

---

# **Directorio de Computación**

***Versión 0.1***

**Carlos Luque**

**15 de noviembre de 2022**



---

## Índice general

---

|   |          |
|---|----------|
| <b>1. Contenidos</b>  | <b>3</b> |
| 1.1. Computación de alto rendimiento o High-Performance Computing (HPC) . . . . . | 3        |
| 1.2. Software . . . . .   | 8        |
| 1.3. Análisis de datos . . . . .  | 9        |
| 1.4. Inteligencia artificial . . . . .  | 9        |
| 1.5. Visualización . . . . .  | 10       |
| 1.6. Acceso a superordenadores en España . . . . .                                | 10       |



Bienvenido al directorio de información de computación del centro de competencia español, [EuroCC\\_Spain](#). En este directorio, encontrarás una breve introducción a términos y temas asociados a *computación de alto rendimiento o High Performance Computing (HPC)*, *análisis de datos* e *inteligencia artificial*. A su vez, se asocian enlaces y videos que profundizan en los términos. El directorio esta disponible en formato pdf, html y epub.

El directorio se ha elaborado para usuarios tanto principiante como experto en estos temas. Para un conocimiento más profundo, puede escribirnos a [eurocc@res.es](mailto:eurocc@res.es).



### 1.1 Computación de alto rendimiento o High-Performance Computing (HPC)

El término *computación de alto rendimiento* hace referencia a cualquier sistema informático que su capacidad de cómputo es mayor que pueden manejar los ordenadores de sobremesa normales. En otras palabras, la computación de alto rendimiento es un conjunto de tecnologías hardware y software que es capaz de tratar información en el menor tiempo posible es decir a gran velocidad. El término *supercomputación* se usa como sinónimo a computación de alto rendimiento.

- Supercomputación en la empresa, rompiendo barreras
  - <https://www.youtube.com/watch?v=TtGpnoRmQLc>
- Supercomputación e Industria - Salud y Medio Ambiente
  - <https://www.youtube.com/watch?v=vX3bMgx7ssw>
- Supercomputación e Industria - Manufactura
  - [https://www.youtube.com/watch?v=T2qQZ\\_SCUZY](https://www.youtube.com/watch?v=T2qQZ_SCUZY)
- Introducción en computación de alto rendimiento
  - <https://www.youtube.com/watch?v=7zJUceJiYxQ>
- Entrevista con Mateo Valero, ¿Para qué se necesita realmente la supercomputación?
  - <https://www.innovaspain.com/bsc-para-que-sirve-supercomputacion/>

Los ordenadores de alto rendimiento contienen todos los componentes habituales de un ordenador de sobremesa - procesador (CPU), memoria principal (RAM), almacenamiento (storage), refrigeración, etc.

En la taxonomía de los ordenadores de alto rendimiento, los *superordenadores* o *supercomputadores* son los más potentes en computación.

Los superordenadores es una agrupación (cluster en inglés) de ordenadores que están interconectado por una red de alto rendimiento y conectado a un sistema de almacenamiento de alto rendimiento.

- Superordenador
  - <https://www.youtube.com/watch?v=nctTZplQY-o>

A cada ordenador en un superordenador es conocido como nodo. Cada nodo puede ser construido con un o más procesadores, una cantidad de memoria principal y una capacidad de disco duro local limitada. Además, puede contener distintos tipos de aceleradores:

- La unidad de procesamiento gráfica (GPU en inglés): Este tipo de hardware es usado generalmente para el procesamiento gráfico. A su vez, es capaz de proporcionar una cantidad elevada de capacidad de cómputo, es decir, cientos de procesos ejecutándose paralelamente. Este tipo de hardware es adecuado para el cálculo algebraico lineal.
- Una matriz de puertas lógicas programable en campo (FPGA en inglés): Es una hardware programable que puede ser utilizado para acelerar algunos procesos específicos que normalmente se llevan a cabo con los procesadores.

Generalmente, los ordenadores de alto rendimiento son gestionados por el sistema operativo Linux.

- ¿Qué es Linux?
  - <https://www.youtube.com/watch?v=UUJ0dFpj1-M>
  - <https://www.youtube.com/watch?v=NDhJfHhe3e4>
  - <https://www.youtube.com/watch?v=OR3tWKKmhEM&t=5s> (en inglés)

Típicamente, la interacción con los ordenadores de alto rendimiento se realiza con la consola o shell.

- Introducción a shell
  - <https://gitlab.com/makhlaghi/smack-talks-iac/-/blob/master/smack-2-shell.md>
  - <https://www.youtube.com/watch?v=48r76WVkJUI>

A su vez, popularmente, las aplicaciones software son gestionadas por el sistema *module*.

- Introducción
  - <https://gitlab.com/makhlaghi/smack-talks-iac/-/blob/master/smack-18-module-py-virtual-env.rst>
- Se distribuyen en dos versiones:
  - TCL: Distribución más expandida en los ordenadores de alto rendimiento actualmente
    - <https://modules.readthedocs.io/en/latest/index.html>
  - LMOD: Distribución en expansión y ofrece una gran versatilidad.
    - <https://lmod.readthedocs.io/en/latest/>

### 1.1.1 Gestor de colas/trabajos

Tradicionalmente, los ordenadores de alto rendimiento emplean algún sistema para la gestión de los recursos del ordenador entre los usuarios. Por lo general, a estos sistemas se conoce con el nombre de *gestor de colas* o *gestor de trabajos*. En un gestor de cola, un *trabajo* o *job* (en inglés) se refiere a una aplicación con los datos de entrada para su ejecución.

Los gestores de colas principales son:

- *SLURM* (Simple Linux Utility for Resources Management)
- *PBS* (Portable Batch System) o *TORQUE* (Terascale Open-source Resource and QUEue Manager) que se distribuye en el estándar abierto, OpenPBS (<https://www.openpbs.org/>)

Hoy en día, el gestor de cola SLURM es el más popular en los ordenadores de alto rendimiento:



- Página oficial:
  - <https://slurm.schedmd.com/>
- Resumen de los comandos:
  - <https://slurm.schedmd.com/pdfs/summary.pdf>
  - <https://www.scayle.es/manual/es/hpc/gestor-de-trabajos>
  - <https://research.iac.es/sieinvens/siepedia/pmwiki.php?n=HOWTOs.LaPalma3UsefulCommands2> (en inglés)

### 1.1.2 Programación paralela

Esta sección, se expone la programación aplicada generalmente en ordenadores de computación de alto rendimiento. El concepto de rendimiento es vago y puede significar varias cosas. Nosotros usaremos el concepto de rendimiento significado la resolución de un problema en el menor tiempo posible. En cambio, la productividad es cuánto trabajo se puede realizar por unidad de tiempo.

En el entorno de computación de alto rendimiento, el rendimiento es la métrica principal que se quiere conseguir optimizar, es decir, la resolución de un problema se complete en el menor tiempo posible.

Tradicionalmente, la *computación en serie* se ha empleado para el desarrollos de las aplicaciones software y consiste en la ejecución o división del trabajo una tras otra. En cambio, a mediado de la década de 2000, se democratizó los ordenadores paralelos motivando el uso de la computación en paralelo.

La *computación en paralelo* aplica el concepto de *paralelismo* que consiste en la realización de múltiples cosas al mismo tiempo. En concreto, se puede definir como la utilización de varios recursos de computación para la resolución de un problema. Este paradigma de computación permite:

- La división de un problema en partes discretas para su ejecución simultáneamente, es decir, en múltiples recursos de computación como los procesadores.

En la computación en paralelo se emplean varios/as modelos/paradigmas de programación, conocido también como programación paralela, de las cuales se resaltan:

- Memoria compartida, OpenMP
- Hilos, (OpenMP o POSIX threads en linux)
- Pasos de mensajes, MPI

Estos paradigmas de programación son abstracciones sobre de arquitecturas del ordenador, y se pueden emplear en cualquier de las mismas.

### Paradigmas de programación

- OpenMP

Un modelo de programación portable y escalable que proporciona a los programadores una interfaz simple y flexible para el desarrollo de aplicaciones paralelas. En otras palabras, es una interfaz de programación de aplicaciones (API) para la programación paralela de memoria compartida.

- Página web oficial
  - <https://www.openmp.org/>
  - Introducción a openMP
    - <https://www.youtube.com/watch?v=nE-xN4Bf8XI&list=PLLX-Q6B8xqZ8n8bwjGdzBJ25X2utwnoEG>

- Resumen de las directivas:
  - <https://www.openmp.org/wp-content/uploads/OpenMP-4.0-C.pdf>

### ■ MPI

MPI (Message Passing Interface) es un estándar que define la sintaxis y la semántica de las funciones contenidas en una biblioteca de paso de mensajes

- Introducción de MPI (inglés)
  - <https://mpitutorial.com/tutorials/mpi-introduction/>
  - <https://www.youtube.com/watch?v=D0-xSWBGNAw> (What is Open MPI)
- Resumen del lenguaje
  - <http://www.netlib.org/utk/people/JackDongarra/WEB-PAGES/SPRING-2006/mpi-quick-ref.pdf>
- GPGPU

La computación de propósito general en la unidad de procesamiento gráfica (GPU) consiste en aprovechar las capacidades de cómputo de las GPUs. La GPU es una aceleradora de computación para la álgebra lineal, como el aprendizaje profundo. Principalmente, los grandes fabricantes de GPUs son Nvidia y AMD.

- Introducción GPU \* <https://www.youtube.com/watch?v=9zfojkQXNoQ>

Varios modelos de programación que son empleados

- Nvidia CUDA (lenguaje de programación para las tarjetas gráficas de Nvidia)
  - <https://developer.nvidia.com/cuda-toolkit>
  - <https://www.youtube.com/watch?v=iwEoVufzqhg>
  - Parallel computing on GPUs with CUDA (curso en inglés)
  - <https://www.youtube.com/playlist?list=PL6lZw8yhOjVh9Hq89p1ECI9hpzjk3ivoq>
- OpenACC (programación se realiza con directivas (pragmas) y biblioteca e independiente del fabricante de la tarjeta gráfica)
  - <https://www.openacc.org/>
- AMD HIP, lenguaje de programación para las tarjetas gráficas de AMD. Este lenguaje también puede ser empleado con las tarjetas gráficas de Nvidia.
  - Introducción:
    - <https://www.youtube.com/watch?v=3ZXbRJVvgJs>
  - Guía de programación
    - <https://docs.amd.com/>
    - [https://rocm-docs.amd.com/en/latest/Programming\\_Guides/HIP-GUIDE.html](https://rocm-docs.amd.com/en/latest/Programming_Guides/HIP-GUIDE.html)
    - [https://github.com/RadeonOpenCompute/ROCm/blob/rocm-4.5.2/AMD\\_HIP\\_Programming\\_Guide.pdf](https://github.com/RadeonOpenCompute/ROCm/blob/rocm-4.5.2/AMD_HIP_Programming_Guide.pdf)

### 1.1.3 Compiladores

La creación de una aplicación software desde el código fuente se consigue a través de los *compiladores*. Este proceso es conocido como *compilación*. Los compiladores principales son:

- **GNU GCC**: es un conjunto de compiladores y librerías de los lenguajes C, C++, Objective-C, Fortran, Ada, Go, and D.
  - Página oficial
    - <https://gcc.gnu.org/>
- **Intel oneAPI**: conjunto de herramientas y librerías para el desarrollo de aplicaciones a través de diferentes arquitectura de computadores.
  - Página oficial
    - <https://www.intel.com/content/www/us/en/developer/tools/oneapi/overview.html#gs.a5u5dz>
    - <https://www.intel.com/content/www/us/en/developer/tools/oneapi/toolkits.html#gs.a5xmdj>
    - Introducción
      - ◇ <https://www.youtube.com/watch?v=NsFGFbdPsh0> (en inglés)
      - ◇ <https://www.youtube.com/watch?v=yнкDh4yYybs>
- **Nvidia HPC SDK**: conjunto de software, librerías y compiladores para los lenguajes C, C++, Fortran, directivas OpenACC y CUDA
  - Página oficial
    - <https://developer.nvidia.com/hpc-sdk>

En la siguiente tabla se muestra la relación de los compiladores y los lenguajes:

| Lenguaje     | GNU GCC  | Intel oneAPI                | Nvidia HPC SDK |
|--------------|----------|-----------------------------|----------------|
| C            | gcc      | Actual: icx; Clásico: icc   | nvc            |
| C++          | g++      | Actual: icpx; Clásico: icpc | nvc++          |
| Fortan       | gfortran | Actual: ifx; Clásico: ifort | nvfortran      |
| Aceleradores | –        | dpc++ (C++ basado en SYCL)  | nvcc (CUDA)    |

### 1.1.4 Buenas prácticas

#### Control de versiones

El control de versiones se define como la gestión de la historia de cambio de un proyecto. Una versión, revisión o edición de un proyecto, es el estado en el que se encuentra el mismo en un momento dado de su desarrollo o modificación. Esta gestión permite el trabajo en equipo en el mismo proyecto sincronizado las contribuciones de cada miembro del equipo.

Generalmente, el control de versiones se realiza a través de un sistema de control de versiones (en inglés Version Control System, VCS). De estos sistemas hay una gran abanico y nosotros resaltaremos el *git*:

- Página oficial de git
  - <https://git-scm.com/>
- Videos de git (en inglés):
  - Principiante: <http://iactalks.iac.es/talks/view/1426>

- Medio: <http://iactalks.iac.es/talks/view/1428>
- Avanzado: <http://iactalks.iac.es/talks/view/1438>

Para mayor información de buenas prácticas puedes leer la [página web](#) (best practices del EuroCC danés).

## 1.2 Software

En la sección se puede encontrar varios tipos de aplicaciones software para diferentes campos:

### ■ Computación de dinámica de fluido

- Ansys Fluent: Software de simulación de fluido
  - Página oficial
    - ◇ <https://www.ansys.com/products/fluids/ansys-fluent>
- OpenFoam: aplicación software de software libre y código abierto
  - Página oficial
    - ◇ <https://www.openfoam.com/>
- Basilisk: programa de software libre para la solución de ecuaciones diferenciales parciales en mallas cartesianas adaptativas
  - Página oficial
    - ◇ <http://basilisk.fr/>
- SU2: SU2 es un conjunto de herramientas de software de código abierto para la solución numérica de ecuaciones diferenciales parciales (PDE en inglés) y la realización de optimización restringida por PDE
  - Página oficial
    - ◇ <https://su2code.github.io/>

### ■ Ciencia de los materiales computacional

- LAMMPS: simulador de dinámica molecular
  - Página oficial
    - ◇ <https://www.lammps.org/#gsc.tab=0>
- Quantum Espresso
  - Página oficial
    - ◇ <http://www.quantum-espresso.org/>
- VASP: Vienna Ab initio Simulation Package
  - Página oficial
    - ◇ <https://www.vasp.at/>
- GROMACS: conjunto de software libre para el análisis dinámica molecular y de resultados de alto rendimiento
  - Página oficial
    - ◇ <https://www.gromacs.org/>

## 1.3 Análisis de datos

El *análisis de datos* es el tratamiento de datos para generar conocimiento en la ayuda de la toma de las decisiones adecuadas. En el entorno de negocio, se usa los términos de *inteligencia de negocio* o *Business Intelligence* y el *análisis de negocio* o *business analytics*

- Qué es inteligencia de negocio
  - [https://www.youtube.com/watch?v=4MG6\\_MGhusM](https://www.youtube.com/watch?v=4MG6_MGhusM)
  - [https://www.youtube.com/watch?v=Pmdps2kK\\_5M](https://www.youtube.com/watch?v=Pmdps2kK_5M)
- Análisis en R
  - <https://github.com/csc-training/da-with-r-remote> (en inglés)
- En python
  - <https://www.utc.fr/~jlaforet/Suppl/python-cheatsheets.pdf>
  - Librería Pandas
    - <http://iactalks.iac.es/talks/view/1463>

### 1.3.1 Big Data

Big data es un tipo de datos en el análisis de datos con operaciones especiales. En español se conoce como datos masivos. Su finalidad es la captura y procesar los datos.

- Diferencias entre Big Data, Business Analytics y Business Intelligence
  - <https://www.campusbigdata.com/big-data-blog/item/148-diferencias-entre-big-data-business-analytics-y-business-intelligence>
- Introducción a Apache Spark
  - <https://www.youtube.com/watch?v=5okC1jHIENw>

## 1.4 Inteligencia artificial

El término engloba a un conjunto de algoritmos con el objetivo de inventar máquinas que presenten las mismas capacidades que el ser humano. Las capacidades como el razonamiento, el aprendizaje, y la creatividad.

- ¿Qué es y cómo funciona la INTELIGENCIA ARTIFICIAL?
  - [https://www.youtube.com/watch?v=\\_tA5cinv0U8](https://www.youtube.com/watch?v=_tA5cinv0U8)
- Diferencias entre Data Science, Inteligencia Artificial, Machine Learning y Deep Learning
  - <https://www.programatically.com/education/las-diferencias-que-hay-entre-data-science-artificial-intelligence-machine-learning>

Las técnicas de la inteligencia artificial se agrupan en dos principales grupos:

- *Machine learning* o *aprendizaje automático*: busca como dotar a las máquinas la capacidad de aprendizaje (La generalización del conocimiento a partir de un conjunto de experiencias).
- *Deep learning* o *aprendizaje profundo*: una modalidad de aprendizaje automático con redes neurales.
  - ¿Qué es el Machine Learning? ¿Y Deep Learning? Un mapa conceptual
    - <https://www.youtube.com/watch?v=KytW151dpqU>
  - ¿Redes Neuronales CONVOLUCIONALES! ¿Cómo funcionan?

- <https://www.youtube.com/watch?v=V8j1oENVz00&list=PL-Ogd76BhmcBaUXZGPJkmQpLgrBgGZ7v0>

## 1.5 Visualización

En esta sección se expone diferentes métodos para la visualización de volúmenes de datos remotamente.

- Virtual Network Computing, VNC
  - TeamViewer \* <https://www.teamviewer.com/es/info/vnc/>
  - TigerVNC \* <https://tigervnc.org/>
  - turboVNC (soporte para GPU) \* <https://turbovnc.org/>
- Paraview: permite el renderizado remotamente y visualización local.
  - <https://www.paraview.org/>

## 1.6 Acceso a superordenadores en España

En esta sección, se puede encontrar centros y convocatorias en España para el acceso a superordenadores.

### 1.6.1 Centros españoles

- Fundación Computación y Tecnologías Avanzadas de Extremadura (COMPUTAEX-CénitS)
- El instituto de Biocomputación y Sistemas Complejos (BIFI)
- Centro de supercomputación de Galicia (CESGA)
- El Consorci de Serveis Universitaris de Catalunya (CSUC),
- Supercomputación Castilla y León (SCAYLE)

### 1.6.2 Convocatorias en España

- **EuroCC Spain testbed**: la convocatoria proporciona acceso gratuito a estos recursos por un tiempo limitado a empresas y administraciones públicas españolas.
- **La Red Española de Supercomputación**: la convocatoria proporciona acceso a recursos de supercomputación y gestión de datos a proyectos de investigación desde grupos de investigación y empresas.

