

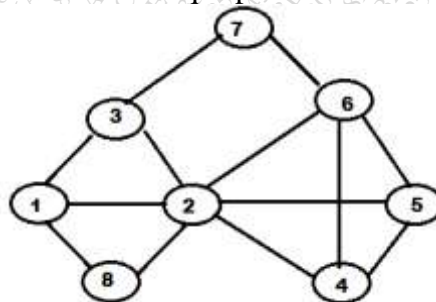
Time: 3 Hours

Marks: 80

Instructions:

- 1) Q.1. is compulsory
- 2) Attempt any three from the remaining
- 3) Assume suitable data

- Q.1. (a) Explain Blooms filter for stream data mining. (5)
 (b) Find the jaccard distance and cosine distance between the following pairs of set: $X=(0,1,2,4,5,3)$ and $Y=(5,6,7,9,10,8)$. (5)
 (c) Explain the steps of the HITS algorithm. (5)
 (d) Explain "Shuffle & Sort" phase and "Reducer Phase" in Map Reduce. (5)
- Q.2. (a) Write a Map reduce pseudo code to multiply two matrices. Illustrate with an example showing all the steps. (10)
 (b) Explain Hadoop Ecosystem with core components. Explain its physical architecture. State the limitations of Hadoop. (10)
- Q.3. (a) Suppose a data stream consists of the integers 1,3,2,1,2,3,4,3,1,2,3,1. Let the Hash function being used is $h(x) = (6x+1) \bmod 5$; estimate the number of distinct in this stream using Flajolet - Martin algorithm. (10)
 (b) Distinguish the following: (10)
 a) PCY, Multistage
 b) Document data store and Column family data store
- Q.4. (a) Give two applications for counting the number of 1's in a long stream of binary values. Using a stream of binary digits, Illustrate how DGIM will find the number of 1's. (10)
 (b) For the given graph show how clique percolation method will find cliques. (10)



- Q.5. (a) Consider the web graph given below with six pages (A, B, C, D, E, F) with directed links as follows. (10)
 $A \rightarrow B, C$
 $B \rightarrow A, D, E, F$
 $C \rightarrow A, F$
 Assume that the PageRank values for any page m at iteration 0 is $PR(m)=1$ and teleportation factor for iterations is $\beta=0.85$. Perform the page rank algorithm and determine the rank for every page at iteration 2.
- (b) Explain clearly how the SON partition based algorithm helps to perform frequent item set mining for large data sets. How does this algorithm avoid false negatives? (10)
- Q.6. (a) Explain collaborative filtering system. How is it different from content based system? (10)
 (b) Clearly explain how CURE algorithm can be used to cluster big data sets. (10)

Instructions

1. Question ONE is compulsory
2. Attempt any three out of remaining five questions.
3. Assume appropriate data and state your reasons
4. Marks are given right of the every question

-
- 1(a) Differentiate between Exhaustive and Effective testing 5M
- 1(b) Define the following with example. 5M
- i) Failure ii) Fault iii) Error
- 1(c) Explain Life cycle of a bug. 5M
- 1(d) Differentiate between CFG and DFG. 5M
- 2(a) A program reads an integer number within the range [1,100] and determines whether it is a prime number or not. Design the test cases for this program using boundary value analysis. 10M
- 2(b) Justify how mutation testing is effective to check the quality of software with an example. 10M
- 3(a) Draw the control flow graph and find the cyclomatic complexity for a program largest of three numbers. 10M
- 3(b) Explain the different types of Incremental Integration Testing Methods. 10M
- 4(a) Briefly explain prioritization techniques. 10M
- 4(b) What is the need of software measurement? Explain different size metrics. 10M
- 5(a) Create a case study on application of ISO 9000 framework to an educational institute. 10M
- 5(b) Explain the guidelines for automated testing. 10M
6. Write Short notes on
- (a) Issues in Object Oriented Testing. 5M
- (b) STLC. 5M
- (c) Challenges in Testing Data Ware house. 5M
- (d) Test Suite Minimization Problem. 5M

(3 hours)

[80 marks]

NOTE: Question No 1 is compulsory. Attempt any three questions from remaining.

Assume suitable data if necessary.

Draw neat labelled diagrams wherever needed.

Q1.

- a. Design and implement ILM for Storage Management system. **10 Marks**
 b. Consider a disk I/O System in which I/O request arrives at the rate of 80 IOPS.

The Disk Service Time is 6 ms.

Compute the following

1. Utilization of IO controller
2. Total Response Time
3. Average Queue Size
4. Total time spent by a request in a queue

10 Marks

Q2 a. An application has 1,000 heavy users at a peak of 2 IOPS each and 2,000 typical users at a peak of 1 IOPS each, with a read/write ratio of 2: 1 . It is estimated that the application also experiences an overhead of 20 percent for other workloads. Calculate the IOPS requirement for RAID 1, RAID 3, RAID 5, and RAID 6.. **10 Marks**

b. Explain FC Protocol Stack and FC SAN topologies. **10 Marks**

Q3 a. Explain in detail the different components required to design Intelligent Storage System. **10 Marks**

b. Explain BC planning lifecycle with an example. **10 Marks**

Q4 a. Explain IP Storage standards. **10 Marks**

b. Explain Multilingual retrieval systems. **10 Marks**

Q5 a. Explain different components of information system and its types. **10 Marks**

b. Explain Network Data Management Protocol (NDMP) **10 Marks**

Q6 Write a short note on **20 Marks**

- a) IP Storage
- b) NAS
- c) Stemming
- d) Symmetric and Asymmetric virtualization
