Student ID (*Matricola*)

0	0	0	0	0	0
1	1	1	1	1	1
2	2	2	2	2	2
3	3	3	3	3	3
4	4	4	4	4	4
5	5	5	5	5	5
6	6	6	6	6	6
7	7	7	7	7	7
8	8	8	8	8	8
9	9	9	9	9	9

Computing Infrastructures

Course 095897

M. Gribaudo, R.Mirandola

30-06-2016

Last Name / Cognome:

. . . .

First Name / Nome:

. . . .

Answers must be given exclusively on the answer sheet (last sheet): DO NOT FILL ANY BOX IN THIS SHEET

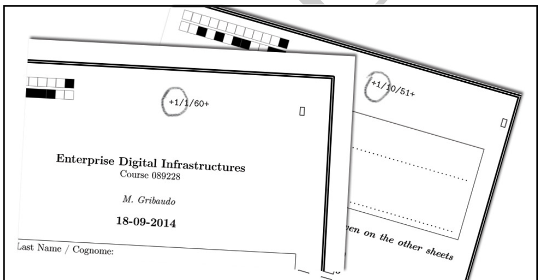
Students must use pen (black or blue) to mark answers (no pencil).
Students are permitted to use a non-programmable calculator.

Students are NOT permitted to copy anyone else's answers, pass notes amongst themselves, or engage in other forms of misconduct at any time during the exam.

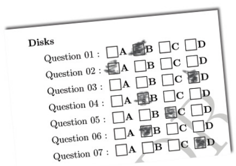
Students are NOT permitted to use mobile phones and similar connected devices.

Scores for the multiple-choice part: correct answers +1 point, unanswered questions 0 points, wrong answers -0.333 points.

You cannot keep a copy of the exam when you leave the room.



Check that the first number in the text and in the answer sheet is the same

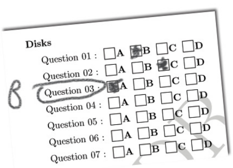


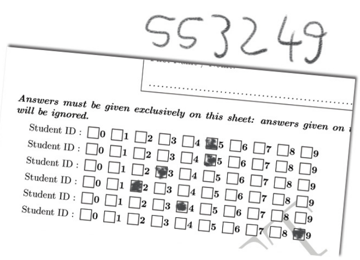
Do not use crosses to mark the answers!

Completely fill the box!

If you make a mistake:

1) circle the word "Question"
2) write the correct answer to its side





Enter the numbers of your "Student ID" from left to right

Disks

Consider the following set of disks connected in RAID 6, with generator $g = 2$:

	Disk 1	Disk 2	Disk 3	Disk 4	Disk 5	Disk 6
	Block A1	Block A2	Block A3	Block A4	Block A5	Block A6
	D0: 0	D1: 5	D2: 1	D3: 3	P:	Q:
	Block B1	Block B2	Block B3	Block B4	Block B5	Block B6
	D0: 1	D1: 1	D2:	P: 5	Q: 23	D3:

Figure 1: A RAID 6 configuration.

Question 1 What data is contained in blocks $A5$ and $A6$?

- [A] $P = 9$, $Q = 76$
- [B] $P = 9$, $Q = 38$
- [C] Cannot be computed
- [D] $P = 9$, $Q = 9$

SOLUTION:
$P = 0 + 5 + 1 + 3 = 9$, $Q = 0 + 5 \cdot 2 + 1 \cdot 4 + 3 \cdot 8 = 10 + 4 + 24 = 38$

Question 2 What data is contained in blocks $B3$ and $B6$?

- [A] $D_2 = 1$, $D_3 = 2$
- [B] $D_2 = 0$, $Q = 3$
- [C] Cannot be computed
- [D] $D_2 = 16$, $Q = 47$

SOLUTION:
$P = D_0 + D_1 + D_2 + D_3$, $Q = D_0 + 2D_1 + 4D_2 + 8D_3$.
$5 = 2 + D_2 + D_3$, $23 = 1 + 2 + 4D_2 + 8D_3$.
$D_3 = 3 - D_2$, $20 = 4D_2 + 24 - 8D_2$. $4D_2 = 4$
$D_2 = 1$, $D_3 = 3 - 1 = 2$

A system administrator has decided to use a stock of disks characterized by $MTTF = 800$ days and $MTTR = 20$ days. If the target lifetime of the system is 3 years:
Question 3 The maximum number of disks that could be used in RAID 0 to have a MTTDL larger the the system lifetime is:

- [A] 1 disk
- [B] None
- [C] $800/20 = 40$ disks
- [D] 2 disks - since RAID 0 requires at least two disks

SOLUTION:
None, since the MTTF of one disks is already less than 3 years, and RAID 0 can only reduce the system's lifetime

Question 4 The maximum number of disks that could be used in RAID 01 to have a MTTDL larger the the system lifetime is:

- [A] No more than 7 disks
- [B] At least 8 disks
- [C] No more than 6 disks
- [D] No more than 58 disks

SOLUTION:
$\frac{2MTTF^2}{N^2 MTTR} > MTTDL$
$N < \sqrt{\frac{2MTTF^2}{MTTDL \cdot MTTR}}$, $N < \sqrt{\frac{2 \cdot 800^2}{3 \cdot 365 \cdot 20}} = 7.6451$. Since the number of disks in a RAID 01 must

be even, the answer is no more than 6 disks.

Virtualization and IaaS

Question 5 Which of the following switches is first traversed by a request starting from a node of a data-center and directed to the WAN, in a three-layers network architecture ?

- A WAN link
- B access
- C aggregation
- D core

SOLUTION:
The first layer is the one directly connected to the physical machines, that is the *access* layer.

Question 6 Which of the following properties of virtualization allows to a virtual machine to be entirely stored in a file?

- A Encapsulation
- B Isolation
- C Partitioning
- D HW-independence

SOLUTION:
Encapsulation: see slides 44 of "L2 - Virtualization"

A program A run in a physical machines is executed in 120 sec, and in a virtualized environment in 121 sec.

Question 7 Knowing that the fraction of instructions that traps is 0.0001%, which is the virtualization overhead?

- A $9900\times$
- B $8333\times$
- C $10000\times$
- D Cannot be computed

SOLUTION:
Since $120(1 + 0.0001/100 \cdot 8333 = 121$, the right answer is $8333\times$.

Question 8 A virtual machine has a network adapter in NAT mode. The virtual machine has IP address 10.0.1.15, while the physical machine has IP address 192.168.5.13. Both the physical and the virtual machines are running a web-server listening on port 80. In order to access the web-server running on the host, the client on the guest must:

- A Use URL `http://10.0.1.15:8080` after setting up a port-forwarding-rule that forwards packet received on port 8080 of the host to port 80 of the guest.
- B Use URL `http://10.0.1.15:80`
- C Use URL `http://192.168.5.13:8080` after setting up a port-forwarding-rule that forwards packet received on port 8080 of the host to port 80 of the guest.
- D Use URL `http://192.168.5.13:80`

SOLUTION:
The connection is from the guest to the host. In the NAT network, outgoing connections are automatically replicated by the VMM on the host network. VMs can reach the web-server using the same address that would be used by a client running on the host, that is `http://192.168.5.13:80`.

Big Data - (4 points)

The following Apache Spark code processes tweets with the aim to understand if they are positive or not

IMPORTANT NOTES:

Comments are as in Java. *Use them to understand what the content of an RDD or the outcome of an instruction should be.*

```


case class Job(JobID:Int,
SubmissionTime: String,
CompletionTime: String)

case class Map(TaskID:Int,
JobID:Int,
ExecutionTimeMillisec: Int)

case class Reduce(TaskID:Int,
JobID:Int,
ReduceTimeMillisec: Int,
ShuffleTimeMillisec: Int)


val job1=Job(1,"08:00:00", "08:10:03")
val job2=Job(2,"08:05:10", "08:09:13")
val job3=Job(3,"09:05:10", "09:09:13")

val jobs=sc.parallelize(List(job1,job2,job3))

val
maps=sc.parallelize(List(Map(1,1,300149),Map(2,1,304190),Map(3,2,30149)))

val
reducers=sc.parallelize(List(Reduce(1,1,198770,100040),Reduce(2,2,19870,100040),
\\Reduce(3,2,32851,180000),Reduce(4,3,19870,100040),Reduce(5,3,32851,180000)))

val mapNumber=maps.map(m => (m.JobID,1)).reduceByKey(_+_)

mapNumber.collect
/* FILL IN */

val reduceNumber=reducers.map(r => (r.JobID,1)).reduceByKey(_+_)
reduceNumber.collect

def sumTasks(tuple:(Int,(Int, Int))) : (Int, Int) = {

val (jobID,(numMap,numReduce))=tuple

(jobID, numMap + numReduce)

}


val totalTask=mapNumber.join(reduceNumber).map(sumTasks)

totalTask.collect

/* FILL IN */

```

```
53
54 /* Compute the average execution time of map tasks */
55 val avgMaps= FILL IN
56
57 avgMaps
58
59 //avgMaps: Float = 166159.25
60
61
62 def isOverlappedJobs(tuple:(Job,Job)) : Boolean = {
63
64 val (j1,j2)=tuple
65
66 val (firstJob, secondJob) = if(j1.SubmissionTime < j2.SubmissionTime)
67 (j1,j2) else (j2,j1)
68
69 (secondJob.SubmissionTime < firstJob.CompletionTime) &&
70 (secondJob.SubmissionTime > firstJob.SubmissionTime)
71
72
73 }
74
75
76 val overlappedJobs=jobs.cartesian(jobs).filter(isOverlappedJobs)
77
78 overlappedJobs.collect
79
80 /* FILL IN */
```

Question 9 Complete line 55.

- [A] val avgMaps= maps.map(m => m.ExecutionTimeMillisec).reduce(_+_)/maps.count
- [B] val avgMaps= maps.map(ExecutionTimeMillisec).reduce(_+_)/maps.count
- [C] val avgMaps= maps.map(m => m.ExecutionTimeMillisec).reduce(_+_).toFloat/maps.count
- [D] val avgMaps= maps.map(m => sum(m.ExecutionTimeMillisec))/maps.count

SOLUTION:
val avgMaps= maps.map(m => m.ExecutionTimeMillisec).reduce(_+_).toFloat/maps.count

Question 10 Select the output compliant with line 32.

- [A] Array[(Int, Int)] = Array((2,1), (1,2))
- [B] Array[(Int, Int)] = Array((2,1), (1,2), (3,2))
- [C] Array[(Int, Int)] = Array((2,1), (1,2), (3,1))
- [D] Array[(Int, Int)] = Array((2,2), (3,2))

SOLUTION:
Array[(Int, Int)] = Array((2,1), (1,2), (3,1))

Question 11 Select the output compliant with line 49.

- [A] Array[(Int, Int)] = Array((2,3), (3,3))
- [B] Array[(Int, Int)] = Array((2,3), (1,3))
- [C] Array[(Int, Int)] = Array((2,4), (1,4), (3,4))
- [D] Array[(Int, Int)] = Array((2,3), (1,3), (3,3))

SOLUTION:
Array[(Int, Int)] = Array((2,3), (1,3), (3,3))

Question 12 Select the output compliant with line 78.

- [A] Array((Job(1,08:00:00,08:10:03),Job(3,09:05:10,09:09:13)))
- [B] Array((Job(1,08:00:00,08:10:03),Job(2,08:05:10,08:09:13)), (Job(2,08:05:10,08:09:13),Job(1,08:00:00,08:10:03)))
- [C] Array((Job(1,08:00:00,08:10:03),Job(3,09:05:10,09:09:13)), (Job(3,09:05:10,09:09:13),Job(1,08:00:00,08:10:03)))
- [D] Array((Job(1,08:00:00,08:10:03),Job(2,08:05:10,08:09:13)))

SOLUTION:
Array((Job(1,08:00:00,08:10:03),Job(2,08:05:10,08:09:13)),
(Job(2,08:05:10,08:09:13),Job(1,08:00:00,08:10:03)))

Big Data and PaaS - (4 points)

Describe the main YARN (Yet Another Resource Negotiator) architectural components implemented within Hadoop 2.x. SOLUTION:

See slides 127-130

Performance - (3 points)

Describe the main issues related to the analysis of the simulation output.
SOLUTION:

See slides 31-56 L07 Simulation

Performance - (7 points)

A computing infrastructure of a company is accessed by N = 100 employees that have a mean think time Z = 20sec. The execution of a typical request requires 10 accesses to the server A, whose service time is S_A = 20ms, utilized at 50% and a certain number of accesses to server B.

1. Define the system model
2. Compute
   (a) the throughput X and the mean response time R of the system
   (b) the number of accesses to the server B by a complete execution of a request knowing that its throughput is X_B = 15 op/sec
   (c) It is known that the mean service time of the server B is S_B = 55 ms. Compute its utilization U_B.
   (d) Compute the maximum throughput that the system may obtain when the number of customers grows to infinite.
   (e) To support the business objective, a maximum throughput of the system of X_{max} = 8job/sec is required. How should the system be changed to meet this new requirement? (describe the characteristics of the new servers)

SOLUTION:

1) We can use a closed model with terminals with Z=20sec and two stations with feedback loops

2)

a) To obtain the throughput X, we focus on Server A. We know that S_A = 20ms and that V_A=10. We obtain then that $D_A = S_A \cdot V_A$=0.02*10=0.2sec. Applying the Utilization law, we have $U_A = X \cdot D_A$, so X=0.5/0.2=2.5 job/sec. The Response time R can be computed as: $R = \frac{N}{X} - Z$, so R=100/2.5-20=20 sec.

b) Applying the forced flow law $X_k = V_k \cdot X$, we have: $V_B = \frac{15}{2.5} = 6$

c) U_B can be computed using the utilization law. We first obtain $D_B = S_B \cdot V_B$ =0.055*6=0.33 sec. Then U_B=2.5*0.33=0.825

d) The bottleneck of the system is server B since it has the highest demand, so we have D_{max} = 0.33sec. The maximum throughput can be obtained as: $X_{\max} = \frac{1}{D_{\max}} = 3.03 job/s$

e) We know that $X_{\max} = \frac{1}{D_{\max}}$, so if we have $X_{\max} = 8j/s$ this implies that $\frac{1}{D_{\max}} = 8job/s$ so $D_{\max} = \frac{1}{8} = 0.125$s. So to be able to obtain this maximum throughput the system should be modified and server B substituted with a new server B_1 with $D_{B_1} \leq 0.125$s and server A too should be substituted with a new server A_1 with $D_{A_1} \leq 0.125$s. A possibility is having a server A_1 twice as fast (with respect to server A), so D_{A_1}=0.1 and a server B_1 thrice as fast so D_{B_1}=0.111

Dependability - (6 points)

In the following questions we will assume that both failure and repair events follow exponential distributions.

We have a four-component system in **series**. The failure rates of all components is the same: $\lambda_A = \lambda_B = \lambda_C = \lambda_D = 0.001 \text{ days}^{-1}$.

1. Calculate the number of days t during which the system has a probability of 0.9 to be continuously working without failure.
2. Calculate the reliability of the system at $t = 20$ days.
3. Calculate the number of these components we can put in series and the system still keeps a probability higher than 0.5 to continuously work for 20 days.
4. Calculate the MTTF of this 4-component system if its configuration is **2-out-of-4**
5. Again in the 4-component **series** configuration, given that $MTTR_A = MTTR_B = MTTR_C = MTTR_D = 50$days. Calculate the system availability.
6. In the system in question 5) , calculate the system MTTR.

SOLUTION:

1) $R(t) = (e^{-0.001t})^4 = 0.9 \Longrightarrow \quad R(t) = e^{-0.004t} = 0.9 \quad \Longrightarrow \frac{ln(0.9)}{-0.004} = t = 26,34$ days
2) $R(20) = (e^{-0.001\cdot 20})^4 = 0.9231$
3) $R(20) = (e^{-0.001\cdot 20})^n = 0.5 \Longrightarrow \quad R(t) = e^{-0.001tn} = 0.9 \quad \Longrightarrow \frac{ln(0.5)}{-0.001\cdot 20} = n = 34.65 \Longrightarrow n = 34$
4) $MTTF = \frac{1}{0.001}(\frac{1}{2} + \frac{1}{3} + \frac{1}{4}) = 1083.33$days
5) $Availability = (\frac{1000}{1000+50})^4 = 0.8227$
6) $Availability = \frac{MTTF}{MTTF+MTTR}$; Availability=0.8227 (calculated in 5)); Then, $MTTF = \frac{1}{4*0.001} =$ 250days; $MTTR = \frac{1-Availability}{Availability} MTTF = 53,87$ days.

Answer sheet:

Last Name / Cognome:

....

First Name / Nome:

....

Answers of the multiple-choice part of the exam must be given exclusively on this sheet

Student ID : ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9
Student ID : ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9
Student ID : ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9
Student ID : ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9
Student ID : ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9
Student ID : ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 ☐8 ☐9

Disks

Question 01 : ☐A ☐B ☐C ☐D
Question 02 : ☐A ☐B ☐C ☐D
Question 03 : ☐A ☐B ☐C ☐D
Question 04 : ☐A ☐B ☐C ☐D

Virtualization and Iaas

Question 05 : ☐A ☐B ☐C ☐D
Question 06 : ☐A ☐B ☐C ☐D
Question 07 : ☐A ☐B ☐C ☐D
Question 08 : ☐A ☐B ☐C ☐D

Spark

Question 09 : ☐A ☐B ☐C ☐D
Question 10 : ☐A ☐B ☐C ☐D
Question 11 : ☐A ☐B ☐C ☐D
Question 12 : ☐A ☐B ☐C ☐D