Project exam

The objective of the project-exam is to experiment parallel programming by means of the implementation of the popular cluster algorithm “k-means”. This algorithm partitions a set of values (usually, d-dimensional real vectors) in k groups, by computing k representatives (usually named centroids): the values in input that share the same nearest centroid belong to the same computed partition.

In the literature there are several variants of this algorithm, that can differ in the way the initial approximations of the centroids are computed, that consider different notions of distance, etc. You can implement any of these variants, or even more than one. Implementation of different variants could be useful, for instance, to compare them in terms of their efficiency or of the quality of the produced solutions.

You can also freely decide on which data to test your implementation(s). In fact, clustering has many possible applications, and you are free to decide the nature of the values to partition. You can apply your implementation on meaningful datasets, as well as on synthesized artificial input data.

Your project work must be described in a PDF document (of at most 6 pages) in which you present the (variants of the) k-means algorithm you have implemented, the design decisions that you have taken to realize a parallel implementation, and the way in which you have evaluated your program. Concerning this last part, you should describe the type of input data that you have decided to use, the approach taken to experiment the performances of your solution(s), and the final results about its (parallel) scalability.

All your experiments should be reproducible, namely, you should submit a zip file with your source code, input data, and scripts to compile and launch your project. As a common execution platform, we will consider the hpc cluster *slurm.cs.unibo.it* offered by the DISI department:
https://disi.unibo.it/it/dipartimento/servizi-tecnici-e-amministrativi/servizi-informatici/utilizzo-cluster-hpc
At the date in which this document was written, the above web page contained instructions to access the service written only in Italian. I kindly ask to the group of students of the course, that includes also italian students, to collaborate in order to help non-italian colleagues in the extraction of the minimal information needed to access the service.

You must submit your zip file via the "virtuale" web page of the course. Your project will be unzipped on *slurm.cs.unibo.it*. Once unzipped, a file named "project.sbatch" should be present, in order to be able to compile and execute your project simply by launching "sbatch project.sbatch". Another file named "project.pdf" will contain the PDF document described above.

You can submit your project exam whenever you want. After you have uploaded your project, please send also an email to *gianluigi.zavattaro@unibo.it* to notify the submission. The final deadline for the submissions is September 30th, 2022.

Additional optional work:

In case you are interested in experimenting also GPU programming, you can use the additional teaching material present on the "virtuale" web page of the course, to learn about the CUDA programming language. You can use CUDA to offload your application on the NVIDIA GPUs available in the *slurm.cs.unibo.it* cluster. In this case, your submitted project will include also CUDA source code, and your sbatch script will have to include also the commands for compiling and executing such code. Obviously, also this additional optional work should be discussed in the submitted PDF document that, in this case, could be 8 pages long.