

Laboratorio real en Cronos

Introducción

El objetivo de este documento es dar un primer impulso para que cada equipo tenga las bases necesarias para desarrollar la primera meta volante. Primero se explicará cómo acceder a los recursos computacionales del curso, y un poco de conocimiento básico de manejo de supercomputadores para que puedan iniciar con las actividades de la primera entrega.

Cronos es un supercomputador que era usado por investigadores de la universidad. Actualmente se usa como laboratorio de pruebas de Apolo. La universidad gentilmente nos habilitó algunos nodos de Cronos para desarrollar parte del curso, y darles la oportunidad de administrar, usar y optimizar un supercomputador real. A cada equipo se le asignará un cluster de dos máquinas al cual podrá ingresar. Cada equipo usará su cluster para desarrollar y solucionar los retos que se proponen en clase. Todos los clusters son iguales y cada equipo solamente tiene acceso a 1 cluster. Con esto nos aseguramos de que todos estén en igualdad de condiciones y se generen resultados comparables y reproducibles. Únicamente se tiene autorizado el acceso a los servidores que se mencionan en este documento. Cada equipo es responsable de la configuración y mantenimiento de su cluster.

A continuación se mostrarán cuáles son los nodos de cada cluster:

Cluster/Equipo	Hostname Nodo1	Hostname Nodo 2
1	compute-1-1	compute-1-2
2	compute-1-5	compute-1-11
3	compute-1-8	compute-1-9
4	compute-1-12	compute-1-13
5	compute-1-15	compute-1-16

Cómo ingresar a cada Cluster

Nota: Recuerde que para poder ingresar al cluster es necesario tener activada y configurada la VPN de Apolo.

Para ingresar a cada nodo se requiere la IP y la contraseña respectiva. Cuando se definan los equipos de trabajo se comparte esta información.

Linux/Mac

1. Abra la terminal de su preferencia: eg. urxvt

```
[user1@sacarch ~]$
```

2. Una vez dentro de la terminal ejecute el siguiente comando donde <ip> es la ip del nodo al cual quiere entrar:

```
$ ssh -J proxyjump@cronos.eafit.edu.co root@<ip>
```

```
[user1@sacarch ~]$ ssh -J proxyjump@cronos.eafit.edu.co root@10.150.7.200
```

3. Se le preguntará si quiere continuar con la conexión; escriba 'yes' para continuar:

```
[user1@sacarch ~]$ ssh -J proxyjump@cronos.eafit.edu.co root@10.150.7.200
The authenticity of host 'cronos.eafit.edu.co (200.12.187.172)' can't be e
stablished.
ED25519 key fingerprint is SHA256:Va+uj9kNpE7+ASoD6w1Tt0bE30f0xGbVqEsi9guJ
0ZA.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
```

4. Luego le pedirá la contraseña de proxy jump la cual es: "7h!5!5jump3r"

```
proxyjump@cronos.eafit.edu.co's password:
```

5. Le volverá a preguntar si quiere continuar con la conexión escriba 'yes' de nuevo
6. Después se le preguntará la contraseña para ingresar al nodo (La daremos por privado a cada equipo):

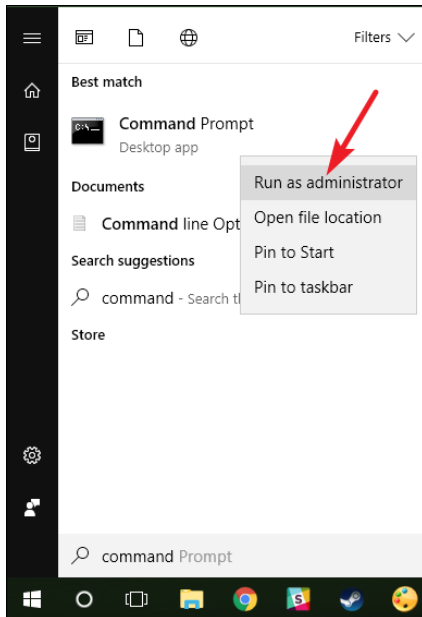
```
root@10.150.7.200's password:
```

7. Si todo fue exitoso le mostrará que esta logueado con root en el compute que deseaba:

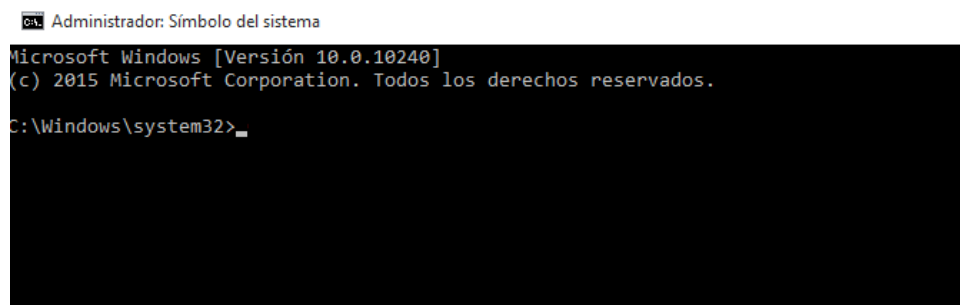
```
[root@compute-1-1 ~]#
```

Windows

Nota: Si está en Windows se recomienda utilizar putty (<https://www.putty.org/>).



1. Presione la tecla Win para abrir el buscador de aplicaciones. Busque la consola de comandos como cmd o la consola de preferencia, luego presione click derecho, esto le abrirá una panel a la derecha, seleccione en ejecutar como administrador:



1. Una vez abierta la consola ejecute el siguiente comando para entrar a cronos.

```
C:\Program Files\PuTTY\PuTTY.exe -ssh proxyjump@cronos.eafit.edu.co
```

2. Luego dentro de cronos ejecute el siguiente comando para acceder al nodo que desee. Tenga en cuenta que <ip> se refiere a la ip del nodo seleccionado:

```
ssh root@<ip>
```

El proceso se hace en dos pasos pero la autenticación sigue siendo igual que en Linux/Mac.

Recomendaciones para la administración de los clusters:

Backups:

Cronos es un cluster que lleva alrededor de 10 años en funcionamiento, que es demasiado para un supercomputador. Debido a esto es posible que los nodos fallen sin posibilidad de encenderlos de nuevo, perdiendo la información del disco y el progreso de cada equipo. Por esto les recomendamos hacer backups de su trabajo regularmente (p. ej. diariamente) en una carpeta especial para cada equipo llamada *cios_hpccourse_cluster[1-5]* en el servidor *cronos.eafit.edu.co* reservadas al usuario *proxyjump*.

Por ejemplo para hacer backup de la carpeta *shared* de *compute-1-1*:

```
[proxyjump@cronos ~]$ rsync -azP --delete root@10.150.7.200:/shared
cios_hpccourse_cluster1
```

Para más detalles sobre *rsync* se pueden referir a [este tutorial](#). Para automatizar esta operación se pueden referir a [este tutorial](#).

Configuración del cluster:

Los siguientes puntos son opcionales (excepto la configuración de la red infiniband); muchos de ustedes los encontrarán de gran ayuda, y esperamos que cada uno de ustedes dedique tiempo en determinar la mejor configuración para maximizar el performance de su cluster:

1. **Infiniband:** Los servidores cuentan con 2 interfaces de red principales; una Ethernet y una Infiniband. Para entender mejor las redes infiniband se pueden referir a [este link](#). Los drivers oficiales de las interfaces Mellanox MT27500 Family [ConnectX-3] se pueden encontrar en [este link](#). Las redes de alta velocidad impactan radicalmente el performance de aplicaciones HPC.
2. **Sistema de archivos compartido:** En el supercómputo los nodos deben sincronizar alguna parte de sus sistemas de archivos para que las aplicaciones HPC que se ejecutan en varios nodos tengan acceso a la misma información. [NFS](#) le permite a un servidor NFS compartir directorios a otros nodos por red. Es una solución común y fácil de usar. Recomendamos el uso de NFS para sincronizar directorios entre los 2 nodos de cada cluster. Hay otras soluciones que pueden o no dar mejor performance cuando se tienen aplicaciones intensivas en escritura a disco, e invitamos a explorarlas: [GlusterFS](#) y [Lustre](#), entre otras.
3. **Administrador de recursos:** En un supercomputador con muchos usuarios su uso es esencial. Cuando se ejecutan varias aplicaciones al mismo tiempo en un computador, el performance de cada aplicación es menor a si se ejecutan consecutivamente (una después de la otra); las aplicaciones se contaminan el performance entre sí. En aplicaciones de HPC esto toma mayor relevancia, porque siempre se busca que el performance sea el máximo. Los administradores de recursos son programas que se encargan de que haya la menor contaminación posible, al aislar los recursos que cada aplicación usa. El administrador de recursos más usado es [Slurm](#), y recomendamos su uso. Hay otras soluciones como [OpenPBS](#).