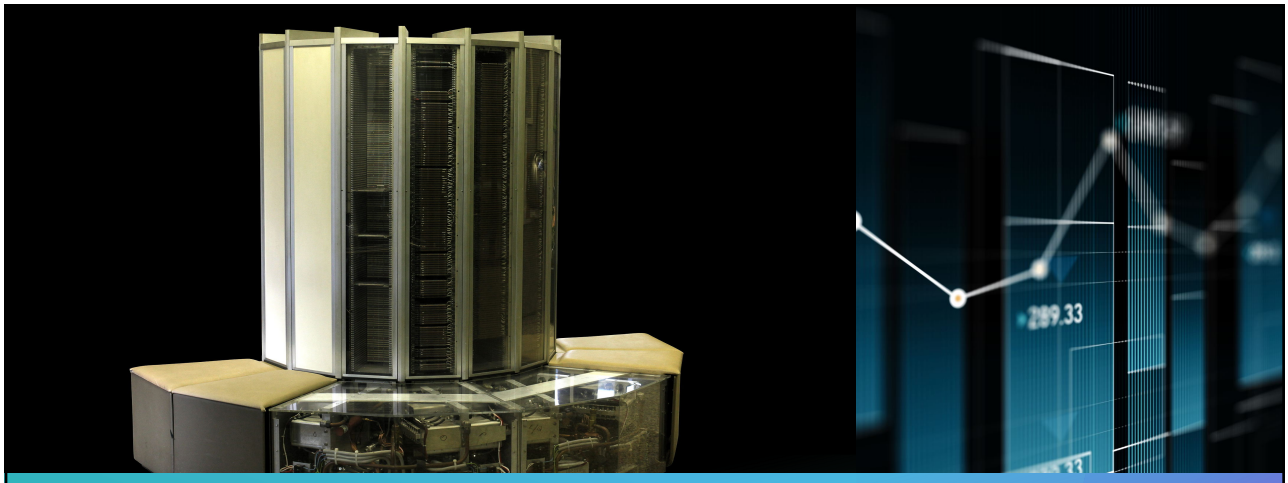
2021

1



MASSIVE COMPUTING

BACHELOR IN DATA SCIENC AND ENGINEERING

2021

2

SHORT RESUME

- Harold Molina-Bulla
  - Signal Theory and Communications
  - Working with HPC and Cluster of Workstations/Servers since 1993
  - Implement a medium size Cluster of Servers with 1250 computational cores and 5.3 Tbytes of shared memory since 2004
    - This facilty is located in the Department of Signal Theory and Communcations
    - It is in constant evolution
- Contact: hymolina@ing.uc3m.es or h.molina@tsc.uc3m.es

2021

3

ABOUT THE COURSE

- GOALS
  - Know the different tools for Massive Computing:
    - Local intensive computing: Use all the cores available in your computer
    - Heterogeneous local intensive computing: Use the cores and your Graphics Card (GPU)
    - Distributed intensive computing: Use several computers to crunch data
  - Use of these tools applied to research
  - Learn when you can use a specific tool
    - Not always use the GPU is the best tool to solve a problem
      - Why not?? This is a good question

2021

4

SYLLABUS

- Introduction to Python and several tools
- Short introduction to Massive Computing/High Performance Computing
  - What is Massive Computing (M.C.)?
  - Short History
  - When can I apply M.C.?
  - Common terms in M.C.

2021

5

SYLLABUS

- Local M.C.:
  - Programing M.C. using Parallel Execution in our computers
  - How can I program to use all the cores in my computer
  - Hazzards when you program in Parallel Execution
    - Hazzards of memory
  - Optimize code for local execution
  - Practical works
    - Basic programs
    - Optimizations
    - Common errors

2021

6

SYLLABUS

- Heterogeneous M.C.
  - Use of GPU
  - Architecture of GPUs
  - When can I use GPU
  - Problems using GPU
  - Tips to programming for GPUs
  - Common errors programing for GPUs
  - Practical Work (many, many small programs)…

2021

7

SYLLABUS

- Distributed M.C.
  - What is Big Data!! We make or we have, or we "what?", Big Data
  - What is our atomic data? Can I distribute all the processes?
  - Map Reduce paradigm
  - Use of State of the Art Frameworks for Distributed Data Processing
  - Introduction to SPARK
  - Machine Learning Algorithms in SPARK

2021

8

GRADES

- 20% Each practical work: The student have to upload the practical work notebook 1 week after the class, with results, code commented, comments and conclussions.
- 3 practical works
  - 1 for each theme
  - Can use previous themes
  - 20% Local M.C.
  - 20% GPU M.C.
  - 40% Distributed M.C. (Emphasis in Distributed Machine Learning)

2021

9

BIBLIOGRAPHY

- **Advanced Python Programming: Build High Performance, Concurrent, and Multi-Threaded Apps with Python Using Proven Design Patterns;** Lanaro, Gabriele ; Nguyen, Quan ; Kasampalis, Sakis
- **Hands-On GPU programming with Python and CUDA;** Brian. Tuomanen
- **Spark: Big Data Processing Made Simple;** Chambers, Bill ; Zaharia, Matei
- **Machine learning with Spark : create scalable machine learning applications to power a modern data-driven business using Spark;** Nick Pentreath
- **Big Data Analytics with Spark A Practitioner's Guide to Using Spark for Large Scale Data Analysis;** Mohammed. Guller.

2021

10