

# 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:	180 Minutes	Total Marks:	80
Paper Date:	24-May-18	Weight	40%
Section:	N/A	Page(s):	5
Exam:	Final	Reg. No.	

## Instruction/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

$$PR(A) = (1-d) + d \sum_{i=1}^n \frac{PR(T_i)}{C(T_i)}$$

**True / False Questions (20 marks)**

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

<b>Question #</b>	<b>Question</b>	<b>Answer</b>
1	Internet Cache Protocol (ICP) messages are transported over TCP	
2	Netflix uses homogeneous software and hardware configuration for all of its video servers	
3	Youtube primarily uses Java programming language	
4	The lexicon in Google Search includes pointers to the repository	
5	A public cloud, such as Amazon AWS, lowers upfront investment in compute infrastructure	
6	From a management point of view, it is preferable to have two different networks in the data center. One for communication between servers and the other for access to network storage	
7	In the Google File System, a chunk that is consistent must also be defined	
8	Youtube servers at colocation sites host moderately played content. These servers are, therefore, underutilized	
9	In the Google Files System, client sends the data to be written to the replica that holds the lease on the respective chunk	
10	A community cloud is owned by a single organization	
11	URL rewriting is essential in case of full-site replication on a CDN	
12	Grid computing applications run on heterogeneous hardware and software	
13	Cloud data centers are also multi-tenant data centers	
14	The control plane functionality for Dropbox runs on their own servers in their own data centers	
15	EdgeStore is the CDN for Netflix	
16	A database contains all building laws that are almost never updated. Resource replication mechanism is not needed in the resource cluster	

Question #	Question	Answer
17	Dropbox provides eventual consistency guarantees	
18	If a client's viewing experience deteriorates, Netflix transcodes the content being watched on the fly	
19	The original Google Search used Hamming Code to encode the hit list	
20	By default, DNS requests persistently carry client IP address during recursive lookups	

### Multiple choice questions (20 marks)

All questions carry equal marks

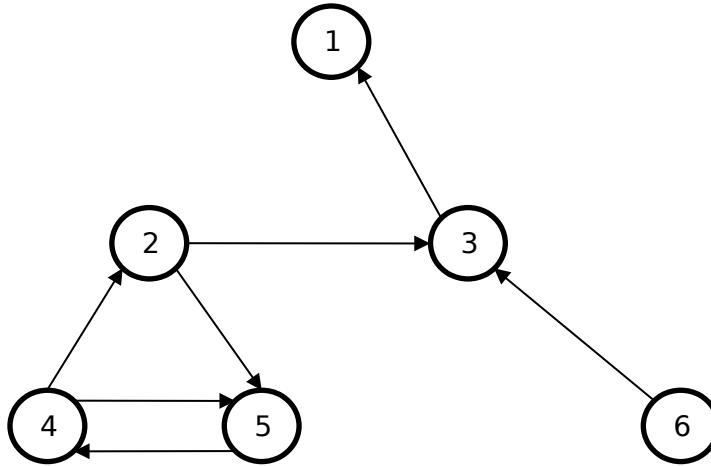
Circle only one correct answer. Overwriting will be awarded no credit

- Which of the following use DNS to redirect clients to their closest server?
  - Netflix
  - Akamai
  - Both of the above
  - None of the above
- In Google Search, when parsing a web document, the hyperlinks are first placed in which of the following?
  - Forward index
  - Anchors file
  - Document index
  - None of the above
- In case of a catastrophic failure in Google Search, which of the following, if survived, can be used to restore all other data structures
  - Inverted index
  - Document index
  - Forward index
  - Repository
- During peak hours, the number of streams from a Netflix video server can be upto
  - 1000
  - 20000
  - 100000
  - 1000000
- A state management database:
  - Makes programs more scalable
  - Increases program throughput
  - Both of the above
  - None of the above
- A cloud service's output is in the form of XML. A device needs to consume this service, but supports JSON formatted data only. Which of the following needs to be used to solve this problem?
  - Resource usage monitor
  - SLA monitor
  - Multi-device broker
  - Load balancer
- In active-active failover mechanism, the client requests are answered by:
  - The primary
  - The replicas
  - Both of the above
  - None of the above
- The original Google Search design used \_\_\_\_ for storing capitalization in a hit
  - 1 bit
  - 3 bits
  - 2 bytes
  - It varies from hit to hit
- An online news startup is likely to:
  - Deploy their own CDN
  - Use a third party CDN
  - Both of the above
  - None of the above

10. Which of the following are likely to affect the performance of a CDN
  - a. Surrogate server locations
  - b. Request routing strategy
  - c. Caching strategy
  - d. All of the above
11. In Openstack, disk files representing volumes attached to VMs are stored in:
  - a. Keystone
  - b. Swift
  - c. Cinder
  - d. Horizon
12. Which of the following is not a CDN
  - a. Akamai
  - b. Limelight Networks
  - c. Both of the above
  - d. None of the above
13. In a CDN, the goal(s) of surrogate placement is (are):
  - a. Minimize client perceived latency
  - b. Minimize overall bandwidth consumption
  - c. Both of the above
  - d. None of the above
14. By default, the swift object storage in Openstack stores objects in:
  - a. MySQL
  - b. Sqlite file
  - c. Cinder store
  - d. Files on local filesystem
15. In Openstack, which of the following picks the physical server to run a VM on?
  - a. Nova-compute
  - b. Nova-scheduler
  - c. Nova-conductor
  - d. Nova-cert
16. The file block / chunk size for Google File System and Dropbox, respectively are:
  - a. 4 MB, 64 MB
  - b. 64 MB, 4 MB
  - c. 4 kB, 64 kB
  - a. 64 kB, 4 kB
17. Which of the following is rare in Google File System?
  - a. A small amount of data read from a random offset
  - b. A large amount of data read from a random offset
  - c. Appends to a file
  - d. Writes to random offsets in a file
18. 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
19. Netflix deploys special cache servers in addition to video servers if:
  - a. There are more than 10000 subscribers in that area
  - b. They have money to spare
  - c. The traffic exceeds 1 Tbps
  - d. They have a good relationship with the host ISP
20. In a web graph, page A has a hyperlink to page B, page B has a hyperlink to page C, page C has a hyperlink to page D and page D has a hyperlink to page A. There are no other hyperlinks. The value of d is chosen as 1. The PageRanks for pages A, B, C and D, respectively are:
  - a. 0.25, 0.25, 0.25, 0.25
  - b. 0.2125, 0.2125, 0.2125, 0.2125
  - c. 0.25, 0.2125, 0.2125, 0.2125
  - d. 0.25, 0.2125, 0.2125, 0.25

### Quantitative and reasoning questions (40 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. [15 marks]



2. With the help of a labelled diagram, describe how a client reads data in the Google File System. Clearly show all the data exchanged between all entities that are involved. [10 marks]
3. Traffic from four client locations A, B, C and D accesses a particular web object. 15%, 25%, 30% and 30% of the object requests originate from locations A, B, C and D, respectively. The clients A, B, C and D are located in the Cartesian space at coordinates (0, 0), (10, 3), (-5, 4) and (6, -3). A CDN operator is considering setting up a surrogate server. Candidate locations are (3, 3) and (2, 3). Which site is better for the surrogate server and why? Show all your working. [10 marks]
4. 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]