

# Computación de Altas Prestaciones (HPC)

## Acceso al clúster y copia de archivos

El acceso se hace a través de una máquina que actúa como *frontend* a través de *ssh*. Para más información, consulta [cómo acceder al clúster y copiar ficheros](#).

Antes de acceder por primera vez, lee atentamente la sección de introducción que sigue para entender cómo funciona el sistema. También encontrarás información que te servirá de referencia más adelante.

## Introducción

### Funcionamiento básico de un clúster de computación

El Clúster de Computación de Altas Prestaciones (HPC, *High Performance Computing*) proporciona a los investigadores una infraestructura compartida que sirve como plataforma adecuada a la ejecución de procesos computacionalmente costosos.

Un **clúster de computación** se compone de un conjunto de nodos interconectados mediante una red dedicada que puede actuar como un único elemento computacional. Esto proporciona una gran potencia, permitiendo la ejecución de trabajos paralelos muy grandes o muchas ejecuciones pequeñas de forma concurrente.

Un **sistema de gestión de colas** (SGC) es un software que planifica la ejecución de los trabajos que se encuentra habitualmente en los sistemas de este tipo, ya que permite una gestión eficiente de los recursos con múltiples usuarios. En este clúster está instalado el sistema **PBS/TORQUE**.

La dinámica de funcionamiento de un sistema de este tipo es la siguiente:

1. El usuario solicita la ejecución de una tarea<sup>1)</sup> con unos recursos determinados.
2. El sistema registra la solicitud en una de sus colas de entrada<sup>2)</sup> y, según la cantidad de recursos solicitados, la deriva a una cola de sistema determinada<sup>3)</sup>.
3. En función de la prioridad de la cola de sistema (a menores recursos necesarios, mayor prioridad) y de la disponibilidad de los recursos en el sistema, la tarea se enviará a uno o varios de los nodos computacionales. Al terminar la ejecución, se devolverá la salida generada.

Lo habitual es que la ejecución tenga que **esperar en la cola** hasta que los recursos estén disponibles y preparados. Además, resulta imposible realizar así ejecuciones de manera interactiva<sup>4)</sup>.

## Descripción del hardware

El clúster **ctcomp2** es un clúster heterogéneo, formado por 8 nodos computacionales *HP Proliant BL685c G7*, 5 nodos *Dell PowerEdge M910* y 5 nodos *Dell PowerEdge M620*.

| Nombre del nodo | Modelo   | Núcleos por nodo (NUMAs, Sockets, Cores/socket, Threads/core) | Memoria RAM         |
|-----------------|--|---|---------------------|
| node1-7         | HP Proliant BL685c G7 - Procesadores AMD         | 64 (8, 4, 8, 2)   | 256GB <sup>5)</sup> |
| inode11-15      | Dell PowerEdge M910 - Procesadores Intel Xeon L  | 32 (4, 4, 8, 1)   | 64GB                |
| inode16-20      | Dell PowerEdge M620 - Procesadores Intel Xeon E5 | 32 (2, 2, 8, 2)   | 64GB                |

Internamente los nodos computacionales están conectados entre sí a través de varias redes 10 GbE dedicadas. La conexión desde el exterior es de 1Gb.

Casi todos los nodos, a excepción de *node1* e *inode20* se apagan cuando no se están utilizando durante un rato, lo que podría causar retrasos de varios minutos en la cola aunque el cluster aparentemente esté desocupado.

[Información técnica de los procesadores y opciones de BIOS](#)

## Descripción del software

El sistema operativo es *Debian GNU/Linux 8.6 (jessie)*.

El sistema de colas que gestiona los trabajos es PBS/Torque, apoyado por CLUES para mejorar la gestión de energía.

- [MAUI 3.3.1](#)
- [Torque 4.1.3](#)
- [CLUES 0.88](#)

El sistema permite compilar y ejecutar código en C++, Java, Python, R, Octave... El uso de determinado software puede requerir su carga a través del sistema *Modules*. Puedes consultar el catálogo completo de software disponible en el siguiente enlace: [Información del software disponible](#)

## Sistema de colas

El sistema de colas tiene cuatro **colas de entrada** y ocho **colas del sistema**. Un usuario envía un trabajo a una cola de entrada y, en función de los recursos solicitados, se envían a una u otra cola de sistema.

Para enviar un trabajo al sistema de colas, se utiliza el comando `qsub`. Por defecto se envía a la cola de entrada *batch*. Si el clúster se encuentra muy saturado, puede ser interesante especificar la cola *short* o *bigmem*.

| Cola  | Número máximo de trabajos encolados <sup>6)</sup> | Características         | Ejemplo |
|-------|---|-------------------------|---------|
| batch | 256   | Es la cola por defecto. |         |

| Cola        | Número máximo de trabajos encolados <sup>6)</sup> | Características  | Ejemplo  |
|-------------|---|--|--|
| short       | 256   | Cola especial con mayor prioridad para trabajos de como máximo 12 horas, 16 procesos, 5 nodos y 32GB   | <code>qsub -q short trabajo.sh</code>                    |
| bigmem      | 8   | Otra cola especial con mayor prioridad para trabajos de, específicamente, 64 procesos y un nodo. Es obligatorio especificar esos recursos ( <code>nodes=1:ppn=64</code> ). | <code>qsub -q bigmem -l nodes=1:ppn=64 trabajo.sh</code> |
| interactive | 1   | Cola especial para trabajar de forma interactiva. Debe llamarse de forma especial.   | <code>qsub -q interactive -I</code>                      |

Los recursos que se pueden solicitar y que determinarán la cola del sistema utilizada son:

- El número de nodos
- El número de núcleos por nodo
- El tiempo de ejecución

La memoria máxima asignada depende de la cola del sistema que se utilice finalmente, así que estará determinada por el resto de parámetros.

Los recursos pueden solicitarse de dos formas: como comentarios al inicio del script, o con el parámetro `-l` (que tiene precedencia). Puedes ver [ejemplos de trabajos](#) para ver cómo se especifican estos recursos.

Comprueba los límites existentes en las colas del sistema, para determinar la mejor elección de recursos posible<sup>7)</sup>:

| Cola     | Límites               |                     |                             |                                 |                       |                          |
|----------|-----------------------|---------------------|-----------------------------|---------------------------------|-----------------------|--------------------------|
|          | Núcleos <sup>8)</sup> | Nodos <sup>9)</sup> | Memoria (GB) <sup>10)</sup> | Trabajos/usuario <sup>11)</sup> | Tiempo máximo (horas) | Prioridad <sup>12)</sup> |
| np1      | 1                     | 1                   | 1,99                        | 300                             | 672                   | 6                        |
| np2      | 2                     | 2                   | 3,75                        | 150                             | 192                   | 5                        |
| np4      | 4                     | 4                   | 7,5                         | 75                              | 192                   | 4                        |
| np8      | 8                     | 5                   | 15                          | 40                              | 192                   | 4                        |
| np16     | 16                    | 5                   | 31                          | 20                              | 192                   | 3                        |
| np32     | 32                    | 5                   | 63                          | 10                              | 288                   | 2                        |
| np64     | 64                    | 5                   | 127                         | 5                               | 384                   | 1                        |
| parallel | 32-160                | 5                   | 64                          | 15                              | 192                   | 3                        |

Puedes leer más sobre [cómo preparar los trabajos para su envío al gestor de colas](#) y [cómo enviar y gestionar trabajos una vez enviados](#).

1)

Normalmente un script de *bash*

2)

Típicamente *batch*

3)

Llamadas *np1*, *np2*, *np4*...

4)

Sin embargo, existe una cola interactiva especial para ejecuciones interactivas, orientada a solucionar problemas en las ejecuciones

5)

node1 128GB

6)

Sólo cuentan los trabajos que no han sido aún enviados a una cola del sistema, así que esto no quiere decir que sólo puedas enviar ese número de trabajos

7)

Si pides una situación imposible, por ejemplo, 8 núcleos con 200 horas, el trabajo permanecerá en la cola de entrada y jamás se llegará a ejecutar

8)

parámetro ppn

9)

parámetro nodes

10)

Determinada por el resto de parámetros

11)

Si se alcanza este número de trabajos, los trabajos restantes permanecerán en la cola de entrada mientras no se liberen trabajos en la cola de sistema

12)

Mayor número = mayor prioridad

From:

<https://wiki.citius.usc.es/> - Wiki do CiTIUS

Permanent link:

<https://wiki.citius.usc.es/es:centro:servizos:hpc>

Last update: **2019/12/03 09:42**

