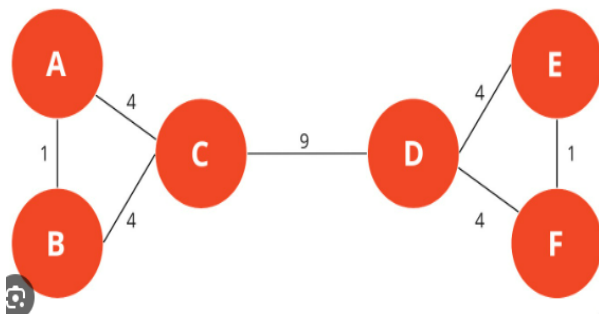


Question Bank:

1. Explain applications of data streams with examples.
2. Elaborate issues in data stream query processing.
3. Explain the sliding window problem with the help of an example.
4. Explain DGIM algorithm for counting ones in stream with given problem $N=24$ and data set is 10101100010111011001011011
5. How bloom filters are useful for big data analytics explain with example.
6. With the help of a diagram explain the data stream management system(DSMS).
7. What are the challenges of querying on large data stream?
8. Suppose the stream is 1,3,2,1,2,3,4,3,1,2,3,1 let $h(x)=6x+1 \bmod 5$ show how the Flajolet-Martin algorithm will estimate the number of distinct elements in this stream.
9. How recommendation is done based on properties of the product?Elaborate with a suitable example.
10. What is jaccard distance and cosine distance in collaborative filtering?
11. A bloom filter with $m=1000$ cells is used to store information about $n=100$ items,using $k=4$ hash functions.Calculate the false positive probability of this instance.will the performance improve by increasing the number of hash function from 4 to 5.Explain your answer.
12. Explain Girvan -Newman algorithm with the help of given example.



13. Enlist and explain different functions used for manipulating and processing data in R.
14. Write the script to sort the values contained in the following vector in ascending order and descending order(23,45,10,34,89,20,67,99).Demonstrate the output.
15. Name and explain the operators used to form data subsets in R.