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

17-06-2016

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

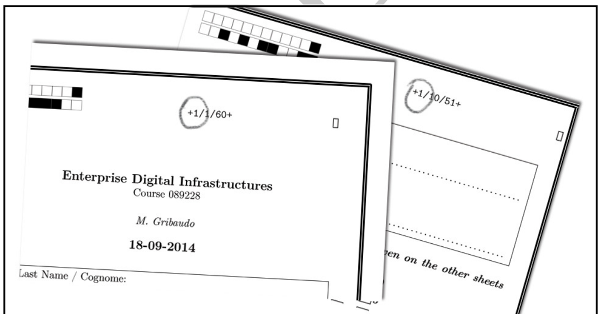
Students must use pen (black or blue) to mark answers (no pencil).
Students are permitted to use a non-programmable calculator.

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.

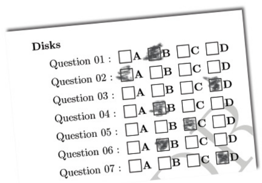
Students are NOT permitted to use mobile phones and similar connected devices.

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

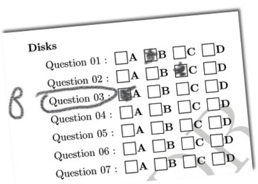


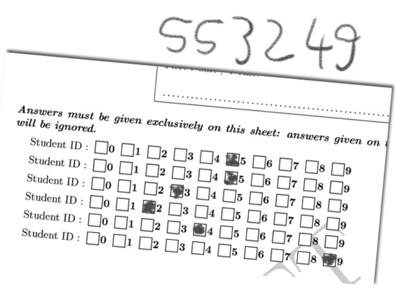
Do not use crosses to mark the answers!

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

For your examination, preferably print documents compiled from auto-multiple-choice.

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 1	D1 1	D2 2	D3 1	P	Q
Block B1	Block B2	Block B3	Block B4	Block B5	Block B6
D0 2	D1 0	D2	P 5	Q	D3 1

Figure 1: A RAID 6 configuration.

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

- [A] $P = 5$, $Q = 5$
- [B] $P = 5$, $Q = 19$
- [C] Cannot be computed
- [D] $P = 5$, $Q = 28$

SOLUTION:
$P = 1 + 1 + 2 + 1 = 5$, $Q = 1 + 1 \cdot 2 + 2 \cdot 4 + 1 \cdot 8 = 1 + 2 + 8 + 8 = 19$

Question 2 What data is contained in blocks $B3$ and $B5$?

- [A] $D_2 = 2$, $Q = 18$
- [B] $D_2 = 2$, $Q = 5$
- [C] Cannot be computed
- [D] $D_2 = 1$, $Q = 5$

SOLUTION:
$D_2 = P - D_0 - D_1 - D_3 = 5 - 2 - 0 - 1 = 2$, $Q = 2 + 0 \cdot 2 + 2 \cdot 4 + 1 \cdot 8 = 2 + 8 + 8 = 18$.

Question 3 If each disk has a MTTF of 1000 days and an MTTR of 30 days, the MTTDL of the RAID 6 will be:

- [A] 9251days
- [B] 18519days
- [C] 5555days
- [D] 1111days

SOLUTION:

$$MTTDL_{RAID6} = \frac{2MTTF^3}{G(G-1)(G-2)MTTR^2} = \frac{2 * 1000^3}{6 \cdot 5 \cdot 4 \cdot 30^2} = 18519 \text{ days}$$

Question 4 If each disk has a capacity of $1TB$ the total storage capacity of the system will be:

- [A] 3TB
- [B] 6TB
- [C] 5TB
- [D] 4TB

SOLUTION:
Since there are 6 disks, and two are used for parity, the total capacity will be 4×1TB, that is 4TB.

Question 5 If the same set of disks as in the previous question would be used as RAID 10, the total capacity would be:

- [A] 4 TB
- [B] 3 TB
- [C] 6 TB
- [D] RAID 10 does not exists

SOLUTION:

Since there are 6 disks, and three are used for data and three for mirroring, the total capacity will be 3×1TB, that is $3TB$

Virtualization and IaaS

Question 6 Which of the following **is not** a layer of the three-layers network architecture of a data-center?

- [A] backbone
- [B] core
- [C] access
- [D] aggregation

SOLUTION:
The three layers are the access, the aggregation and the core. So the wrong one is the *backbone*

Question 7 Which of the following properties of virtualization allows to run multiple O.S. on the same physical machine:

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

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

Consider a physical machine, whose IP address is 192.168.5.33, running two virtual machines: VM_1 with a network adapter in NAT mode, and VM_2 with a network adapter in Bridge mode.

Question 8 Which of the following could be a possible IP address assignemnt for the two VMs?

- [A] $VM_1 = 192.168.5.34$, $VM_2 = 192.168.4.35$
- [B] $VM_1 = 10.0.1.15$, $VM_2 = 10.0.1.16$
- [C] $VM_1 = 10.0.1.15$, $VM_2 = 10.0.1.16$
- [D] $VM_1 = 10.0.1.15$, $VM_2 = 192.168.5.34$

SOLUTION:
NAT addresses are given in a different network, for example $VM_1 = 10.0.1.15$ is ok. Bridge addresses must be given in the same network as the host, for example $VM_2 = 192.168.5.34$

Question 9 If we add VM_3 on the same physical machine, with a Bridge network adapter with IP = 192.168.5.38, and a port-forwarding rule that forwards packet to port 8080 to port 80, how can VM_1 contact the webserver on VM_3 ?

- [A] 192.168.5.33:8080
- [B] 192.168.5.38:8080
- [C] No other choice is correct
- [D] 10.0.1.15:8080

SOLUTION:
Port-forwarding rules are used only for NAT network. Moreover, since it is in bridge mode, the correct choice would be 192.168.5.38:80, which is not given in the list. For this reason the correct answer is "No other choice is correct".

Big Data

The following Apache Spark code processes a csv file containing the measurements coming form sensors in multiple buildings. Buildings 17 and 18 in *area 1* are the one selected for the analysis
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.*

```
/* Let's start by creating a RDD for data coming from a csv file*/

val eventsFile =
sc.textFile("hdfs:///user/zeppelin/SensorDemo/HVAC.csv")

/* Base Event class */
case class Event(Date: String,
Time: String,
TargetTemp: Float,
ActualTemp: Float,
System: String,
SystemAge: Int,
BuildingID: String)


//Build events RDD
val eventsRDD = eventsFile. map(s=> s.split(",")).map(
s => Event(s(0),
s(1),
s(2).toFloat,
s(3).toFloat,
s(4),
s(5).toInt,
s(6)
)
)

eventsRDD.count
// res1284: Long = 8000

eventsRDD.take(2)
/* res1286: Array[Event] =
Array(Event(6/1/13,0:00:01,66.0,58.0,13,20,4),
Event(6/2/13,1:00:01,69.0,68.0,3,20,17))
*/

/* Select critical events, such that ActualTemp and the TargetTemp
are within 2 degrees */
val criticalTempRDD = FILL

criticalTempRDD.cache

// Get events of Building 1
val eventsOfInterest= criticalTempRDD.filter(e => e.BuildingID ==
"1")

eventsOfInterest.take(2)
```

Question 10 Complete line 39.

- [A] eventsRDD.filter(e => e.TargetTemp-e.ActualTemp <2)
- [B] eventsRDD.forall(if (e.TargetTemp-e.ActualTemp <2)) => (e.BuildingID,1)
- [C] eventsRDD.collect
- [D] eventsRDD.filter(TargetTemp-ActualTemp <2)

SOLUTION:
eventsRDD.filter(e => e.TargetTemp-e.ActualTemp <2)

Question 11 Select the output compliant with line 47.

- [A] Array(Event(6/13/13,18:13:19,67.0,68.0,2,14,1), Event(6/7/13,0:33:07,71.0,72.0,18,8,1))
- [B] Array(Event(6/13/13,18:13:19,67.0,68.0,2,14,1) Event(6/7/13,0:33:07,68.0,72.0,18,8,1), Event(6/13/13,18:14:19,67.0,68.0,2,14,1))
- [C] Array(Event(6/13/13,18:13:19,67.0,68.0,2,14,2), Event(6/7/13,0:33:07,68.0,72.0,18,8,1))
- [D] Array(Event(6/13/13,1), Event(6/7/13,1))

SOLUTION:
Array(Event(6/13/13,18:13:19,67.0,68.0,2,14,1), Event(6/7/13,0:33:07,71.0,72.0,18,8,1))

Question 12 Which line is variable CriticalTempRDD (line 39) actually computed for the first time (either partially or totally)?.

- [A] 39
- [B] 47
- [C] 41
- [D] 44

SOLUTION:
47

Big Data and PaaS - (4 points)

Discuss the main performance benefits introduced by Spark with respect to the Hadoop 2.0 framework (remind that this latter includes Tez).

SOLUTION:

Many languages are available, SQL is supported almost natively and Graph and Machine Learning libraries are available.

From the performance perspective, Spark executors are persistent and Spark Jobs run as threads. Operations are pipelined to reduce I/O and RDD can be cached in memory for reuse. Jobs include many tasks (thousands) which last few milliseconds in optimized deployments, reducing the effects of stragglers.

Performance - (10 points)

A cloud SaaS application has a fixed number $N = 100$ of registered users, having a mean think time $Z = 15$sec. During its execution, the application utilizes a server A, having a mean service time $S_A = 0.1$sec, a server B with mean service time $S_B = 0.05$sec and a server C with $S_C = 0.01$sec. The complete execution of a transaction requires one access to server A, 3 accesses to server B, and 6 accesses to server C. The utilization of the server B has been measured, its value is $U_B = 0.9$.

1. Define the system model
2. Compute:
   (a) The throughput of the system
   (b) The throughputs of servers A, B and C
   (c) The demands of servers A,B,C
   (d) The utilizations of servers A and C
   (e) The system response time R
   (f) The system response time when $N = 150$ users
   (g) If the bottleneck server will be substituted with a new one twice as fast, which will be the new bottleneck of the system? With this new configuration, would it be possible to have a system throughput of 12j/s or greater?

SOLUTION:

1) We can use a closed model with a terminal station (time-sharing system) with think time $Z = 15$sec, with $N = 100$ jobs, three stations with the following characteristics:
$S_A = 0.1$sec, $v_A = 1$
$S_B = 0.05$sec, $v_B = 3$
$S_C = 0.01$sec, $v_C = 6$

2)
a) Applying the utilization and the forced flow law to server B we have $U_B = X_B \cdot S_B = X \cdot v_B \cdot S_B$, so $X = \frac{U_B}{v_B \cdot S_B} = \frac{0.9}{3 \cdot 0.05} = 6$j/s
b) Applying the forced flow law $X_k = X \cdot v_k$, we have: $X_A = 6 \cdot 1 = 6$ j/s, $X_B = 6 \cdot 3 = 18$ j/s, $X_C = 6 \cdot 6 = 36$ j/s.
c) Applying the definition of demand $D_k = S_k \cdot v_k$, we have: $D_A = 0.1 \cdot 1 = 0.1$ sec, $D_B = 0.05 \cdot 3 = 0.15$ sec, $D_C = 0.01 \cdot 6 = 0.06$ sec.
d) Applying the utilization law: $U_k = X \cdot D_k$, we have: $U_A = 6 \cdot 0.1 = 0.6$, $U_C = 6 \cdot 0.06 = 0.36$
e) Applying the response time law $R = \frac{N}{X} - Z$ we have: $R = 100/6 - 15 = 1.666$sec.
f) With the following data, we cannot compute the response time at $N = 150$, since X would be different and it is not given as input data.
g) The bottleneck of the system server B since it has the highest demand. If we replace with one twice fast, then $D'_B = 0.075$ sec, and the new bottleneck will become server A. Asymptotically, the maximum throughput of the system can be $X_{\max} = \frac{1}{D_{\max}} = \frac{1}{0.1} = 10$j/s. So the system cannot reach a throughput of 12j/s.

Dependability - (6 points)

The analysis of the failure behavior of a two-component system reveals that the system is down only when both its components are down. One of the components is of TypeA and the other of TypeB, and they have the following failure rates: $\lambda_{TypeA} = 0.005\text{days}^{-1}$, $\lambda_{TypeB} = 0.002\text{days}^{-1}$.

Assuming that failure events follow exponential distribution, calculate:

1. The probability that the system is able to work without any failure up to at time $t = 100$ days.
2. How many additional replicas of component TypeB do we need to achieve a reliability at $t = 100$ of 0.999
3. MTTF of the system

If $MTTR_{TypeA} = MTTR_{TypeB} = 10\text{days}$ and also follow the exponential distribution, calculate:

4. The availability of the system considering repairs.
5. The MTTF of the system considering repairs.

SOLUTION:

This is a parallel system

1) $R(100) = 1-(1-R_A(100))\cdot(1-R_B(100)) = e^{-100\cdot 0.005} + e^{-100\cdot 0.002} - e^{-100\cdot 0.005}\cdot e^{-100\cdot 0.002} = e^{-0.5} + e^{-0.2} - e^{-0.7} = 0.92868$

2) $(1-R_B(100))^n > \frac{1-0.999}{1-R_A(100)}$, $n > \frac{\ln(0.001)-\ln(1-e^{-0.5})}{\ln(1-e^{-0.2})}$, $n > 3.4987$. 3 additional TypeB components are needed.

3) $MTTF = \int_0^\infty R(t)\mathrm{d}t = \int_0^\infty 1-(1-e^{-\lambda_A t})(1-e^{-\lambda_B t})\mathrm{d}t = \int_0^\infty e^{-\lambda_B t} + e^{-\lambda_A t} - e^{-(\lambda_A+\lambda_B)t}\mathrm{d}t = \frac{1}{-\lambda_B}e^{-\lambda_B t} + \frac{1}{-\lambda_A}e^{-\lambda_A t} - \frac{1}{-(\lambda_A+\lambda_B)}e^{-(\lambda_A+\lambda_B)t}\Big|_0^\infty = (0+0-0) - (\frac{1}{-\lambda_B} + \frac{1}{-\lambda_A} - \frac{1}{-(\lambda_A+\lambda_B)}) = \frac{1}{\lambda_A} + \frac{1}{\lambda_B} - \frac{1}{\lambda_A+\lambda_B} = \frac{1}{0.005} + \frac{1}{0.002} - \frac{1}{0.005+0.002} = 557.14$ days.

4) $A = 1 - \left(1-\frac{200}{200+10}\right)\cdot\left(1-\frac{500}{500+10}\right) = 0.99907$

5) $MTTR = \frac{1}{\frac{1}{10}+\frac{1}{10}} = 5$. $A = \frac{MTTF}{MTTR+MTTF} \rightarrow MTTF = MTTR\cdot(\frac{A}{1-A})$. $MTTF = 5\cdot(\frac{0.99907}{1-0.99907}) = 5350$ 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
Question 05 : ☐A ☐B ☐C ☐D

Virtualization and Iaas

Question 06 : ☐A ☐B ☐C ☐D
Question 07 : ☐A ☐B ☐C ☐D
Question 08 : ☐A ☐B ☐C ☐D
Question 09 : ☐A ☐B ☐C ☐D

Spark

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