

## TRACCIA 3

### Question 1 (9 points)

One of the Principal Investigators of the Theoretical Physics group wants to build a webserver on top of an analysis software written in C. The code should run with parameters passed by the web interface; its output should then be displayed on the web interface as well. To launch the execution a user should input the parameters on the web interface and click run. Currently, the code is written in a single C file more than 2000 lines long. The file contains 10 functions. Analysis parameters are read from the standard input interactively, and the output is written on a text file in which different observables correspond to different columns. As the observable to be computed are decided at runtime, the number of columns can vary between execution - if an observable is not computed, its corresponding column is not present in the output file.

From the given information, describe what you think are the problems with the code if it has to become the backend for a web server, and what information is missing that you need to ask the original author. Describe how you would change the code in order to make it maintainable; how you would identify performance issues and possible runtime problems; the workflow you would follow together with the scientists in order to reshape the code.

### Question 2 (9 points)

Define a template set of directories (which could be created for example by an ad-hoc Python or Bash script, or through cookiecutter) suitable for the development of reproducible computational Physics projects. A typical Computational Physics project includes data from multiple simulations performed with a publicly available simulation package, a set of analysis scripts written in Python, a report in Latex. Each simulation is based on a different set of input files (parameters/data). The typical workflow of a scientist in the Theoretical Physics group involves multiple refinement cycles in which new analyses and/or simulations are added, thus modifying both the Python code, the input scripts, and the amount of data in output from the simulations. What would be the best way to structure the directory tree for such a project? Are there parts that could be versioned? Pay particular attention to the reproducibility of the results and data storage. Discuss how to maintain documentation.

### Question 3 (3 points)

Consider a parallel, MPI code which is used to simulate the properties of a large, homogeneous physical system (e.g. a fluid) inside a cubic box with periodic boundary conditions. By using more MPI cores to simulate the system, it is possible to divide the box in smaller cubic regions, each of which has the same size and is taken care of by a single core, resulting in a speedup. How does this speedup scale, in general, with the number of MPI cores? Why? How can we expect the speedup to change if the system does not have a homogeneous density, but contains some empty regions?

### Question 4 (3 points)

Consider the following C function. Write what it does.

```
double *d1t (int n1)
{
    double *p, *a;
    int i;
    if ((p = (double *) malloc ((size_t) n1 * sizeof (double)))==NULL)
        exit(1);
    for (i = 0, a = p; i < n1; i++)
        *a++ = 0;
    return p;
}
```

### Question 5 (1 point)

which mysql command would you use to backup the contents of the database EXPERIMENT?

- A) mysqldump -u root -p EXPERIMENT > experiment.sql
- B) mysqladmin -u root -p drop EXPERIMENT > experiment.sql
- C) mysqladmin -u root -p version EXPERIMENT > experiment.sql
- D) mysql -u root -p EXPERIMENT < experiment.sql

### Question 6 (1 point)

which MPI function would you use in order to send, from one root process to all processes, the same block of data for each receiver:

- A) MPI\_Bcast
- B) MPI\_Gather
- C) MPI\_Scatter
- D) MPI\_Reduce

### Question 7 (1 point)

The GDPR provides that, in the event that the processing is carried out by a public authority or body, except for courts acting in their judicial capacity, the Data Controller (il titolare del trattamento):

- A) could designate a Data Processor Officer (responsabile della protezione dei dati)
- B) cannot designate a Data Processor Officer
- C) shall designate a Data Processor Officer
- D) shall designates a Data Processor Officer only if the processing presents a high risk for the rights and freedoms of the Data Subject (l'interessato)

### Question 8 (1 point)

The Unix command:

```
find . -type d -name "*.txt"
```

- A) finds all files in the current directory and in the subdirectories of the current directory with the extension "txt"
- B) finds all files in the current directory with "txt" extension
- C) finds all directories in the current directory and in the subdirectories of the current directory with the extension "txt"
- D) finds all files and directories present in the current directory and in the subdirectories of the current directory with the extension "txt"

### Question 9 (1 point)

Consider the following python code:

```
import numpy as np
a1 = np.arange(10) # array([ 0,1,2,...9])
a2 = a1
for i in range(10):
    a2[i] = a2[i] // 2
a1s = a1.sum() # sum all elements of the array
a2s = a2.sum()
```

What are the values stored in the variables a1s and a2s?

- A) a1s = 45 ; a2s = 22.5
- B) a1s = 45 ; a2s = 20
- C) a1s = 20 ; a2s = 20
- D) a1s = 55 ; a2s = 25

### Question 10 (1 point)

The Unix command:

```
touch file_name
```

- A) returns the value 1 if there is a file named file\_name
- B) creates an empty file if the file file\_name does not exist
- C) deletes the file file\_name if that file is empty
- D) displays the access time and modification time of the file file\_name