Unit I

1. What is Big Data? Which fields comes under big data?
2. What is Big Data? What are the benefits and applications of big data?
3. What is Big Data? Explain characteristics of big data?
4. What is Big Data? What are advantages of big data?
5. What is Big Data? Explain sources of big data?
6. What is Big Data? What are the challenges of big data?
7. Explain types of big data technologies in detail.
8. Short note on Operational Big data Technologies. Give Example.
9. Short note on Analytical Big data Technologies. Give Example.
10. Explain top big data technologies in detail.
11. What is Big Data? Give Examples of Big data.
12. Explain three types of data in big data with example.
13. Explain 4 V’s of Big data.

Unit II

1. What is Hadoop? Explain applications of hadoop.
2. Explain Hadoop Ecosystem in detail.
3. What is Hadoop? Explain Components of Hadoop.
4. Briefly explain Hadoop HDFS. Also explain features of HDFS.
5. Briefly explain Hadoop MapReduce. Give example. Explain with diagram.
6. Briefly explain Hadoop YARN with diagram.
7. Explain YARN architecture with diagram in detail.
8. What is Distributed File System? What is the need of it.
9. What is physical organization of compute nodes and large scale file system in DFS.
10. Explain MapReduce architecture in detail.
11. Define following terms :

a. Map task

b. Reduce task.

c. Grouping by key.

1. Explain HDFS commands in detail.
2. Short note on:
   1. HDFS
   2. YARN
   3. MapReduce
   4. Sqoop
   5. Flume
   6. Pig
   7. Hive
   8. Spark
   9. OOzie

Unit III

9. What is similarity of documents. Give examples.

10. Define

a. Jaccard Similarity of sets.

b. Jaccard-bag Similarity of sets.

11. i) Compute the Jaccard and Jaccard-bag similarities of each pair of the following three sets :

{1,2,3,4},{2,3,5,7} and {2,4,6}

ii) Compute the Jaccard and Jaccard-bag similarities of each pair of the following three sets :

{1,1,1,2},{1,1,2,2,3} and {1,2,3,4}

ii) Compute the Jaccard and Jaccard-bag similarities of each pair of the following two sets :

{a,a,a,b} and {a,a,b,b,c}

12. Define Collaborative filtering as a similar sets problem with examples.

13. Define k-shingles. Give an example. What would be the singles size?

14. How do we built shingles from words. Explain with example.

15. Define Minhashing. How to built characteristics matrix from given sets. Explain with example.

16. Given Universal set {a,b,c,d,e} and sets S1={a,d}, S2={c}, s3={b,d,e} and S4={a,c,d}

Compute characteristic matrix.

17. Write a short note how to minhash a set. Explain with example.

18. Given Universal set {a,b,c,d,e} and sets S1={a,d}, S2={c}, s3={b,d,e} and S4={a,c,d}

Compute characteristic matrix and compute minhash of each set.

19. What are the rules for computing minhash signature?

20. Given has functionsh1(x)=(x+1) mod 5 h2(x)= (3x+1) mod 5 and matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Row | S1 | S2 | S3 | S4 |
| 0 | 1 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 |
| 2 | 0 | 1 | 0 | 1 |
| 3 | 1 | 0 | 1 | 1 |
| 4 | 0 | 0 | 1 | 0 |

compute minhash signature.

21. Given has functions h1(x)=(2x+1) mod 6 , h2(x)= (3x+2) mod 6 , h3(x)=(5x+2) mod 6 and matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Row | S1 | S2 | S3 | S4 |
| 0 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 0 | 0 |
| 2 | 1 | 0 | 0 | 1 |
| 3 | 0 | 0 | 1 | 0 |
| 4 | 0 | 0 | 1 | 1 |
| 5 | 1 | 0 | 0 | 0 |

compute minhash signature.

20. Short note on Locality Sensitive Hashing.

Unit IV

1. Short note on Data-Stream-Management System Architecture with diagram.

2. What is stream data model? Give examples.

3. Define with example :

a. 0th moment

b. 1st moment

c. 2nd moment

4. Write short note on Alon-Matias-Szegedy Algorithm for Second Moment with example.

5. Write short note on working of Alon-Matias-Szegedy Algorithm.

6. Define higher-order moments.

7. Write a short note on Data-Gionis-Indyk-Motwani Algorithm.

8. Explain with example how to maintain DGIM conditions.