National Institute of Technology, Silchar
End-semester UG Examination, May 2023

Subject Code: CS-484
Semester: 8th B.Tech.
Duration: 2 hours

Subject: Cloud Computing
Department: Open Elective (All)
Total Marks: 50

Figures in the right-hand margin indicate full marks.
Answer any FIVE questions.

---

1. (a) Given that you have an HDFS cluster with NameNode and DataNode. Each node is configured with 16GB RAM and 1TB HDD. There are 100 such nodes in the HDFS cluster. One node from the cluster is designated as NameNode and others are designated as DataNode. Given that the data size is equal to the metadata size which is approximately 128KB. What is the total capacity of the cluster? Derive the capacity of the HDFS cluster in GB.

   (b) What do you mean by replication? How does it help HDFS? 6+4=10

2. Assume, you are using the LATE Scheduling Algorithm. Given four tasks T_1, T_2, T_3 and T_4. The progress score of T_1 is 0.03 in 5s, the progress score of T_2 is 0.8 in 60s, the progress score of T_3 is 0.4 in 25s, and the progress score of T_4 is 0.06 in 12s. Answer the following questions-

   (a) Calculate the time-to-complete for the task T_1, T_2, T_3, and T_4 using LATE algorithm.

   (b) Which task is the fastest? 8+2=10

3. (a) Identity MapReduce program is the default map and reduce program provided by Hadoop. When you submit MapReduce job, mapper and reducer classes will be invoked automatically when no map and reducer class are specified in MapReduce driver class. Alternatively, it will just collect the output of the program. What will be the output if you input a set of words into the identity MapReduce program?

   (b) What is the difference between vendor lock-in and data lock-in?

   (c) Why do we need a federated service level agreement (FLA)? Justify.

   (d) How does QoS affect the FLA?

   (e) Differentiate between data portability and VM portability.

   2+2+2+2+2=10

4. Assume that an identity manager, called IdM, manages the identity of users. IdM manages the users' passwords, user IDs, phone numbers, etc., and provides authentication services. In this context, answer the following questions-

   (a) What is the domino effect in IdM?

   (b) Can we use OTP as a password for the IdM?

   (c) What is the purpose of digital signature in IdM?

   (d) How do you use public key cryptography in IdM? 2+2+2+4=10

5. Let us assume an email service, called Emailer. Emailer strictly prohibits access the emails by any other devices and users including the administrative staff. Therefore, Emailer has numerous challenges.

   (a) What are the key challenges of Emailer? Describe two challenges.

   (b) Do state-of-the-art email services restrict the access of the users' emails by the service providers? Prove yes or no by citing a suitable example. 6+4=10

6. Let us assume that you are building a datacenter. The data are duplicated on the left and right nodes in a circular double linked list as shown in Figure 1. Assume that there are n nodes in the cluster. Answer the following questions-

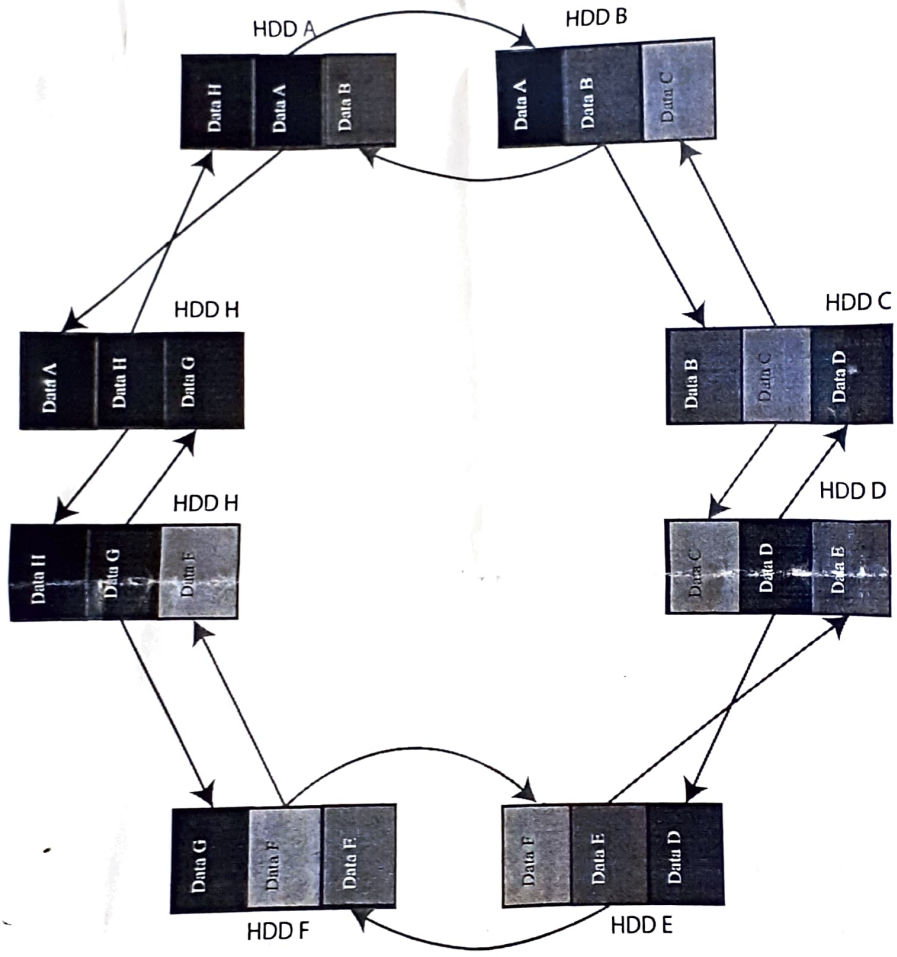

Figure 1: Data storage in a circular double-linked list structure.

(a) In what condition the cluster will cause data loss?

(b) How many node failures can be tolerated by the cluster without data loss in a given time? Express your answer in terms of n.

(c) How many consecutive node failures can be tolerated by the cluster without data loss in a given time?

3+4+3=10

******** END ********