

National University of Computer and Emerging Sciences, Lahore Campus



Course:	Introduction to Cloud Computing	Course Code:	CS-499
Program:	BSc (Computer Science)	Semester:	Spring 2018
Duration:	150 Minutes	Total Marks:	55
Paper Date:	20-July-18	Weight	40%
Section:	A and B	Page(s):	5
Exam:	Final (Retake)	Reg. No.	

Instruction/Notes:

- The exam is closed book, closed notes
- Answer all the questions
- Attempt the first two sections on the question paper
- Question paper must be returned with the answer script
- Read the questions carefully before answering them
- If you are certain that something is unclear, make a *reasonable assumption*, mention it and answer the question
- Use of a simple calculator is allowed. Calculators in laptops or smart devices are not allowed

True / False Questions (15 marks)

All questions carry equal marks. Write T (for TRUE) or F (for FALSE) in the “Answer” column. Overwriting will be awarded no credit

Question #	Question	Answer
1	TCP is appropriate for live video streaming whereas UDP is appropriate for pre-recorded video streaming	
2	Netflix uses homogeneous software and hardware configuration for all of its video servers	
3	Youtube prefers Python programming language because it is the most efficient one	
4	The Google Search repository data structure stores compressed copies of all fetched web documents	
5	Cloud service providers offer cost effective solutions because their infrastructure deployments benefit from economies of scale	
6	Typical data center networks use the spanning tree protocol	
7	Netflix uses BGP protocol to connect the client to the nearest video streaming server	
8	A private cloud can not be a multi-tenant cloud	
9	Organizations that have time varying computing demand can gain the most from using a public cloud	
10	Increasing the amount of RAM in a server is an example of scaling out	
11	When a CDN uses URL rewriting to redirect clients to their nearest server, we call it IP AnyCast	
12	All cloud services are free of cost	
13	In the Google File System, a chunk that is consistent must also be defined	
14	Heroku is an example of SaaS	
15	Audit / legal requirements may prohibit the use of a public cloud	

Multiple choice questions (15 marks)

All questions carry equal marks

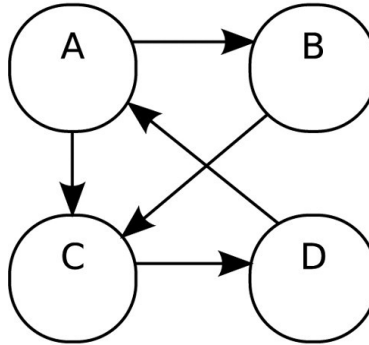
Circle only one correct answer. Overwriting will be awarded no credit

1. Which of the following are true about the Coral CDN?
 - a. Free
 - b. Peer to peer
 - c. Both of the above
 - d. None of the above
2. Which of the following best describes Cache Digest (CD)?
 - a. Caches exchange summary of cached content
 - b. A central directory of cache content
 - c. Both of the above
 - d. None of the above
3. In a CDN, topology informed replica placement:
 - a. Selects replica nodes based on least cost
 - b. Selects nodes with highest outdegree as replicas
 - c. Selects the sites for replica that have the highest traffic near them
 - d. Finds a minimum spanning tree of the underlying network graph
4. In non-cooperative push-based approach, in case of a cache miss:
 - a. Surrogate servers fetch objects from the origin server
 - b. Surrogate servers query a central directory
 - c. Surrogate servers query the nearest replica
 - d. None of the above
5. What characterizes the typical file size in Google File System?
 - a. One huge file
 - b. Mostly huge files
 - c. Mostly small files
 - d. Mostly moderately sized files
6. Which of the following is true about files deleted from Google File System?
 - a. The file chunks are immediately garbage collected
 - b. The file name is immediately removed from metadata
 - c. The file is renamed in the metadata
 - d. None of the above
7. Dropbox added *namespace ID* column to the primary key in their metadata table to:
 - a. Reduce the table size
 - b. Provide better consistency guarantees
 - c. Enable all important data to be kept in memory
 - d. Enable faster searches for a specific user's files
8. Clients can be redirected to the nearest server using which of the following?
 - a. DNS
 - b. IP routing
 - c. HTTP
 - d. All of the above
9. In Google Search, what best describes the hit list?
 - a. A list of documents in which a word occurs
 - b. A list of occurrences of a word within a document
 - c. Both of the above
 - d. None of the above
10. In Google Search, which of the following choice was made for the order of documents in the inverted index for a given word?

- a. Sorted by Document ID
 - b. Sorted by ranking of word occurrence
 - c. Both of the above
 - d. None of the above
11. The problem with maintaining millions of client connections on the DropBox Notification Servers is?
- a. Too much server resource usage
 - b. Restoring from failure becomes a problem
 - c. Both of the above
 - d. None of the above
12. Which of the following best describes the benefit(s) of using commodity, commonly used hardware instead of high-performance but novel hardware?
- a. Economical
 - b. Abundant knowledge base
 - c. Both of the above
 - d. None of the above
- a.
13. When the DropBox Notification Servers maintained connections to every database server, which of the following problems occurred?
- a. Notification servers crashed
 - b. Database servers got overloaded
 - c. Both of the above
 - d. None of the above
14. DropBox users can share file chunks for which of the following reasons?
- a. Conserving space usage
 - b. Security
 - c. Both of the above
 - d. None of the above
15. In a web graph, page A has a hyperlink to page B and page B has a hyperlink to page C. There are no other hyperlinks. In each PageRank algorithm iteration, the PageRank values from the previous iteration are used. Using the defaults, the PageRanks for the pages A, B and C, respectively are:
- a. 0.333, 0.333, 0.333
 - b. 0.433, 0.433, 0.433
 - c. 0.333, 0.433, 0.433
 - d. 0.333, 0.433, 0.518

Quantitative and reasoning questions (10 marks)

1. Calculate the PageRank for all the nodes in the following graph. In each iteration, when calculating the PageRank for any node, use the PageRanks from the previous iteration. Use $d=0.85$ and choose the initial PageRank for each node as discussed in class. Perform at least 10 iterations or until convergence is reached. [10 marks]



2. With the help of a labelled diagram, describe how a client writes data in the Google File System. Clearly show all the data exchanged between all entities that are involved. [10 marks]
3. A startup is developing a cloud based social network in the form of a photo album which allows users to share photos with their “friends”. Each photo uploaded by a user will be replicated on three servers. In the context of this replication, what kind of consistency guarantees should the application provide? [5 marks]