



Computing Infrastructures

September 5, 2024

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points.** An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 Warehouse-scale computers are primarily used in small businesses with limited computing needs.

A False

B True

Question 2 Given their efficiency, accelerators like GPUs and TPUs have no impact on the cooling requirements of datacenters.

A True

B False

Question 3 Hypervisors can only be used to run virtual machines with the same amount of memory as the physical machine on which they are hosted.

A True

B False

Question 4 Leaf-spine topologies are not suitable for highly virtualized environments with a large number of virtual machines.

A True

B False

Question 5 GPUs in datacenters are primarily used for accelerating graphics rendering and video processing tasks.

A True

B False

Question 6 RAID 10 provides better random write performance than RAID 6.

A True

B False

Question 7 Virtualization can reduce costs by enabling the sharing of hardware resources between virtual machines.

A True

B False

Question 8 RAID architectures are used to substitute regular backups

A False

B True

Question 9 A UPS provides backup power to equipment during a power outage.

A True

B False

Question 10 Cooling towers are not suitable for high-density datacenters.

A True

B False



Exercises

Correct answer: +2, No answer: 0.

The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

Question 11

A monitoring system for a rack inside a data center is designed to be composed of 2 hardware sensors (M_{HW}), all necessary to extract information on the use of each blade. Each of the sensors is characterized by a MTTF of 27 days and a MTTR of 1 day. If the monitoring system were to be updated to include a purely software system (M_{SW}), from which it would be possible to extract the same usage information obtained with the HW sensors, what would be the availability of the entire blade-usage monitoring system ($M_{HW} + M_{SW}$) if the availability of the software system (M_{SW}) was 96%? Use always at least 5 decimal digits for each calculation.

$$A_{HW} = \frac{MTTF}{MTTF + MTTR} = \frac{27}{27+1} = 0.96429$$
$$A_{SW} = 96\%$$
$$A_{sys} = 1 - (1 - 0.96429) \times (1 - 0.96429^2) = 0.9971$$

Question 12

A company uses a temperature sensor to monitor a critical environment within an industrial plant. The sensor has a MTTF (Mean Time To Failure) of 4000 hours. If the company decides to replace the original sensor with a redundant version consisting of two sensors like the previous one, what is the overall reliability of the monitoring system after 2000 hours? Use always at least 4 decimal digits for each calculation.

$$R(2000) = e^{-\frac{2000}{4000}} = 0.6065$$
$$R_{sys} = 1 - (1 - 0.6065)^2 = 0.8452$$

**Question 13**

Consider the following RAID 1+0 setup composed of 6 disks, each one with an MTTF equal to 410 days and an MTTR equal to 2 days, and a single mirror case for the RAID 1 part. What is the MTTF of the storage infrastructure?

$$\text{MTTF}_{\text{RAID}10} = \frac{(\text{MTTF}_{\text{single}})^2}{N \times \text{MTTR}} = \frac{410^2}{6 \times 2} = 14.008$$

Question 14

You are tasked with performing capacity planning for an application cluster. Your system can be modeled using a closed model. On average, there are $N = 30$ users logged into the system, with a think time of $Z = 10$ seconds. Your objective is to determine whether it is more efficient to introduce:

- 10 servers, **each** with a service demand of $D_{10} = 10$ seconds, or
- A single server that is 10 times faster than each server in the previous case, with a service demand of $D_1 = 0.1$ seconds.

Calculate the lower bound on response time for both scenarios.

$$R_A = \frac{N \cdot D}{m} - Z = \frac{30 \times 10}{10} - 10 = 20$$
$$R_B = D = 0.1$$

**Question 15**

Based on the system described in question 14, calculate the upper bound on throughput for both scenarios.

Case A:

Due to think time

$$X \leq \frac{N}{D+z} = \frac{30}{10+10+10} = 2.7272$$

Due to hardware bottleneck

$$X \leq \frac{1}{D_{max}} = \frac{1}{1} = 1$$

Case B:

Due to think time:

$$X \leq \frac{h}{D+z} = \frac{30}{0.1+10} = 2.970$$

Due to hardware bottleneck

$$X \leq \frac{1}{D_{max}} = \frac{1}{0.1} = 10$$

Question 16

Considering the system in question 14, compute the number of users N^* that determines whether the system should be analyzed using the light load or heavy load optimistic bounds.



Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ Describe characteristics, advantages, and disadvantages of Direct Attached Storage (DAS), Network Attached Storage (NAS) and Storage Area Networks (SAN).

Question 18

⇒ Describe, comment, and contextualize the four main types of Clouds

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.



Computing Infrastructures - September 5, 2024

Answer Sheets (Page 1)

First Name (CAPITAL LETTERS):

Last Name (CAPITAL LETTERS):

Student ID (Codice Persona):

Question 17

- ⇒ Describe characteristics, advantages, and disadvantages of Direct Attached Storage (DAS), Network Attached Storage (NAS) and Storage Area Networks (SAN).



Computing Infrastructures - September 5, 2024

Answer Sheets (Page 2)

Question 18

⇒ Describe, comment, and contextualize the four main types of Clouds



Computing Infrastructures - September 5, 2024

Answer Sheets (Page 3)

Student ID (Codice Persona):

True/False Questions

Question 01 : A BQuestion 02 : A BQuestion 03 : A BQuestion 04 : A BQuestion 05 : A BQuestion 06 : A BQuestion 07 : A BQuestion 08 : A BQuestion 09 : A BQuestion 10 : A B

Exercises

0.8871

Question 11 :

0.8952

Question 12 :

16008 days

Question 13 :

Question 14 : 20 sec | 0.1 sec

1 R~~Y~~ sec

0.1 sec

 $\frac{100}{sec}$

Question 15 : 2.87 |

..... = 101

Question 16 : N* = 20 |

N* = 20

= 101



Computing Infrastructures

January 13, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points.** An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 SANs are primarily used for block-level access to data, while NAS devices provide file-level access.

A False

B True

Question 2 RAID can be used with both hardware and software implementations.

A False

B True

Question 3 GPUs (Graphics Processing Units) are primarily used in datacenters for accelerating graphical applications and gaming.

A False

B True

Question 4 Cloud computing applications do not require any server to run.

A True

B False

Question 5 Availability zones within a Compute Region can have a roundtrip time larger than few milli seconds.

A False

B True

Question 6 Edge computing is a distributed computing paradigm that brings computation and data storage closer to the location where it is needed.

A False

B True

Question 7 The power consumption of an average datacenter is in the order of a few GigaWatts.

A False

B True

Question 8 Leaf-spine topologies are more suitable for small-scale datacenter deployments with limited server connections.

A False

B True

Question 9 Warehouse-scale computers are typically composed of a large set of homogeneous servers.

A False

B True

Question 10 Virtualization can help simplify IT management by reducing the number of physical machines that need to be managed.

A True

B False



Exercises

Correct answer: +2, No answer: 0.

The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

Question 11

A scientific computation that needs to be carried out within the PoliMi data center uses a server composed of 2 CPUs and 4 GPUs. Knowing that:

- The computation takes 8 days to complete,
- The computation requires at least one CPU and all GPUs within the server to be operational to complete successfully;
- $MTTF_{CPU} = 180$ days and $MTTF_{GPU} = 120$ days.

How many parallel instances of the computation must be launched to ensure a probability higher than 98% that at least one computation produces results successfully? Notes: (i) Use at least 4 decimal places for all intermediate calculations. (ii) All other components of the server can be considered ideal.

$$\begin{aligned}
 R_{CPU}(8) &= e^{-\frac{8}{180}} = 0.9565 \\
 R_{GPU}(8) &= e^{-\frac{8}{120}} = 0.9355 \\
 R_{\text{instance}} &= [1 - (1 - 0.9565)^2] \times 0.9355^4 = 0.7645 \\
 1 - (1 - 0.7645)^n &= 98\% \rightarrow n = 3 \\
 0.2355
 \end{aligned}$$

Question 12

If the time required to recover and replace a component of the server described in the previous exercise (whether it is a CPU or a GPU) is equal to 24 days, what is the total availability of the server? Notes: (i) Use at least 5 decimal places for all intermediate calculations. (ii) All other components of the server can be considered ideal.

$$\begin{aligned}
 A_{CPU} &= \frac{180}{180+24} = 0.88235 \\
 A_{GPU} &= \frac{120}{120+24} = 0.83333 \\
 [1 - (1 - 0.88235)^2] \times 0.83333^4 &= 0.4756
 \end{aligned}$$

**Question 13**

A company is planning to use a RAID 5 array composed of 8 disks for critical data storage. The desired Mean Time to Failure for the entire RAID system ($MTTF_{RAID5}$) is 16 years. Each disk in the array has a $MTTF_{disk}$ equal to 500 days. What should be the MTTR required *in hours* to meet the target RAID $MTTF_{RAID5}$.

Question 14

Consider a system composed of three stations: the CPU that is characterized by $V_{CPU} = 100$ visits and an average service time of $S_{CPU} = 10\text{ms}$; the disk, characterized by a throughput of 12 IOPS, and a demand of $D_{DISK}=150\text{ms}$; and the GPU whose demand is $D_{GPU}=40\text{s}$ and the number of visits $V_{GPU}=10$. Finally, the system throughput is $X=20\text{ jobs/min}$ while the response time when there are $N = 20$ end-users in the system is $R = 55\text{s}$.

Compute the CPU demand and the GPU throughput.

Write in the answer sheet: $D_{CPU} = \dots; X_{GPU} = \dots$

$$D_{CPU} = V_{CPU} \cdot S_{CPU} = 100 \times 10 = 1\text{s}$$

$$D_{GPU} = V_{GPU} \cdot S_{GPU} \rightarrow S_{GPU} = \frac{D_{GPU}}{V_{GPU}} = \frac{40}{10} = 4\text{s}$$

$$X_{GPU} = X \cdot V_{GPU} = 20 \times 10 = 200 / 60 = 3.33\text{s}$$

**Question 15**

Considering the system described in Question 14, what is the users' think time Z?

$$D_{CPU} = V_{CPU} \cdot S_{CPU} = 100 \times 10 = 1s$$

$$D_{GPU} = V_{GPU} \cdot S_{GPU} \rightarrow S_{GPU} = \frac{D_{GPU}}{V_{GPU}} = \frac{40}{10} = 4s$$

$$X_{GPU} = X \cdot V_{GPU} = 20 \times 10 = 200 / 60 = 3.33s$$

$$R = \frac{N}{X} - z \quad 55 = \frac{\cancel{20}/\cancel{60} 20}{20/60} - z \Rightarrow z = 5s$$

Question 16

Based on the system in Question 14, the number of end users is predicted to reach 40 in one month. Considering the response time lower bound, which option is better?

- Upgrade the system by adding one more GPU (you can assume the new GPU is equal to the one initially available and to balance evenly the GPU processing).
- Replace the GPU with one 2.5 times faster than the original one.

Provide the estimated bounds in the two scenarios to motivate your answer.

Write in the answer sheet: A or B ; $R_{LOW}^A = \dots$; $R_{LOW}^B = \dots$

Case A: $40s \rightarrow D_{GPU}' = 20s$

$$D_{GPU} = 40s \rightarrow D_{GPU} = 20s$$

$$R_A = N \cdot D_{GPU}' - z$$

$$\therefore = 40 \times 20 - 5 = 795s$$

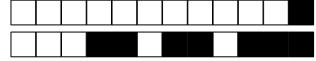
Case B: $D_{GPU}' = \frac{D_{GPU}}{2.5} = 16s$

$$R_B = N \cdot D_{GPU}' - z = (6 \times 40) - 5$$

$$= 635s$$

$\approx 5s$

So Case B is better.



Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ What are the advantages and disadvantages of using Type 1 hypervisors versus Type 2 hypervisors in a virtualized environment?

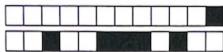
Question 18

⇒ What are the implications of hardware heterogeneity on software stack development in datacenters?

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.

Type 1 : Runs on bare metal machine Advantages : 1. Runs on bare metal machine , which means higher performance 2. The security is strong 3. Use case = enterprise , production 4. Setup complexity : complex
Type 2 : Runs on host operating system Advantages : 1. Easy to install and setup



Computing Infrastructures - January 13, 2025

Answer Sheets (Page 3)

Student ID (Codice Persona):

True/False Questions

Question 01 : A BQuestion 02 : A BQuestion 03 : A BQuestion 04 : A BQuestion 05 : A BQuestion 06 : A BQuestion 07 : A BQuestion 08 : A BQuestion 09 : A BQuestion 10 : A B

Exercises

3

Question 11 :

0.4756 σ 0.8558

Question 12 :

18,35 h (σ 0.7644 giorni)Question 14 : $D_{CPU} = 1 \frac{sec}{sec}$ $X_{GPU} = 200 \frac{sec}{mm} \sigma 3.33 \frac{sec}{sec}$ Question 15 : $Z = 5 \frac{sec}{sec}$ Question 16 : B ; $R_{MIN}^A = 785 \frac{sec}{sec}$; $R_{MIN}^B = 635 \frac{sec}{sec}$



Computing Infrastructures

February 3, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points.** An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 Data centers requires specialized fire suppression systems.

A False

B True

Question 2 In current datacenters, east-west traffic is typically less bandwidth-intensive than north-south traffic.

A False

B True

Question 3 Cloud computing does not require the user to manage the underlying hardware infrastructure, instead relying on the cloud provider to manage it.

A True

B False

Question 4 RAID 6 offers better random read performance than RAID 5.

A True

B False

Question 5 Cloud architectures can help reduce the cost of IT infrastructure by enabling pay-as-you-go pricing.

A True

B False

Question 6 In-rack cooling involves placing cooling units directly in the server racks.

A False

B True

Question 7 RAID 5 has a higher storage overhead than RAID 4.

A True

B False

Question 8 Oversubscription is not used in TOR switches

A False

B True

Question 9 A three-tier network architecture is never used in leaf-spine architectures

A False

B True

Question 10 Raised floor systems in data centers help with cable management and airflow distribution.

A False

B True



Exercises

Correct answer: +2, No answer: 0.

The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

Question 11

Let us consider a set of requests in the disk queue referring to the following cylinders of the disk: 65, 23, 50, 8, 39. Consider the initial position of the disk head at cylinder 52 and it is moving from inside (lower cylinder number) to outside (higher cylinder number). If no further requests arrive, write the order of the served requests (from the first to the last) if the disk head scheduling algorithm adopted is SCAN? Use the cylinder number to refer to the request.

65, 50, 39, 23, 8

Question 12

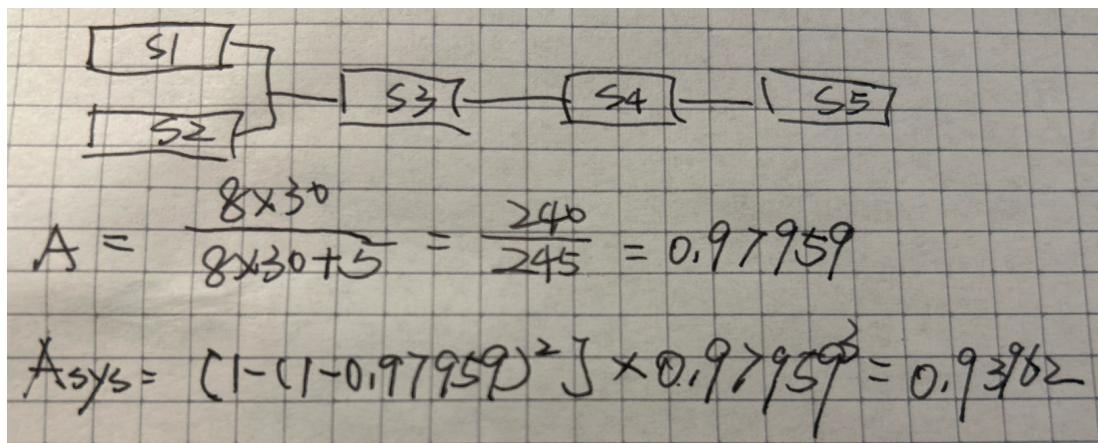
A company uses a computing system composed of three servers: one front-end server and two back-end servers. The system remains operational even if one of the two back-end servers fails. In all other cases, the system is considered offline. All servers have a Mean Time To Failure (MTTF) of 8 months. (i) Compute the probability that at least one of the three servers fails within the first 40 days, and (ii) compute the overall reliability of the computing system after 40 days. Notes: Use at least 4 decimal places for each intermediate calculation.

$$R_F(40) = \frac{1}{e^{\frac{40}{8 \times 30}}} = \frac{1}{e^{\frac{1}{6}}} = 0.846$$
$$R_B(40) = \frac{1}{e^{\frac{40}{8 \times 30}}} = 0.846$$
$$P(\text{at least one fails}) = 1 - 0.846^3 = 0.3945$$
$$R_{\text{sys}} = 0.846 \times [1 - (1 - 0.846)^2] = 0.826$$



Question 13

A company uses a distributed data storage system composed of five storage units, organized in a *primary* and *secondary* storage groups, according to the following scheme: (i) Two redundant *primary* storage units (S_1 and S_2), meaning that the system continues to work as long as at least one of the two remains operational; (ii) Three *secondary* storage units (S_3 , S_4 , and S_5) in a configuration that requires that all three secondary units must be operational; (iii) Both storage groups should be operational to have the data storage system working. Each storage unit has the following parameters: Mean Time To Failure (MTTF) = 8 months; Mean Time To Repair (MTTR) = 5 day. Compute the overall availability of the system. Notes: Use at least 5 decimal places for each intermediate calculation.



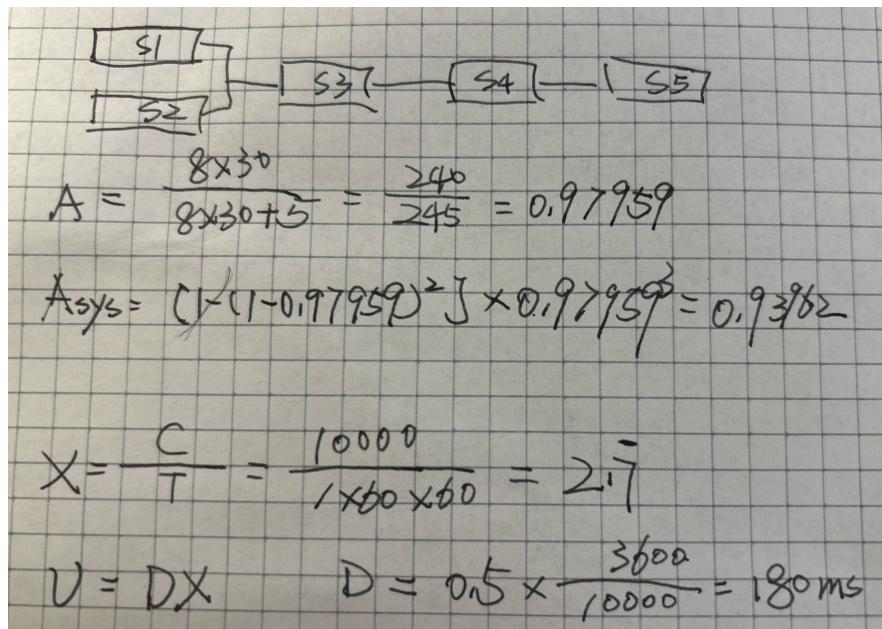
Question 14

You have been assigned the responsibility of optimizing the web server of your company. Through the analysis of your system logs, you have collected the following data (assume that this data is gathered during the peak hour, i.e., $T=1\text{h}$, on a representative business day).

Number of completed requests $C=10000$

Average system utilization $U=0.5$

What is the system throughput X and the requests service demand D ?





Question 15

Considering the system described in Question 14, what is the maximum throughput that the system can achieve?

$$1 / 0.18 = 5.5$$

Question 16

Given the system described in Question 14, assume an expected peak workload growth of 4x (i.e. four times more requests within the same time interval). Determine the minimum number of additional servers required to ensure that the system utilization does not exceed 60%. Assume that all new servers have the same configuration as the existing ones and that the incoming requests are evenly distributed across all servers. The answer should include the *total number of servers* required, including the original one.

$$\lambda_{sys} = (1 - (1 - 0.97959)^2) \times 0.97959^3 = 0.93962$$

$$X = \frac{C}{T} = \frac{10000}{1 \times 60 \times 60} = 2.7$$

$$U = DX \quad D = 0.5 \times \frac{3600}{10000} = 180 \text{ ms}$$

New throughput $X' = 4 \times 10000 = 40000 \text{ requests/hour}$

$$U_{\text{per server}} = U' = D - \frac{X'}{N} \leq 0.6$$

$$\Rightarrow N = 4$$



Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ Explain the concept of Wear Leveling in the context of SSD.

Question 18

⇒ Discuss the concept of Datacenter Availability, and define what are the differences among the Tier Levels established by the Uptime Institute.

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.



Computing Infrastructures - February 3, 2025

Answer Sheets (Page 3)

SOL.

Student ID (Codice Persona):

True/False Questions

Question 01 : A BQuestion 02 : A BQuestion 03 : A BQuestion 04 : A BQuestion 05 : A BQuestion 06 : A BQuestion 07 : A BQuestion 08 : A BQuestion 09 : A BQuestion 10 : A B

Exercises

65-50-38-23-8

Question 11 :

0.3335 - 0.8265

Question 12 :

0.83862

Question 13 :

 $X = 2.7 \frac{R_{EQ}}{S}$ $D = 180 \text{ ms}$ **$X_{MAX} = 5.5 \frac{R_{EQ}}{S_{BC}}$** **$N=4$**

Question 16 :



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

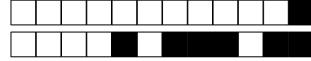
Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points.** An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 Cloud architectures can be deployed on-premises, in a public cloud, or in a hybrid environment.

A True

B False

Question 2 TPUs require specialized software libraries and frameworks to fully utilize their capabilities.

A False

B True

Question 3 Virtualization is only suitable for running non-critical applications and workloads.

A True

B False

Question 4 Blade servers universally reduce hardware expenses, ensuring they cost far less than rack servers in every scenario.

A False

B True

Question 5 FPGAs are widely adopted for all large-scale HPC and AI workloads, replacing GPUs in most scenarios.

A True

B False

Question 6 Edge computing nodes often cache and compute on locally produced data to reduce latency and congestion.

A False

B True

Question 7 Data centers require specialized fire suppression systems.

A False

B True

Question 8 DAS storage provides low throughput due to its reliance on a shared network.

A True

B False

Question 9 Direct open-loop cooling is most suitable for hot, humid climates.

A False

B True

Question 10 RAID 5 uses parity data to provide fault tolerance.

A True

B False



Exercises

Correct answer: +2, No answer: 0.

The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

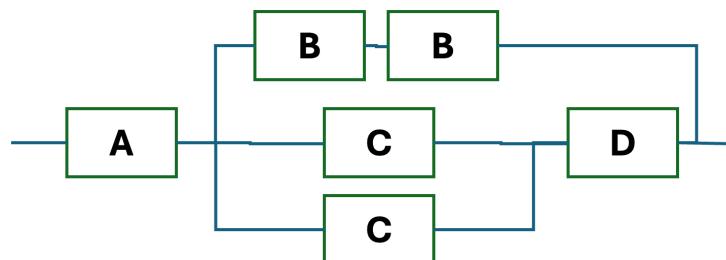
Question 11

Let us consider a set of requests in the disk queue referring to the following cylinders of the disk: 65, 23, 58, 48, 39. Consider the initial position of the disk head at cylinder 53 and it is moving from outside (higher cylinder number) to inside (lower cylinder number). If no further requests arrive, write the order of the served requests (from the first to the last) if the disk head scheduling algorithm adopted is SCAN? Use the cylinder number to refer to the request.

48, 39, 23, 58, 65

Question 12

Suppose we have a computer system composed of 6 different components, and designed to have an RBD as shown in the image below. The four types of components (A, B, C, and D) have different reliability characteristics. We know that the availability of the components B, C, and D are respectively $Av_A = Av_B = 0.8$, $Av_C = 0.7$, and $Av_D = 0.9$. What is the availability of the entire system? Use always at least 4 decimal digits for each calculation.



$$A_{\text{sequential } B} = Av_B \times Av_B = 0.8 \times 0.8 = 0.64$$

$$A_{\text{parallel } C} = 1 - (1 - 0.7)(1 - 0.7) = 1 - 0.09 = 0.91$$

$$A_{\text{sequential } CD} = 0.91 \times Av_D = 0.91 \times 0.9 = 0.819$$

$$0.8 \times [1 - (1 - \underline{0.64}) \cdot (1 - 0.819)] = 0.748$$



Question 13

A scientific computation uses a server composed of 2 CPUs and 4 GPUs. Knowing that the $MTTF_{CPU} = 380\text{days}$ and $MTTF_{GPU} = 260\text{days}$, and the computation to work requires both CPUs and one GPU within the server to work properly. What is the reliability value after 1/2 years, $R(0.5\text{y})$? Notes: (i) Use at least 4 decimal digits for all the intermediate calculations; (ii) All the other components within the server can be considered as ideal.

$$\begin{aligned}
 & V_{S2} = D_{S2}/X \rightarrow D_{S2} = 0.6668 \\
 & R_{CPU} = e^{-\frac{0.5 \times 365}{380}} \approx 0.6186 \\
 & R_{GPU} = e^{-\frac{0.5 \times 365}{260}} \approx 0.4955 \\
 & R_{System} = R_{CPU} \cdot R_{GPU} = 0.6186^2 \times [1 - (1 - 0.4955)^4] \\
 & = 0.3578 \approx 0.358
 \end{aligned}$$

Question 14

A video rendering system consists of three components: a GPU Server (GS), which processes rendering tasks, a Model Cache Server (MCS) which manages 3D assets, and a Frame Buffer Server (FBS) which handles output frames. The main data obtained from the logging system are reported below:

Service time: $S_{GS} = 20\text{s}$, Visits: $V_{GS} = 21$ visits

Service time: $S_{MCS} = 40\text{s}$, Visits: $V_{MCS} = 12$ visits

Service time: $S_{FBS} = 10\text{s}$, Visits: $V_{FBS} = 80$ visits

Additionally, the system serves $N=5$ users characterized by a think time $Z = 600\text{s}$

What is the system bottleneck (i.e. GS, MCS or FBS)?

FBS



Question 15

Considering the system described in Question 14, compute: a) the maximum system throughput in *jobs/min*, b) the minimum response time in *minutes*.

Question 16

Given the system described in Question 14, you now have the opportunity to enhance its performance by adding exactly one additional server. However, you can only duplicate one of the three existing servers: GS, MCS or FBS. The new server will be identical to the other of the same type (homogeneous), and you can distribute the workload (visits) evenly between them. Assume that all the other system monitoring metrics remain unchanged. Answer the following: a) which one do you choose? (i.e. GS, MCS, FBS) b) What will be the new minimum response time in *minutes*?

$$\begin{aligned}
 D_{GS} &= S_{GS} \cdot V_{GS} = 20 \times 21 = 420 \\
 D_{MCS} &= S_{MCS} \cdot V_{MCS} = 40 \times 12 = 480 \\
 D_{FBS} &= S_{FBS} \cdot V_{FBS} = 10 \times 80 = 800 \\
 \text{Due to think time:} \\
 X &= \frac{5}{D + x} = 0.002 \\
 \text{Due to device bottleneck} \\
 X &= \frac{1}{D_{max}} = \frac{1}{800} = 0.00125 \times 60 = 0.075 \\
 R(N) &\geq \max(D, N \times D_{max} - x) = 56.67 \\
 \text{Duplicate FBC} \\
 \downarrow \\
 D_{max} &= 480 \rightarrow \max(D, N \times D_{max} - x) = 30 \text{ min}
 \end{aligned}$$



Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 1)

First Name (CAPITAL LETTERS):

Last Name (CAPITAL LETTERS):

Student ID (Codice Persona):

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?

From compute infrastructure perspective,

Hardware:

GPU: Use Nvidia GPU like A100

TPU: Use TPU for tensor computation, but we can only use it as cloud service

CPU: Use high throughput CPU for I/O, multiple core CPU

Interconnects:

NVLink: high bandwidth for GP-GPU communication

PCIe: Connection between CPU, GPU, memory to reduce latency

Storage:

NVMe SSDs: For training dataset or checkpoint because of its high performance

HDD array: ~~SSD~~ storage less-frequently accessed data

Network:

Spine-leaf networking = high-bandwidth, redundant design.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 2)

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.

CHOOSE RAID 1+0:

① The application is write-intensive. Because RAID 1+0 has more powerful write performance than RAID 5

② You can afford to trade storage efficiency for speed. Because the storage efficiency is 50%, lower than RAID 5's

③ Need high fault tolerance, RAID 1+0 can tolerate multiple drive failures.

CHOOSE RAID 5:

① The workload is primarily read-intensive

② The budget is limited, because the storage efficiency is $\frac{N-1}{N}$



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 3)

Student ID (Codice Persona):

True/False Questions

Question 01 : A B

Question 02 : A B

Question 03 : A B

Question 04 : A B

Question 05 : A B

Question 06 : A B

Question 07 : A B

Question 08 : A B

Question 09 : A B

Question 10 : A B

Exercises

48 39 23 58 65

Question 11 :

0,748

Question 12 :

0,358

Question 13 :

FBS

Question 14 :

0,075 Jobs/min. 56,667 min

Question 15 :

FBS. 30min

Question 16 :



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

⇒ The remaining part of this page has been intentionally left blank ←

If needed, you can use this page for notes. Any answer written here will be ignored.



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

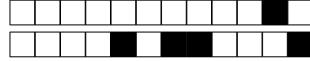
Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points.** An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 Direct open-loop cooling is most suitable for hot, humid climates.

- A False B True

Question 2 Edge computing nodes often cache and compute on locally produced data to reduce latency and congestion.

- A True B False

Question 3 TPUs require specialized software libraries and frameworks to fully utilize their capabilities.

- A True B False

Question 4 Cloud architectures can be deployed on-premises, in a public cloud, or in a hybrid environment.

- A True B False

Question 5 Data centers require specialized fire suppression systems.

- A True B False

Question 6 FPGAs are widely adopted for all large-scale HPC and AI workloads, replacing GPUs in most scenarios.

- A True B False

Question 7 RAID 5 uses parity data to provide fault tolerance.

- A True B False

Question 8 DAS storage provides low throughput due to its reliance on a shared network.

- A False B True

Question 9 Blade servers universally reduce hardware expenses, ensuring they cost far less than rack servers in every scenario.

- A True B False

Question 10 Virtualization is only suitable for running non-critical applications and workloads.

- A True B False



Exercises

Correct answer: +2, No answer: 0.

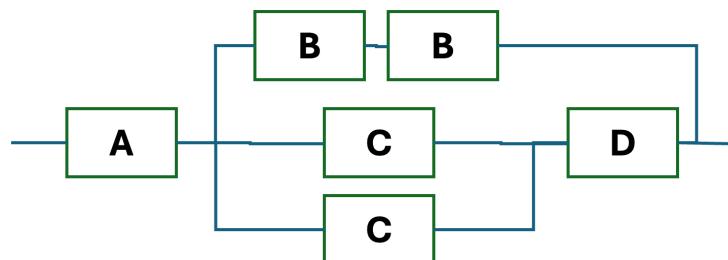
The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

Question 11

Let us consider a set of requests in the disk queue referring to the following cylinders of the disk: 65, 23, 58, 48, 39. Consider the initial position of the disk head at cylinder 53 and it is moving from outside (higher cylinder number) to inside (lower cylinder number). If no further requests arrive, write the order of the served requests (from the first to the last) if the disk head scheduling algorithm adopted is C-SCAN? Use the cylinder number to refer to the request.

Question 12

Suppose we have a computer system composed of 6 different components, and designed to have an RBD as shown in the image below. The four types of components (A, B, C, and D) have different reliability characteristics. We know that the availability of the components B, C, and D are respectively $Av_A = Av_B = 0.8$, $Av_C = 0.7$, and $Av_D = 0.9$. What is the availability of the entire system? Use always at least 4 decimal digits for each calculation.



**Question 13**

A scientific computation uses a server composed of 2 CPUs and 4 GPUs. Knowing that the $MTTF_{CPU} = 380\text{days}$ and $MTTF_{GPU} = 260\text{days}$, and the computation to work requires both CPUs and one GPU within the server to work properly. What is the reliability value after 1/2 years, $R(0.5y)$? Notes: (i) Use at least 4 decimal digits for all the intermediate calculations; (ii) All the other components within the server can be considered as ideal.

Question 14

A video rendering system consists of three components: a GPU Server (GS), which processes rendering tasks, a Model Cache Server (MCS) which manages 3D assets, and a Frame Buffer Server (FBS) which handles output frames. The main data obtained from the logging system are reported below:

Service time: $S_{GS} = 20\text{s}$, Visits: $V_{GS} = 21$ visits

Service time: $S_{MCS} = 40\text{s}$, Visits: $V_{MCS} = 14$ visits

Service time: $S_{FBS} = 10\text{s}$, Visits: $V_{FBS} = 80$ visits

Additionally, the system serves $N=7$ users characterized by a think time $Z = 600\text{s}$

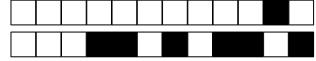
What is the system bottleneck (i.e. GS, MCS or FBS)?

**Question 15**

Considering the system described in Question 14, compute: a) the maximum system throughput in *jobs/min*, b) the minimum response time in *minutes*.

Question 16

Given the system described in Question 14, you now have the opportunity to enhance its performance by adding exactly one additional server. However, you can only duplicate one of the three existing servers: GS, MCS or FBS. The new server will be identical to the other of the same type (homogeneous), and you can distribute the workload (visits) evenly between them. Assume that all the other system monitoring metrics remain unchanged. Answer the following: a) which one do you choose? (i.e. GS, MCS, FBS) b) What will be the new minimum response time in *minutes*?



Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 1)

First Name (CAPITAL LETTERS):

Last Name (CAPITAL LETTERS):

Student ID (Codice Persona):

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 2)

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 3)

Student ID (Codice Persona):

True/False Questions

Question 01 : A B

Question 02 : A B

Question 03 : A B

Question 04 : A B

Question 05 : A B

Question 06 : A B

Question 07 : A B

Question 08 : A B

Question 09 : A B

Question 10 : A B

Exercises

48 39 23 65 58 (accepted also 58 65 23 39 48)

Question 11 :

0,748

Question 12 :

0,358

Question 13 :

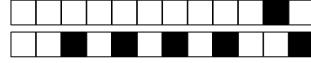
FBS

Question 14 :

0,075 Job/min. 83,333 min

Question 15 :

Question 16 : ...**FBS. 55,333 min.**.....



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

⇒ The remaining part of this page has been intentionally left blank ←

If needed, you can use this page for notes. Any answer written here will be ignored.



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

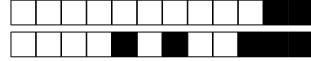
Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points.** An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 Edge computing nodes often cache and compute on locally produced data to reduce latency and congestion.

A False

B True

Question 2 FPGAs are widely adopted for all large-scale HPC and AI workloads, replacing GPUs in most scenarios.

A True

B False

Question 3 Virtualization is only suitable for running non-critical applications and workloads.

A True

B False

Question 4 DAS storage provides low throughput due to its reliance on a shared network.

A False

B True

Question 5 RAID 5 uses parity data to provide fault tolerance.

A False

B True

Question 6 Data centers require specialized fire suppression systems.

A True

B False

Question 7 Cloud architectures can be deployed on-premises, in a public cloud, or in a hybrid environment.

A False

B True

Question 8 TPUs require specialized software libraries and frameworks to fully utilize their capabilities.

A False

B True

Question 9 Direct open-loop cooling is most suitable for hot, humid climates.

A False

B True

Question 10 Blade servers universally reduce hardware expenses, ensuring they cost far less than rack servers in every scenario.

A False

B True



Exercises

Correct answer: +2, No answer: 0.

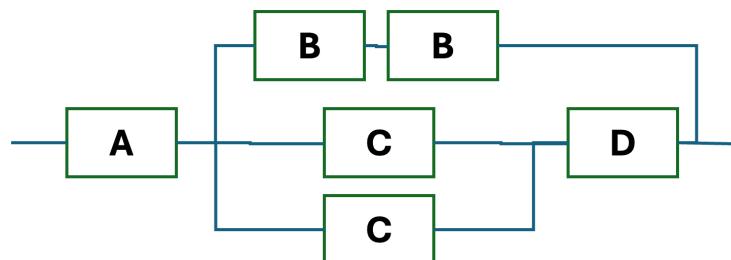
The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

Question 11

Let us consider a set of requests in the disk queue referring to the following cylinders of the disk: 65, 23, 58, 48, 39. Consider the initial position of the disk head at cylinder 52 and it is moving from inside (lower cylinder number) to outside (higher cylinder number). If no further requests arrive, write the order of the served requests (from the first to the last) if the disk head scheduling algorithm adopted is C-SCAN? Use the cylinder number to refer to the request.

Question 12

Suppose we have a computer system composed of 6 different components, and designed to have an RBD as shown in the image below. The four types of components (A, B, C, and D) have different reliability characteristics. We know that the availability of the components B, C, and D are respectively $Av_A = 0.8$, $Av_B = Av_C = 0.6$, and $Av_D = 0.7$. What is the availability of the entire system? Use always at least 4 decimal digits for each calculation.



**Question 13**

A scientific computation uses a server composed of 2 CPUs and 4 GPUs. Knowing that the $MTTF_{CPU} = 380\text{days}$ and $MTTF_{GPU} = 260\text{days}$, and the computation to work requires both CPUs and one GPU within the server to work properly. What is the reliability value after $1/2$ years, $R(0.5y)$? Notes: (i) Use at least 4 decimal digits for all the intermediate calculations; (ii) All the other components within the server can be considered as ideal.

Question 14

A video rendering system consists of three components: a GPU Server (GS), which processes rendering tasks, a Model Cache Server (MCS) which manages 3D assets, and a Frame Buffer Server (FBS) which handles output frames. The main data obtained from the logging system are reported below:

Service time: $S_{GS} = 20\text{s}$, Visits: $V_{GS} = 21$ visits

Service time: $S_{MCS} = 40\text{s}$, Visits: $V_{MCS} = 12$ visits

Service time: $S_{FBS} = 10\text{s}$, Visits: $V_{FBS} = 80$ visits

Additionally, the system serves $N=10$ users characterized by a think time $Z = 600\text{s}$

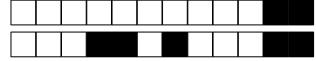
What is the system bottleneck (i.e. GS, MCS or FBS)?

**Question 15**

Considering the system described in Question 14, compute: a) the maximum system throughput in *jobs/min*, b) the minimum response time in *minutes*.

Question 16

Given the system described in Question 14, you now have the opportunity to enhance its performance by adding exactly one additional server. However, you can only duplicate one of the three existing servers: GS, MCS or FBS. The new server will be identical to the other of the same type (homogeneous), and you can distribute the workload (visits) evenly between them. Assume that all the other system monitoring metrics remain unchanged. Answer the following: a) which one do you choose? (i.e. GS, MCS, FBS) b) What will be the new minimum response time in *minutes*?



Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 1)

First Name (CAPITAL LETTERS):

Last Name (CAPITAL LETTERS):

Student ID (Codice Persona):

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 2)

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 3)

Student ID (Codice Persona):

True/False Questions

Question 01 : A B

Question 02 : A B

Question 03 : A B

Question 04 : A B

Question 05 : A B

Question 06 : A B

Question 07 : A B

Question 08 : A B

Question 09 : A B

Question 10 : A B

Exercises

58 65 23 39 48

Question 11 :

0,5891

Question 12 :

0,358

Question 13 :

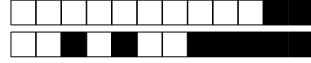
FBS

Question 14 :

0,075 Job/min. 123,333 min

Question 15 :

Question 16 : ...**FBS.**...**70 min**.....



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

⇒ The remaining part of this page has been intentionally left blank ←

If needed, you can use this page for notes. Any answer written here will be ignored.



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points**. An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 Data centers require specialized fire suppression systems.

- A True B False

Question 2 Edge computing nodes often cache and compute on locally produced data to reduce latency and congestion.

- A True B False

Question 3 RAID 5 uses parity data to provide fault tolerance.

- A True B False

Question 4 DAS storage provides low throughput due to its reliance on a shared network.

- A True B False

Question 5 Direct open-loop cooling is most suitable for hot, humid climates.

- A False B True

Question 6 FPGAs are widely adopted for all large-scale HPC and AI workloads, replacing GPUs in most scenarios.

- A True B False

Question 7 Blade servers universally reduce hardware expenses, ensuring they cost far less than rack servers in every scenario.

- A False B True

Question 8 Virtualization is only suitable for running non-critical applications and workloads.

- A True B False

Question 9 TPUs require specialized software libraries and frameworks to fully utilize their capabilities.

- A True B False

Question 10 Cloud architectures can be deployed on-premises, in a public cloud, or in a hybrid environment.

- A False B True



Exercises

Correct answer: +2, No answer: 0.

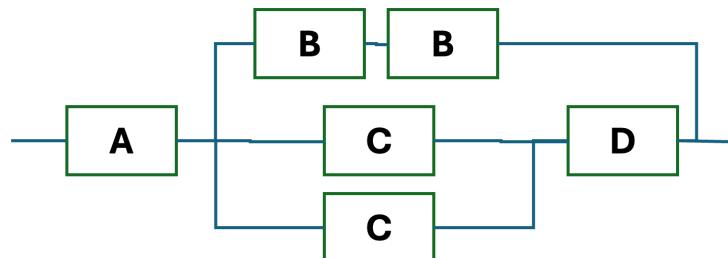
The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

Question 11

Let us consider a set of requests in the disk queue referring to the following cylinders of the disk: 65, 23, 58, 48, 39. Consider the initial position of the disk head at cylinder 53 and it is moving from outside (higher cylinder number) to inside (lower cylinder number). If no further requests arrive, write the order of the served requests (from the first to the last) if the disk head scheduling algorithm adopted is C-SCAN? Use the cylinder number to refer to the request.

Question 12

Suppose we have a computer system composed of 6 different components, and designed to have an RBD as shown in the image below. The four types of components (A, B, C, and D) have different reliability characteristics. We know that the availability of the components B, C, and D are respectively $Av_A = Av_B = 0.8$, $Av_C = 0.7$, and $Av_D = 0.9$. What is the availability of the entire system? Use always at least 4 decimal digits for each calculation.



**Question 13**

A scientific computation uses a server composed of 2 CPUs and 4 GPUs. Knowing that the $MTTF_{CPU} = 380\text{days}$ and $MTTF_{GPU} = 260\text{days}$, and the computation to work requires both CPUs and one GPU within the server to work properly. What is the reliability value after 1/2 years, $R(0.5y)$? Notes: (i) Use at least 4 decimal digits for all the intermediate calculations; (ii) All the other components within the server can be considered as ideal.

Question 14

A video rendering system consists of three components: a GPU Server (GS), which processes rendering tasks, a Model Cache Server (MCS) which manages 3D assets, and a Frame Buffer Server (FBS) which handles output frames. The main data obtained from the logging system are reported below:

Service time: $S_{GS} = 20\text{s}$, Visits: $V_{GS} = 21$ visits

Service time: $S_{MCS} = 40\text{s}$, Visits: $V_{MCS} = 12$ visits

Service time: $S_{FBS} = 10\text{s}$, Visits: $V_{FBS} = 80$ visits

Additionally, the system serves $N=5$ users characterized by a think time $Z = 600\text{s}$

What is the system bottleneck (i.e. GS, MCS or FBS)?

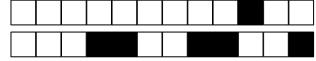


Question 15

Considering the system described in Question 14, compute: a) the maximum system throughput in *jobs/min*, b) the minimum response time in *minutes*.

Question 16

Given the system described in Question 14, you now have the opportunity to enhance its performance by adding exactly one additional server. However, you can only duplicate one of the three existing servers: GS, MCS or FBS. The new server will be identical to the other of the same type (homogeneous), and you can distribute the workload (visits) evenly between them. Assume that all the other system monitoring metrics remain unchanged. Answer the following: a) which one do you choose? (i.e. GS, MCS, FBS) b) What will be the new minimum response time in *minutes*?



Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 1)

First Name (CAPITAL LETTERS):

Last Name (CAPITAL LETTERS):

Student ID (Codice Persona):

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 2)

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 3)

Student ID (Codice Persona):

True/False Questions

Question 01 : A B

Question 02 : A B

Question 03 : A B

Question 04 : A B

Question 05 : A B

Question 06 : A B

Question 07 : A B

Question 08 : A B

Question 09 : A B

Question 10 : A B

Exercises

48 39 23 65 58 (accepted also 58 65 23 39 48)

Question 11 :

0,748

Question 12 :

0,358

Question 13 :

FBS

Question 14 :

0,075 Job/min. 56,667 min

Question 15 :

Question 16 : ...**FBS..30 min**.....



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

⇒ The remaining part of this page has been intentionally left blank ←

If needed, you can use this page for notes. Any answer written here will be ignored.



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points.** An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 FPGAs are widely adopted for all large-scale HPC and AI workloads, replacing GPUs in most scenarios.

A False

B True

Question 2 Blade servers universally reduce hardware expenses, ensuring they cost far less than rack servers in every scenario.

A False

B True

Question 3 Edge computing nodes often cache and compute on locally produced data to reduce latency and congestion.

A True

B False

Question 4 Direct open-loop cooling is most suitable for hot, humid climates.

A True

B False

Question 5 DAS storage provides low throughput due to its reliance on a shared network.

A True

B False

Question 6 Cloud architectures can be deployed on-premises, in a public cloud, or in a hybrid environment.

A True

B False

Question 7 RAID 5 uses parity data to provide fault tolerance.

A True

B False

Question 8 Data centers require specialized fire suppression systems.

A False

B True

Question 9 TPUs require specialized software libraries and frameworks to fully utilize their capabilities.

A False

B True

Question 10 Virtualization is only suitable for running non-critical applications and workloads.

A False

B True



Exercises

Correct answer: +2, No answer: 0.

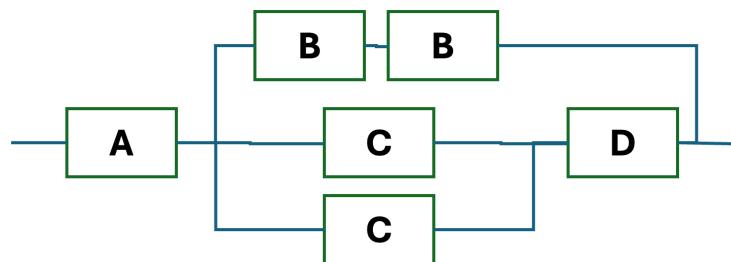
The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

Question 11

Let us consider a set of requests in the disk queue referring to the following cylinders of the disk: 65, 23, 58, 48, 39. Consider the initial position of the disk head at cylinder 53 and it is moving from outside (higher cylinder number) to inside (lower cylinder number). If no further requests arrive, write the order of the served requests (from the first to the last) if the disk head scheduling algorithm adopted is C-SCAN? Use the cylinder number to refer to the request.

Question 12

Suppose we have a computer system composed of 6 different components, and designed to have an RBD as shown in the image below. The four types of components (A, B, C, and D) have different reliability characteristics. We know that the availability of the components B, C, and D are respectively $Av_A = Av_B = 0.8$, $Av_C = 0.7$, and $Av_D = 0.9$. What is the availability of the entire system? Use always at least 4 decimal digits for each calculation.



**Question 13**

A scientific computation uses a server composed of 2 CPUs and 4 GPUs. Knowing that the $MTTF_{CPU} = 380\text{days}$ and $MTTF_{GPU} = 260\text{days}$, and the computation to work requires both CPUs and one GPU within the server to work properly. What is the reliability value after 1/2 years, $R(0.5y)$? Notes: (i) Use at least 4 decimal digits for all the intermediate calculations; (ii) All the other components within the server can be considered as ideal.

Question 14

A video rendering system consists of three components: a GPU Server (GS), which processes rendering tasks, a Model Cache Server (MCS) which manages 3D assets, and a Frame Buffer Server (FBS) which handles output frames. The main data obtained from the logging system are reported below:

Service time: $S_{GS} = 20\text{s}$, Visits: $V_{GS} = 21$ visits

Service time: $S_{MCS} = 40\text{s}$, Visits: $V_{MCS} = 14$ visits

Service time: $S_{FBS} = 10\text{s}$, Visits: $V_{FBS} = 80$ visits

Additionally, the system serves $N=7$ users characterized by a think time $Z = 600\text{s}$

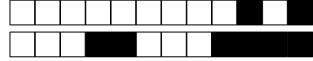
What is the system bottleneck (i.e. GS, MCS or FBS)?

**Question 15**

Considering the system described in Question 14, compute: a) the maximum system throughput in *jobs/min*, b) the minimum response time in *minutes*.

Question 16

Given the system described in Question 14, you now have the opportunity to enhance its performance by adding exactly one additional server. However, you can only duplicate one of the three existing servers: GS, MCS or FBS. The new server will be identical to the other of the same type (homogeneous), and you can distribute the workload (visits) evenly between them. Assume that all the other system monitoring metrics remain unchanged. Answer the following: a) which one do you choose? (i.e. GS, MCS, FBS) b) What will be the new minimum response time in *minutes*?



Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 1)

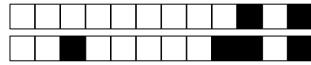
First Name (CAPITAL LETTERS):

Last Name (CAPITAL LETTERS):

Student ID (Codice Persona):

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 2)

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 3)

Student ID (Codice Persona):

True/False Questions

Question 01 : A B

Question 02 : A B

Question 03 : A B

Question 04 : A B

Question 05 : A B

Question 06 : A B

Question 07 : A B

Question 08 : A B

Question 09 : A B

Question 10 : A B

Exercises

48 39 23 65 58 (accepted also 58 65 23 39 48)

Question 11 :

0,748

Question 12 :

0,358

Question 13 :

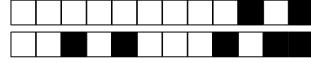
FBS

Question 14 :

0,075 Job/min. 83,333 min

Question 15 :

Question 16 : ...**FBS. 55,333 min.**.....



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

⇒ The remaining part of this page has been intentionally left blank ←

If needed, you can use this page for notes. Any answer written here will be ignored.



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

Exam Duration: 1hour and 30min

Students are not permitted to use mobile phones and similar connected devices. Course materials and programmable devices (e.g. programmable calculators) cannot be used as well. **Any violation of the rules is considered a cheating action.**

Answers must be given on the Answer Sheets and in English. Any box filled or answer provided on the other sheets will be ignored. Students must use a pen (black or blue) to mark the answers (no pencil).

Write the LAST and FIRST name in CAPITAL LETTER, and in this order, in all places where requested. **Where it is requested only the STUDENT ID (Codice Persona), do not write your name.**

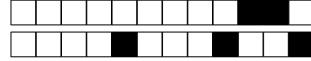
Check that the first number of the code for the Answer Sheet is the same as for the other sheets. The code can be found in the top-right corner of each page in the form +NN/KK/XX+. The parts that should correspond is ONLY the first digit NN.

Mark clearly the box corresponding to your answers, without overlapping on other boxes. If you make a mistake on them, circle the word *Question* together with the related number, and write the correct letter to its side.

Numerical exercises require writing the formulas and procedure used to solve the problem just after the question in the left space. Exercises without the procedure used to reach the result will not be considered for the evaluation. Only the numeric answer and its unit should be reported on the corresponding dotted line in the Answer Sheet.

The answers to the *Open Questions* should be written using ONLY the space available on in the boxes within the Answer Sheets. The answers should be readable by the professor. Unreadable answers will not be considered for the evaluation.

Scores: correct answers take positive points, unanswered questions take 0 points, **wrong answers can have negative points.** An indication of the points is available at the beginning of each section. The final score can be re-modulated at the end of the evaluation.

**True false questions**

Correct answer: +1, No answer: 0, Wrong Answer -0.5

Answers must be given on the ANSWER SHEETS. Any box filled here will be ignored. Pay attention to the position (A or B) of the True/False answers, since they are not always in the same position.

Question 1 Blade servers universally reduce hardware expenses, ensuring they cost far less than rack servers in every scenario.

[A] False

[B] True

Question 2 TPUs require specialized software libraries and frameworks to fully utilize their capabilities.

[A] True

[B] False

Question 3 FPGAs are widely adopted for all large-scale HPC and AI workloads, replacing GPUs in most scenarios.

[A] True

[B] False

Question 4 Data centers requires specialized fire suppression systems.

[A] False

[B] True

Question 5 RAID 5 uses parity data to provide fault tolerance.

[A] False

[B] True

Question 6 Direct open-loop cooling is most suitable for hot, humid climates.

[A] True

[B] False

Question 7 Virtualization is only suitable for running non-critical applications and workloads.

[A] True

[B] False

Question 8 DAS storage provides low throughput due to its reliance on a shared network.

[A] False

[B] True

Question 9 Edge computing nodes often cache and compute on locally produced data to reduce latency and congestion.

[A] True

[B] False

Question 10 Cloud architectures can be deployed on-premises, in a public cloud, or in a hybrid environment.

[A] False

[B] True



+6/3/8+

Exercises

Correct answer: +2, No answer: 0.

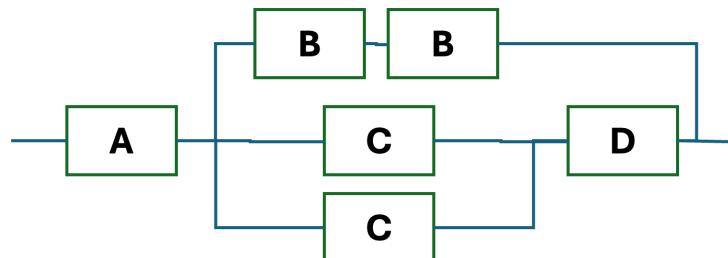
The formulas and procedures used to solve the exercises should be included here close to the question. The numeric answer, and only that, must be given on the ANSWER SHEETS. Any number written only here will be ignored. The correct number is ONLY a necessary condition for a correct answer. If the formulas are not available after each exercise, they will be considered as not answered.

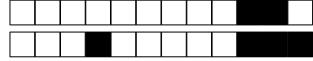
Question 11

Let us consider a set of requests in the disk queue referring to the following cylinders of the disk: 65, 23, 58, 48, 39. Consider the initial position of the disk head at cylinder 52 and it is moving from inside (lower cylinder number) to outside (higher cylinder number). If no further requests arrive, write the order of the served requests (from the first to the last) if the disk head scheduling algorithm adopted is C-SCAN? Use the cylinder number to refer to the request.

Question 12

Suppose we have a computer system composed of 6 different components, and designed to have an RBD as shown in the image below. The four types of components (A, B, C, and D) have different reliability characteristics. We know that the availability of the components B, C, and D are respectively $Av_A = Av_B = 0.8$, $Av_C = 0.7$, and $Av_D = 0.9$. What is the availability of the entire system? Use always at least 4 decimal digits for each calculation.





+6/4/7+

Question 13

A scientific computation uses a server composed of 2 CPUs and 4 GPUs. Knowing that the $MTTF_{CPU} = 380\text{days}$ and $MTTF_{GPU} = 260\text{days}$, and the computation to work requires both CPUs and one GPU within the server to work properly. What is the reliability value after 1/2 years, $R(0.5y)$? Notes: (i) Use at least 4 decimal digits for all the intermediate calculations; (ii) All the other components within the server can be considered as ideal.

Question 14

A video rendering system consists of three components: a GPU Server (GS), which processes rendering tasks, a Model Cache Server (MCS) which manages 3D assets, and a Frame Buffer Server (FBS) which handles output frames. The main data obtained from the logging system are reported below:

Service time: $S_{GS} = 20\text{s}$, Visits: $V_{GS} = 21$ visits

Service time: $S_{MCS} = 40\text{s}$, Visits: $V_{MCS} = 12$ visits

Service time: $S_{FBS} = 10\text{s}$, Visits: $V_{FBS} = 80$ visits

Additionally, the system serves $N=5$ users characterized by a think time $Z = 600\text{s}$

What is the system bottleneck (i.e. GS, MCS or FBS)?



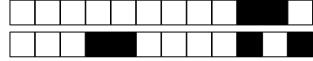
+6/5/6+

Question 15

Considering the system described in Question 14, compute: a) the maximum system throughput in *jobs/min*, b) the minimum response time in *minutes*.

Question 16

Given the system described in Question 14, you now have the opportunity to enhance its performance by adding exactly one additional server. However, you can only duplicate one of the three existing servers: GS, MCS or FBS. The new server will be identical to the other of the same type (homogeneous), and you can distribute the workload (visits) evenly between them. Assume that all the other system monitoring metrics remain unchanged. Answer the following: a) which one do you choose? (i.e. GS, MCS, FBS) b) What will be the new minimum response time in *minutes*?



+6/6/5+

Open Questions

Correct answer: +5, No answer: 0. Points are modulated considering the written text

Write the answer using ONLY the space available in the boxes on the ANSWER SHEETS. The answers should be readable by the professor. Unreadable answers will be considered wrong.

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.

!!!ANY ANSWER PROVIDED ON THIS PAGE WILL BE IGNORED!!!

If needed, you can use the space hereafter to organize your answer.



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 1)

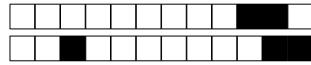
First Name (CAPITAL LETTERS):

Last Name (CAPITAL LETTERS):

Student ID (Codice Persona):

Question 17

⇒ Deep learning is transforming data-center technology. From a technological standpoint, how would you design a data center purpose-built for deep-learning workloads?



Computing Infrastructures - June 12, 2025

Answer Sheets (Page 2)

Question 18

⇒ In a data-center storage context, focusing exclusively on write performance, under what circumstances would you choose RAID 1+0 and under what circumstances RAID 5? Please explain your reasoning.



+6/9/2+

Computing Infrastructures - June 12, 2025

Answer Sheets (Page 3)

Student ID (Codice Persona):

True/False Questions

Question 01 : A B

Question 02 : A B

Question 03 : A B

Question 04 : A B

Question 05 : A B

Question 06 : A B

Question 07 : A B

Question 08 : A B

Question 09 : A B

Question 10 : A B

Exercises

58 65 23 39 48

Question 11 :

0,748

Question 12 :

0,358

Question 13 :

FBS

Question 14 :

0,075 Job/min. 56,667 min

Question 15 :

Question 16 : ...**FBS.** **30 min**.....



Computing Infrastructures

June 12, 2025

Course Section: Prof. Ardagna Prof. Palermo Prof. Roveri

Student ID (Codice Persona):

Last Name:
(LAST NAME IN CAPITAL LETTERS)

First Name:
(FIRST NAME IN CAPITAL LETTERS)

⇒ The remaining part of this page has been intentionally left blank ←

If needed, you can use this page for notes. Any answer written here will be ignored.