysis.
- Ques 2. List out the best practices of Big Data Analytics.
- Ques3. Discuss 5V's of Big Data
- Ques4. Define Mahout
- Ques5. Define Spark
- Ques6. How mahout plays an important role in data mining(Hadoop)?
- Ques 7. How Zookeeper plays an important role in data mining(Hadoop)?
- Ques8. Differentiate between Hive and Pig
- Ques 9. Discuss Characteristics of Big Data
- Ques 10. Define YARN
- Ques 11. List out risks in Thread Pools
- Ques12. What do you mean by Thread Pooling
- Ques 13. Differentiate between Lazy Learner and Eager Learner
- Ques 14. How Classification Algorithm is different from Regression algorithm
- Ques 15. Briefly describe Decision Tree
- Ques 16. What are Pros and Cons Of KNN algorithm
- Ques 17. Differentiate between Lazy Learner and Eager Learner
- Ques 18. How Traditional Approach is different from Streaming Analytics
- Ques 19. What do you mean by Thread Pooling
- Ques 20. Briefly Discuss Streaming Analytics use cases
- Ques 21. Differentiate between Scale Up and Scale Out
- Ques 22. State concept of Segmentation
- Ques23. State the usage of 'filters', 'group', 'orderBy', 'distinct' keywords in pig scripts.
- Ques 24. Define Hive Metastore
- Ques25. Define Stream Computing
- Ques 26. Compare :Horizontal Scalability & Vertical Scalability
- Ques 27. List Characteristics of Stream Computing .
- Ques 28. Define Segmentation
- Ques 29. Compare : Linear Regression and Logistic Regression
- Ques 30. Define JSON
- Ques 31. Compare : PIG and HIVE
- Ques 32. Describe the concept of Clustering in data analytics and provide a real world scenario where it is applied
- Ques 33. List Issues in Data Stream Query Processing

## SECTION B

Ques1. Discuss the benefits of Big Data?How Big Data Analytics can be useful in the development of smart cities?

Ques 2. Discuss challenges of Conventional System

Ques 3. Explain 5V's of Big Data

Ques 4. Explain HDFS with the help of an example

Ques5. Explain the ecosystem of Hadoop

Ques 6. Discuss the use cases of Big Data analytics

Ques 7. Discuss operations of PIG

Ques 8. Discuss: a)MongoDB b) Cassandra c)Spark d) kafka

Ques 9. How a secondary name node differs from the name node in HDFS.

Ques10. Explain the following : a)Mapper Class b) Scaling Out

Ques 11. Explain K-Means Clustering with the help of an Algorithm

Ques 12. Calculate the regression coefficient and obtain the lines of regression for the following data

X : 1 , 2 , 3 , 4 , 5

Y: 6 , 7 , 8 , 9 , 10.

Ques 13. Discuss Challenges of Conventional System .

Ques14. Compare: Linear regression and Logistic Regression.

Ques15. What are Benefits of Streaming Analytics

Ques16. Explain Real Time Analytics Platform

Ques 17. Explain Distributed file system

Ques 18. Explain Naïve Bayes Classifier with the help of an example

Ques 19. Discuss Hive Architecture

Ques 20. Compare the following: HDFS and GPFS

Ques 21. Apply Bayes theorem on following Table to Calculate  $P(\text{Yes} / \text{Overcast})$  ,  $P(\text{No} / \text{Sunny})$

	Outlook	Play
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	NO
10	Sunny	Yes

Ques22. Compare the following: Linear Regression and Logistic Regression

Ques23. Discuss Decision Tree with the help of an example .

Ques24. Explain Real Time Sentiment Analysis

Ques25. Explain working of following phases of Map Reduce with one common example :

a. Map Phase

b. Combiner Phase

c. Shuffle and Sort Phase

d. Reducer Phase

Ques26. Explain a) Classification b) Clustering