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

25-07-2016

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

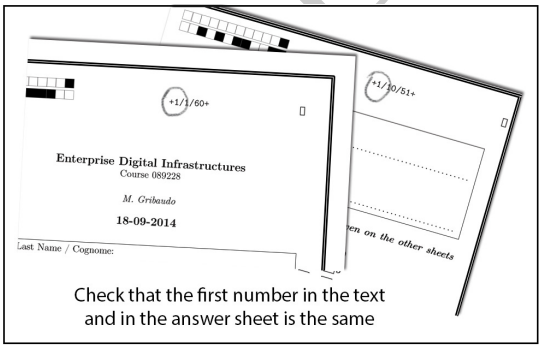
Students must use pen (black or blue) to mark answers (no pencil).
Students are permitted to use a non-programmable calculator.

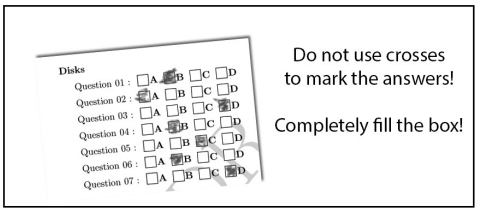
Students are NOT permitted to copy anyone else's answers, pass notes amongst themselves, or engage in other forms of misconduct at any time during the exam.

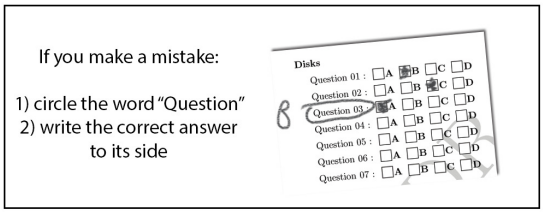
Students are NOT permitted to use mobile phones and similar connected devices.

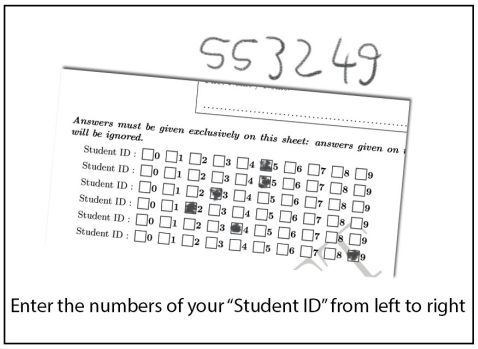
Scores for the multiple-choice part: correct answers +1 point, unanswered questions 0 points, wrong answers -0.333 points.

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









Disks

noindent A HDD has a rotation speed of 6000 RPM, an average seek time of 3 ms, a negligible controller overhead and transfer time of 128 MB/s. Files are stored into blocks whose size is 4 KB.

Question 1 The rotational latency of the disk is:

A 6 ms B 5 ms C 10 ms D 3 ms

SOLUTION:
The rotational latency is half of the time required to perform one rotation:
$T_l = \frac{60000}{2 \cdot 6000} = 5$ ms.

Question 2 The average time required to read a 4 KB block is:

A 8.0305 ms B 8.2441 ms C 13.0305 ms D 13.244 ms

SOLUTION:
The total average transfer time is:
$T_a = T_l + T_s + F/r_t = 5 + 3 + 4/(128 * 1024) * 1000 = 8.0305$ ms

Question 3 The time required to read a 128 KB file with a locality of 96.875% is:

A 248.98 ms B 8.9766 ms C 256.98 ms D 13.977 ms

SOLUTION:
The file is composed by $128/4 = 32$ blocks. Then we have:
$T_a = 32 \cdot ((1 - l) \cdot (T_l + T_s) + B/r_t) = 32 \cdot ((1 - 0.96875) \cdot (5 + 3) + 4/(256 * 1024) * 1000) = 8.9766$ ms

Question 4 The minimum number of disks that have to fail in a RAID 15 to have a data loss is:

A 3 B 2 C 1 D 4

SOLUTION:
A RAID 15 is a RAID 5 applied to couple of RAID 1 disks. The first level (RAID 1) will have a failure if two disks fails in the same group at the same time. The second level will fail (RAID 5) if there are two groups failed at the same time. To summarize, the system will fail if two groups, of two disks each, fail at the same time, which corresponds to a minimum of 4 disks.

Virtualization and IaaS

Question 5 Which of the following equipment **is not** usually put in a rack inside a data center?

A Power distribution units
B Batteries (UPS units)
C Servers
D Storage units

SOLUTION:
Batteries, since they are usually stored in a specific room to allow a separate maintenance with respect to other IT equipment installed in the racks.

Question 6 Which of the following properties of virtualization allows provisioning / migration of a given VM on a given physical server?

A HW-independence
B Isolation
C Partitioning
D Encapsulation

SOLUTION:
HW-independence: see slides 44 of "L2 - Virtualization"

Question 7 In the context of virtualization, "Microkernel" refers to:

- [A] The core of a modified operating system used to perform para-virtualization
- [B] A hypervisor running directly on the hardware, where device drivers are part of a service virtual machine
- [C] A hypervisor running on an existing operating system
- [D] A hypervisor running directly on the hardware, where device drivers are part of its code

SOLUTION:
A hypervisor running directly on the hardware, where device drivers are part of a service virtual machine: see slides 36 of "L2 - Virtualization"

Question 8 A cloud in which a company holds its own cloud infrastructure, but can rent virtual machines when needed is a:

- [A] Public Cloud
- [B] Private Cloud
- [C] Community Cloud
- [D] Hybrid Cloud

SOLUTION:
Hybrid Cloud: see slides 33 of "L3 - Cloud Computing"

Big Data - (4 points)

The following Apache Spark code processes the logs of a mobility system, used to monitor a highway. In particular logs include data on drivers, weather conditions and possibly traffic violations.
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.*

```
1
2
3  case class Event(eventType: String,
4  isCertifiedDriver: String,
5  paymentScheme: String,
6  hoursDriven: Int,
7  milesDriven: Int,
8  lat: Double,
9  long: Double,
10 isFoggy: Int,
11 isRainy: Int,
12 isWindy: Int)
13
14
15 val event1 = Event("Normal","N","miles",70,3300,-95.01,36.73,0,1,1)
16 val event2 = Event("Lane
17 Departure","N","miles",70,3300,-91.99,37.94,0,0,0)
18 val event3 = Event("Overspeed","N","miles",60,3400,-92.99,37.34,1,0,0)
19 val event4 = Event("Unsafe following
20 distance","Y","miles",70,3300,-91.18,38.22,1,1,1)
21 val event5 = Event("Unsafe following
22 distance","N","miles",0,30,-89.63,39.84,1,0,0)
23
24 val eventsRDD =
25 sc.parallelize(List(event1,event2,event3,event4,event5))
26
27 eventsRDD.take(2)
28 /* res41: Array[Event] =
29 Array(Event(Normal,N,miles,70,3300,-95.01,36.73,0,1,1),
30 Event(Lane Departure,N,miles,70,3300,-91.99,37.94,0,0,0))
31 */
32
33 // Select certified drivers events
34 certifiedDriversEvents = FILL IN
35 certifiedDriversEvents.collect
36
37 /* Array(Event(Unsafe following
38 distance,Y,miles,70,3300,-91.18,38.22,1,1,1)) */
39
40
41 // Determine events total miles
42 eventsMilesRDD = FILL IN
43 eventsMilesRDD.collect
44 /* Array[(String, Int)] = Array((Unsafe following distance,3330),
45 (Lane Departure,3300), (Overspeed,3400), (Normal,3300)) */
46
47
48 // Verify if for overspeed violations, miles per hours are greater
49 // than 55
50 def verifyOverspeed(e : Event) : Boolean = {
51 (e.milesDriven.toDouble/e.hoursDriven > 55 )
52
```

```
53 }
54
55 verifiedOverspeedEvents = FILL IN
56 verifiedOverspeedEvents.collect
57 /* Array[Event] =
58 Array(Event(Overspeed,N,miles,60,3400,-92.99,37.34,1,0,0)) */
59
60 // Compute total miles for non-certified drivers filtering overspeed
61 // events
62 nonCertifiedMiles = FILL IN
63 nonCertifiedMiles.collect
64 /* Array((N,3430)) */
```

Question 9 Complete line 34.

- [A] var certifiedDriversEventsRDD.map(e => e.isCertifiedDriver == "Y")
- [B] var certifiedDriversEventsRDD.filter(isCertifiedDriver == "Y")
- [C] val certifiedDriversEventsRDD.filter(e => e.isCertifiedDriver == "Y")
- [D] val certifiedDriversEventsRDD.filter(isCertifiedDriver)

SOLUTION:
val certifiedDriversEventsRDD.filter(e => e.isCertifiedDriver == "Y")

Question 10 Complete line 42.

- [A] var eventsMilesRDD = eventsRDD.map(e => (e.eventType, e.milesDriven)). reduce (_ + _)
- [B] var eventsMilesRDD = eventsRDD.map(e => (e.eventType, e.milesDriven)). reduce (_ + _)
- [C] val eventsMilesRDD = eventsRDD.map(e => (e.eventType, e.milesDriven)). reduceByKey (_ + _)
- [D] val eventsMilesRDD = eventsRDD.map(e => (e.eventType, e.milesDriven)). reduce (_ + _)

SOLUTION:
val eventsMilesRDD = eventsRDD.map(e => (e.eventType, e.milesDriven)).
reduceByKey (_ + _)

Question 11 Complete line 55.

- [A] val verifiedOverspeedEvents = eventsRDD.filter(verifyOverspeed)
- [B] var verifiedOverspeedEvents = eventsRDD.filter(e => e.eventType == "Overspeed").map(verifyOverspeed)
- [C] var verifiedOverspeedEvents = eventsRDD.map(e => e.eventType == "Overspeed").filter(verifyOverspeed)
- [D] val verifiedOverspeedEvents = eventsRDD.filter(e => e.eventType == "Overspeed").filter(verifyOverspeed)

SOLUTION:
val verifiedOverspeedEvents = eventsRDD.filter(e => e.eventType == "Overspeed").filter(verifyOverspeed)

Question 12 Complete line 62.

A val nonCertifiedMiles = eventsRDD.filter(verifyOverspeed).filter(e => e.isCertifiedDriver == "N").reducebyKey(_ + _)

B val nonCertifiedMiles = eventsRDD.filter(verifyOverspeed).filter(e => e.isCertifiedDriver == "N").map(e => (e.isCertifiedDriver, e.milesDriven)).reduceByKey(_ + _)

C val nonCertifiedMiles = eventsRDD.filter(verifyOverspeed).filter(e => e.isCertifiedDriver == "N").map(e => (e.isCertifiedDriver, e.milesDriven)).reduce(_ + _)

D var nonCertifiedMiles = eventsRDD.filter(verifyOverspeed).filter(e => e.isCertifiedDriver == "N").map(e => (e.isCertifiedDriver, e.milesDriven)).reduce(_ + _)

SOLUTION:
val nonCertifiedMiles = eventsRDD.filter(verifyOverspeed).filter(e => e.isCertifiedDriver == "N").
map(e => (e.isCertifiedDriver, e.milesDriven)).reduceByKey(_ + _)

Performance

Question 13 The service time S_k of a service center k can be defined as:

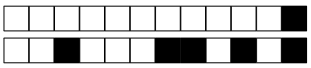
- [A] The time spent by a job in a single visit at service center k
- [B] The ratio between the utilization of service center k and the system throughput
- [C] None of the other answers
- [D] The total time spent by a job when visiting service center k

SOLUTION:
The service time S_k is defined as the time spent by a job for each single visit to server k

Question 14 Considering a Queueing Network model, which of the following sentences is true?

- [A] The service discipline characterising a queue is always Firt-in-first-out
- [B] The jobs arriving at a service center k should always wait in a queue before being served
- [C] All the service centers are single servers
- [D] None of the other sentences

SOLUTION:
The jobs can wait in a queue only if the service center is not an Infinite server and if there is another job being served by the server when it arrives The service centers can be single server, multiple servers or infinite servers the service disciplines are FIFO, LIFO, Priority, Random,... So the correct answer is None of the other answers



Big Data and PaaS - (4 points)

Define what a *Resilient Distributed Dataset* is and how it is managed by the Spark framework.
See slides 245, 247-248, 263

Performance - (8 points)

Let us consider a computing infrastructure that consists of Servers A,B,C and D and that can be accessed by a large number of users. The execution of a single request must pass through: server A, then server B, then it is directed to server C for the 50% of the time and to server D for the remaining 50%. After the service in C or D it goes back to server B before leaving the system. During a period of observation T=30 minutes we have collected the following information:

- Number of system completions: 450
- Number of system completions at server A: 450
- Number of system completions at server B: 900
- Number of system completions at server C: 225
- Number of system completions at server D: 225
- Server A Busy time: 1080 sec
- Server B Busy time: 810 sec
- Server C Busy time: 675 sec
- Server D Busy time: 675 sec

1. Define the system model
2. Compute
   (a) the throughput X of the system
   (b) The service demands of all the servers and determine the bottleneck
   (c) The utilizations of all the servers
   (d) The number of visits at all servers
3. Study and discuss the effect of the following changes on the maximum system throughput and minimum response time:
   (a) Server A is substituted with one having double speed
   (b) A new server A is added equal to the original one and the load is balanced between the two servers

SOLUTION:

1) We can define an open model like the one in Figure 1
2)

a) $X = \frac{C}{T} = \frac{450}{1800} = 0.25 job/s$
b)$D_k = \frac{B_k}{C}$, so we obtain $D_A = \frac{1080}{450} = 2.4$, $D_B = \frac{810}{450} = 1.8$, $D_C = \frac{675}{450} = 1.5$ $D_D = \frac{675}{450} = 1.5$. The bottleneck is Server A since $D_{\max} = 2.4$
c) $U_k = X \cdot D_k$, so we obtain $U_A = 0.25 \cdot 2.4 = 0.6$, $U_B = 0.25 \cdot 1.8 = 0.45$, $U_C = U_D = 0.25 \cdot 1.5 = 0.375$
d)$V_k = \frac{C_k}{C}$, so we obtain $V_A = \frac{450}{450} = 1$, $V_B = \frac{900}{450} = 2$, $V_C = \frac{225}{450} = 0.5$, $V_D = \frac{225}{450} = 0.5$.

3)

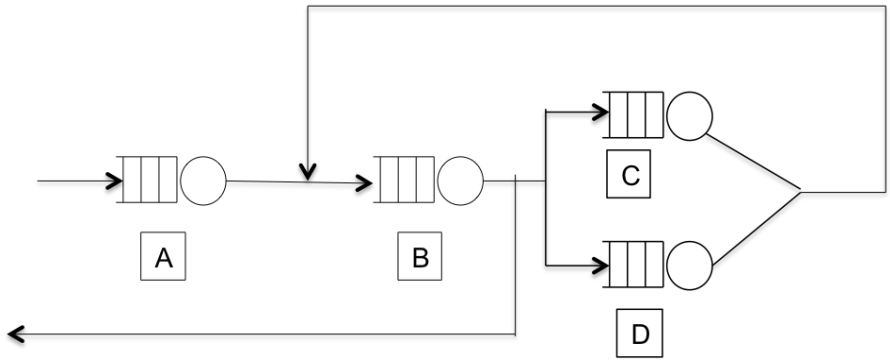


Figure 1: QN model

The original system has the following characteristics: $X_{\max} = \frac{1}{D_{\max}} = 0.4167 job/s$ and R > D, so R > 7.2 sec

a) Server A is substituted with one having double speed: so we have $D_{Anew} = 1.2$, the new $D_{\max}$ is now the $D_B = 1.8$, so the first change leads to these new values: $X_{new1\,\max} = \frac{1}{D_B} = 0.555 job/s$ and $R >$ D, so $R_{new1} > 6$ sec

b) A new server A is added equal to the original one and the load is balanced between the two servers: $D_{A1} = D_{A2} = 1.2$ the new $D_{\max}$ is now the $D_B = 1.8$, so the first change leads to these new values: $X_{new2\,\max} = \frac{1}{D_B} = 0.555 job/s$ As concerns the response time, now D is unchanged with respect to the original system so $R_{new2} > 7.2$ sec

+1/11/50+

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

Dependability - (6 points)

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

We have components of TypeA whose measured reliability at time $t = 2$ days is $R(2) = 0.9$.

1. Calculate the reliability at time 5 of a system made of two components of TypeA in **series**
2. Calculate the reliability at time 10 of a system made of three components of TypeA in **parallel**
3. Calculate the reliability at time 5 of a system made of four components of TypeA in a configuration **2-out-of4**
4. Calculate the availability of the system *System*1 in Figure 2 being the MTTR of a component of TypeA 2 days.

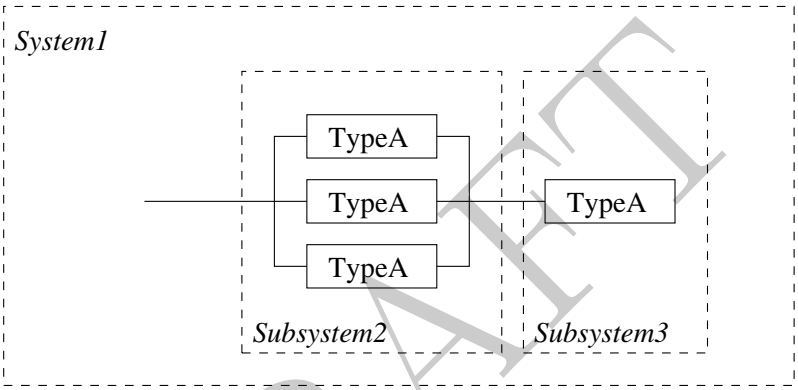


Figure 2: *System*1 model including *Subsystem*2

5. Calculate the MTTF of the box called *Subsystem*2 in Figure 2 being the MTTR of a component of TypeA 2 days.
6. Calculate how many components of TypeA we can append in **series** to the *Subsystem*3 in Figure 2 so that the system availability is still higher than 0.5.

SOLUTION:

1) $R_A(2) = e^{-\lambda 2} = 0.9 \implies \lambda = \frac{-ln(0.9)}{2} = -0.05268$ days $\implies R(5) = (e^{-0.05268 \cdot 5})^2 = 0.5905$ OR $R(5) = (R_A(2)^{\frac{5}{2}}2)$
2) $R(5) = 1 - (1 - e^{-0.05268 \cdot 10})^3 = 0.9313$
3) $R(5) = 1 - 1 \cdot (e^{-0.05268 \cdot 5})^0 (1 - e^{-0.05268 \cdot 5})^4 - 4 \cdot (e^{-0.05268 \cdot 5})^1 (1 - e^{-0.05268 \cdot 5})^3 = 0,95896$
4) $MTTF = \frac{1}{0.05268} = 18,98$; $Availability_{TypeA} = \frac{18,98}{18,98+2} = 0.90468$; $Availability = (1 - (1 - Availability_{TypeA})^3) Availability_{TypeA} = 0.90389875$
5) $Availability_{SubSystem2} = 1 - (1 - Availability_{TypeA})^3 = 0.999134$; $MTTR_{Subsystem2} = \frac{1}{\frac{1}{2}+\frac{1}{2}+\frac{1}{2}} = 0.6666$ days; $MTTF_{Subsystem2} = \frac{Availability_{SubSystem2}}{1 - Availability_{SubSystem2}} MTTR_{Subsystem2} = 769.15$ days
6) $0.5 = Availability_{SubSystem2} \cdot (Availability_{TypeA})^n \implies n = \lfloor \frac{log(\frac{0.5}{Availability_{SubSystem2}})}{log(Availability_{TypeA})} \rfloor = \lfloor 6.91095 \rfloor = 6$. We can append 6 components in total, or 5 more.

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

Performance

Question 13 : ☐A ☐B ☐C ☐D
Question 14 : ☐A ☐B ☐C ☐D