

**DI/FCT/NOVA**  
**Mestrado Integrado em Engenharia Informática**  
**Mestrado em Engenharia Informática**

**Cloud Computing Systems**  
**1st Semester, 2021/2022**

**Midterm Test (13/November/2021)**

Wrong answers to V/F questions discount up to the equivalent of the corresponding right answer value. For multiple choice questions, the discount is  $\$1/(n-1)\$$ , with  $\$n\$$  the number of choices. The penalty only accumulates in the context of the same question. For each question, the first wrong answer does not count.

1) Answer the following questions.

- a) \_\_\_\_ (V/F) For a cloud provider, it is good that all customers are from the same industries and have similar requirements.
- b) \_\_\_\_ (V/F) In a cloud platform, a region typically spans multiple continents and it is composed by multiple availability zones.
- c) \_\_\_\_ (V/F) In web application services, the key objective of staging is to avoid downtime when deploying new versions of the software.
- d) \_\_\_\_ (V/F) For implementing a high performance, low latency, blob store, erasure coding is preferable to (standard) replication techniques.
- e) \_\_\_\_ (V/F) For a service with unknown scalability requirements (i.e., there is no bound on the number of users that will be using the service), scaling up (for more powerful machines) is preferable to scaling out (to more machines).
- f) \_\_\_\_ (V/F) Amazon Dynamo get operation returns a context value that can be passed to the following put operation. The context is used to know which keys were accessed, in order to decide if the put can be committed or needs to be rolled back.
- g) \_\_\_\_ (V/F) When a non-master replica receives a request with bounded staleness consistency level, it might need to contact the master or not.
- h) \_\_\_\_ (V/F) For enforcing an RPO equals to 0 in a single master, multi-region deployment, data replication must be performed synchronously.  
Note: The recovery point objective (RPO) is the period of time for which updates might get lost in a failure.

2) Answer the following questions.

- a) \_\_\_\_ (V/F) When Memcached runs in a configuration with a fixed set of servers, each key is stored only in a single server.
- b) \_\_\_\_ (V/F) Unlike Memcached, where it is important to use keys that include a differentiating part for different object types (e.g. the key of a user starts with "u:"), in Redis this is not necessary as Redis natively supports data types.
- c) \_\_\_\_ (A/B/C/D) Which of the following features implemented by CDNs is more important for improving latency for end-users: (1) route optimization; (2) object prefetching; (3) TCP optimizations; (D) all of the previous features.
- d) \_\_\_\_ (V/F) Azure durable functions provide support for orchestrating the execution of multiple Azure functions, with the possibility of feeding the result of one function to other function.
- e) \_\_\_\_ (V/F) In platforms that support serverless functions, the "active pool" consists in a set of VMs that have been used to run a function and have all code necessary to run another call to the same function immediately. NOTE: these platforms maintain a "warm pool" and an "active pool".
- f) \_\_\_\_ (V/F) One key requirement of platforms for supporting micro-service-based applications is to include a mechanism to easily send messages to a micro-service without having to know exactly in which machine the micro-service is running.

- g) \_\_\_\_ (V/F) In Map-reduce, for improving performance, it is possible to have a reducer emit outputs before all mappers finish their execution.
  - h) \_\_\_\_ (V/F) In Map-reduce, the Master node maintains information about the state of each map and reduce task.
- 3) There is no major Cloud provider that is originally from Europe. Do you think we can expect this situation to change in the near future? Justify.

4) Consider you want to develop a service similar to Discord, as the one being developed in this course's project. This service maintains information about users, channels and messages. Channels can be public or private. A private channel has a list of members. Users can subscribe any public channel. Most common operations include accessing the list of message of a channel (with pagination), creating a message, accessing information about users and channels. **NOTE: the following questions must be answered considering this specific service – avoid generic comments. Specify your assumptions regarding the service, when relevant.**

- a) In some countries, there are laws that require providers of communication services to keep the messages exchanged by users of the services for an extended period of time (several years), which can be accessed by the police given an authorization issued by a court.  
Discuss how you could support this requirement for a service similar to Discord – in your answer, consider both channels that remain active and those that users decide to delete, and make sure you explain which cloud services you would use (and any special configuration of the service).

- b) Discuss the differences in the users' experience when using your Discord-based service between using the session consistency and the eventual consistency model – give examples of these differences in the context of accesses to the messages of a channel.

- c) Azure CosmosDB is developing an integrated cache feature. This integrated cache has two separate parts: (1) a cache of items – updates when an individual object is created, updated, deleted or read, the object is added to the cache; (2) a cache of queries, that stores the results of previous queries – a result stored in the cache is invalidated based on a TTL.

With this feature, using a cache based on Redis becomes redundant when implementing the Discord service? Justify explaining how the integrated cache can support all operation or which operations need to be supported using Redis.

- d) When deleting a channel, it is necessary to delete all messages of the channel and to update the information about subscribed channels maintained for each user. Consider the following alternatives for implementing such feature: (1) include a REST method in the application that deletes the channel object from the database and records the information that the channel is to be deleted; additionally include a Timer function that runs periodically, once a day, to update the other objects (messages and users); (2) include a HTTP-trigger based function that executes all the updates to the database. Present advantages and problems of each of these solutions and choose the solution you would select – justify.

Solution 1 – advantages / problems

Solution 2 – advantages / problems

My choice is (justify):

- 5) In Map-Reduce, the result of a map function is written in a set of files, one for each of the reducers that will run. This requires a reducer to read data from multiple files. Discuss whether it would be preferable having all mappers writing in the same set of files, and having reducers reading from a single file. Suggestion: in your answer, consider how Map-reduce handle normal executions, faults and stragglers.