**DI/FCT/NOVA**

**Mestrado Integrado em Engenharia Informática**

**Cloud Computing Systems**

**1st Semester, 2020/2021**

**Final Test (8/January/2021)**

**IMPORTANT NOTE: you should answer with a reasonable letter size. The idea of boxes is not to see how small you can write** ☺

**As an indication, the box for the first question should not hold more than 4-5 lines (and concise and precise answers will be highly appreciated).**

1) In the context of the first project of the course, it would be possible to replicate images to a different region by using an Azure function that, whenever an image was added to the local blob store, would write the image also in the remote blob store. What are the interesting properties of this approach? Justify (contrasting this approach with having the REST method to write the image in both the local and remote blob store) .

|  |
| --- |

2) “In Map-Reduce, a reducer can only start executing after all mappers have finished”. State if this statement is true or false and justify.

Suggestion: consider the logical execution steps of a map-reduce task, in particular, the steps between map and reduce.

| True, because… / False, because… |
| --- |

3) Assume a data set including a log of transfer transactions in a banking system, with the following format, where: **Date** is the date of the transfer, **Source** and **Destination** are identifiers of the source and destination accounts (with the first three digits in an identifier identifying the bank of the account), **Amount** is the amount transferred, **Commission** is the commission paid for the transaction and **State** is either 0 for failed transfers and 1 for successful transactions. The element are separated by a tab (’\t’). Example:

Date Source Destination Amount Commission State

2016-12-06T08:58:35.318+0000 032001234 034007890 100.00 1.25 1 ...

The following Python program processes information from the log of transfer transactions using Spark RDD interface.

result = sc.textFile(’log.log’)

 .map( lambda line: line.split(’\t’))

 .map( lambda row: (row[1], float(row[4]) ) ) .reduceByKey( lambda v1, v2: v1 + v2)

 .filter( lambda pair : pair[0].startWith( ’032’)) a) Explain the problem that the presented program is solving.

|  |
| --- |

b) Do you think it would be possible to optimize this program, making it more efficient, by reordering the operators of the program? Justify.

|  |
| --- |

4) Computing whether a road is congested or not is a typical example of a computation that is performed using stream processing systems. Can this computation be performed using a mini-batch model or should it be performed using a continuous processing model? Justify.

| Can use mini-batch model because… / Need continuous model because… |
| --- |

5) Briefly explain what is paravirtualization and why it was necessary in x86 architectures.

|  |
| --- |

6) A VM can run with a managed disk (or virtual hard disk) stored in a blob store service. Present two techniques used by IaaS services to make access to these virtual disks efficient?

| 1. 2. |
| --- |

7) Consider you want to run some given software (e.g. database) in a given computer. Present reasons to use a solution based on containers and, alternatively, on Virtual Machines.

| Possible reasons to use containers: Possible reasons to use VMs: |
| --- |

8) Consider a container service where users can run their containers. Suppose this service is implemented by running one (or a small number of) VMs in each physical computer, with each VM running multiple containers at the same time. Explain why the use of Copy-on-write File systems helps in making the service running efficiently.

|  |
| --- |

9) Kubernetes is often used to support micro-service platforms. Discuss why this is the case (consider the properties of Kubernetes that could be used for supporting micro-services).

|  |
| --- |

10) Consider that a company wants to deploy a new multi-user online game. The company has a (small) private cloud (data center) where it will host the servers of the game. In which conditions do you think it would be interesting to consider using an hybrid cloud solution? Note: consider both hybrid monocloud and hybrid multicloud approaches in your reply.

|  |
| --- |