

TRACCIA 1

Question 1 (9 points)

One of the Principal Investigators of the Theoretical Physics group wants to make publicly available a set of computational analysis based on jupyter notebooks (interactive notebooks similar to matlab notebooks) developed by their students, transforming them into a Python package or C program. Currently, each notebook implements a single analysis, without defining functions. All notebooks work on datafiles in a specific format which are read through a publicly available parser (a python package). The notebooks often used explicit loops, as the code is written in a syntax similar to C, resulting in poor performance.

From the given information, describe what you think are the problems with the code and what information is missing that you need to ask the original author. Describe how you would proceed to rewrite the code in order to make it maintainable; how you can improve the performance of the software; the workflow you would follow together with the scientists in order to reshape the code and make it distributable.

Question 2 (9 points)

The Theoretical Physics group is composed by about 20 Principal Investigators (PI) working on related research topics. Each PI usually has a few master and/or PhD students, and collaborates with other PI, both within the group, within UniTN and outside UniTN. Several research lines include computer simulations, often run on remote clusters. The data from such simulations needs to be saved in such a way to satisfy the following requirements.

- It must be accessible by all collaborators inside the TP group and, upon request, by other scientists and collaborators.
- It must be stored in such a way as to ensure its identification, integrity, and reproducibility.

The candidate should argue what solutions are preferable among NAS storage, cloud storage provided by UniTN, cloud storage through personal subscriptions (e.g. dropbox) or others, like self-hosting a data server. The candidate should also describe what they think are the best practices to ensure the reproducibility of the data.

Question 3 (3 points)

Examine the following C function; indicate which lines, if any, have an error and, if so, what are the errors and their consequences.

```
int return_series_lover_n2 ( int n) {
    int *array = malloc(n * sizeof(int));
    for (int i=1; i<=n; i++)
        array[i] = 1/(i*i);
    return array;
}
```

Question 4 (3 points)

Given the following fragment of MPI code:

```
MPI_Comm_rank (comm, &my_rank);
if (my_rank == 0) {
    MPI_Send (sendbuf, count, MPI_INT, 1, tag, comm);
    MPI_Recv (recvbuf, count, MPI_INT, 1, tag, comm, &status);
} else if (my_rank == 1) {
    MPI_Send (sendbuf, count, MPI_INT, 0, tag, comm);
    MPI_Recv (recvbuf, count, MPI_INT, 0, tag, comm, &status);
}
```

briefly discuss which problems could arise by using it, and how to avoid them.

Question 5 (1 point)

Assume that you need to extract all values between V_min and V_max from column VALUE of the table RESULTS of the database EXPERIMENT . Which mysql command would you use:

- A) EXTRACT VALUE from EXPERIMENT.RESULTS where (VALUE >= V_min and VALUE < V_max);
- B) EXTRACT * from EXPERIMENT.RESULTS where (VALUE >= V_min and VALUE < V_max);
- C) SELECT * from EXPERIMENT.RESULTS where (VALUE >= V_min and VALUE < V_max);
- D) SELECT VALUE from EXPERIMENT.RESULTS where (VALUE >= V_min and VALUE < V_max);

Question 6 (1 point)

which MPI function would you use in order to send from one given process to all processes a different block of data for each receiver:

- A) MPI_Bcast
- B) MPI_Gather
- C) MPI_Scatter
- D) MPI_Reduce

Question 7 (1 point)

In the event of a personal data breach, the GDPR requires that the Data Processor (responsabile del trattamento) shall:

- A) notifies the Data Controller (titolare del trattamento) of the violation without undue delay
- B) notifies the violation to the competent supervisory authority without undue delay
- C) notifies the violation to the public security authority within 72 hours
- D) notifies the violation to the interested party without undue delay

Question 8 (1 point)

The Unix command:

```
find . -type f -name "*.txt" -exec grep -l Hello {} \;
```

- A) is syntactically incorrect
- B) lists all the files which are present in the current directory with extension "txt", and contain the word Hello
- C) lists all the files which are present in the current directory and in the subdirectories of the current directory with extension "txt" and contain the word Hello
- D) list all the files which are present in the current directory and in the subdirectories of the current directory that contain the word Hello

Question 9 (1 point)

In Python, tuples (...) are immutable objects, meaning that their contents can not be modified. The following code snippet, when executed will:

```
my_tuple = (1, 2.19, "ciao", [0, 1, 2] )  
my_tuple[-1][0] = 1
```

- A) Raise a TypeError exception.
- B) Raise a ValueError exception.
- C) Raise an IndexError exception.
- D) None of the above.

Question 10 (1 point)

Executing the Unix command:

```
ls -lS | head -5
```

The result will be:

- A) the list of files in the current directory that contain character 5
- B) the first 5 files in the current directory, in alphabetical order, which contain the S character
- C) the list of files in the current directory, sorted by size, which contain the character 5
- D) the first 5 files in the current directory, sorted by size