1

A cluster has three racks R1, R2, and R3, each having three servers. R1 has servers C11, C12, C13; R2 has servers C21, C22, C23; and R3 has servers C31, C32, C33. You have a file of 10 MB and you want to keep 3 copies of data in the cluster. What would be the best possible allocation of the data?

a.

C11, C12, C22

b.

C11, C22, C23

c.

C12, C22, C23

d.

C13, C23, C31

2

A map-reduce job runs five map tasks (numbered 1,2,3,4,5) and three reduce tasks. The i-th map task produces 10*i records. There are total 10 distinct keys produced by all the map tasks and each of these map tasks produce these keys with uniform distribution. What is the maximum number of records a reduce task has to process?

3

Choose the correct option in terms of the memory reference time.

a.

Main memory > L1 cache > L2 cache

b.

L2 cache > L1 cache > Main memory

c.

L1 cache > L2 cache > Main memory

d.

Main memory > L2 Cache > L1 cache

4

A cluster has 45000 servers and the lifetime of a server is 1500 days. What is the expected number of server failure per day in the cluster?

5

A node dies just after finishing a reduce task. What does the map-reduce do?

6

If a worker in Map-reduce job dies, what should the right strategy be followed?

a.

Take the job out of the dead worker and assign it to any other active worker chosen randomly

b.

Take the job out of the dead worker and assign it to a live worker who is idle

c.

Take the job out of the dead worker, divide the job into smaller parts, and assign it to other active workers

d.

Wait until all other workers have finished their job and then assign it to any one of such workers

7

A map-reduce job runs five map tasks (numbered 1,2,3,4,5) and three reduce tasks. The i-th map task produces 10*i records with i distinct keys. What is the maximum number of records a reduce task has to process?

8

When do you need to write a combiner to optimize the code?

9

Consider a 10 node graph where nodes are numbered 1,2,… 10. There is an edge between node 5 and 6 also between 4 and 6. The graph has two complete subgraphs of nodes (1,2,3,4) and (6,7,8,9,10). Compute the edge betweenness of the edge 4,6. A complete subgraph is one where every pair of the nodes in the subset have edges between them.

10

A map-reduce job runs five map tasks (numbered 1,2,3,4,5) and three reduce tasks. The i-th map task produces 10*i records with i distinct keys. What is the minimum number of records a reduce task has to process?