



Introducción al uso del cluster de cómputo ADA

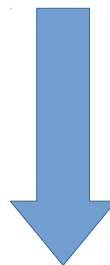
Ing. Luis Alberto Aguilar



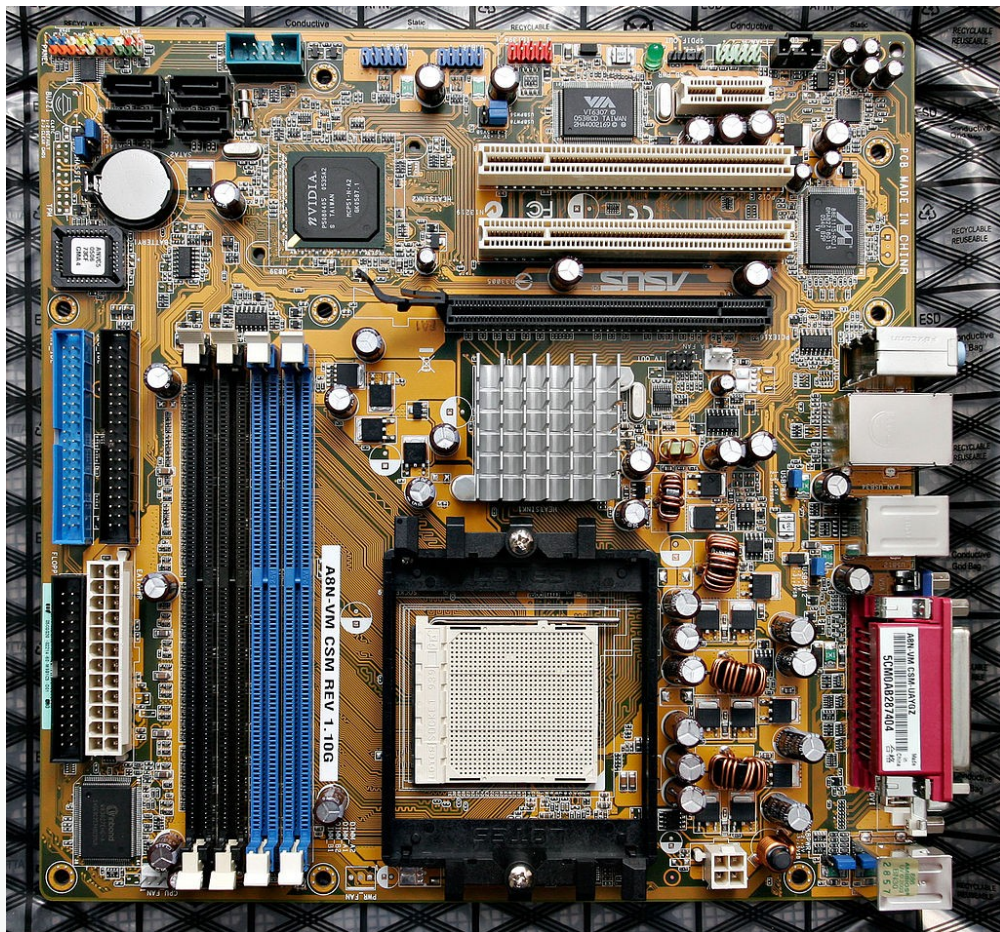
Temario

- ¿Cómo funciona una computadora?
- ¿Qué es el SO?
- ¿Qué es un cluster?
- ¿Cómo usar los recursos del cluster?
- ¿Cómo y dónde guardar archivos?
- ¿Cómo usar los programas del cluster?

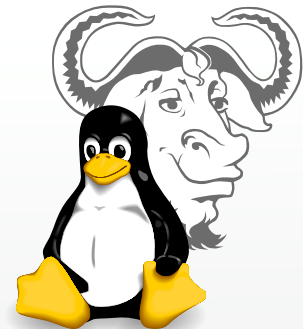
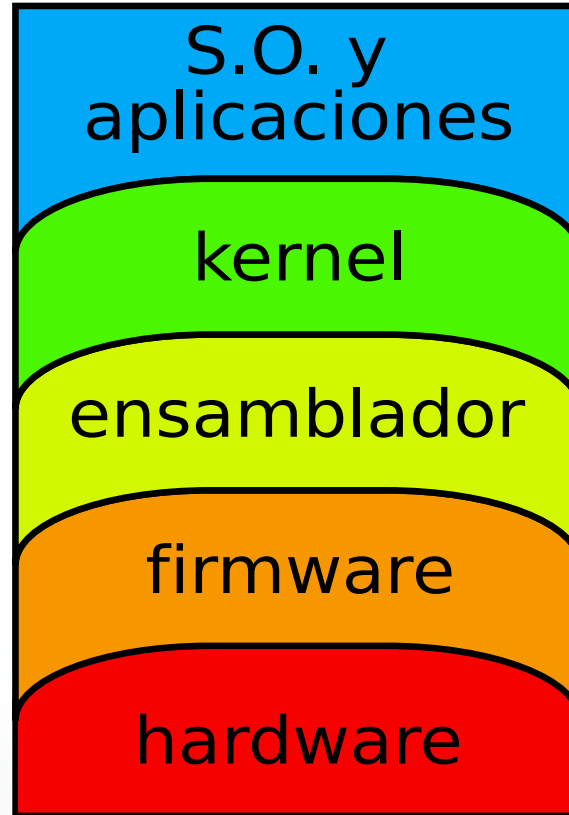
¿Cómo funciona una computadora?



¿Cómo funciona una computadora?

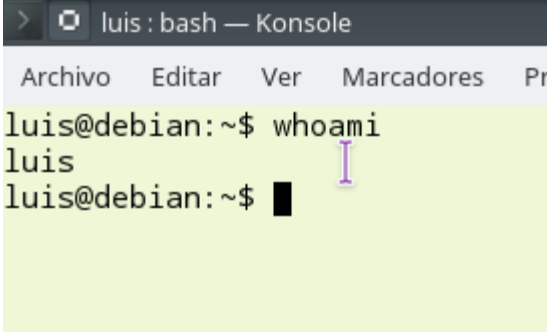


¿Qué es el Sistema Operativo?



GNU/Linux

- Interfaz gráfica y terminal



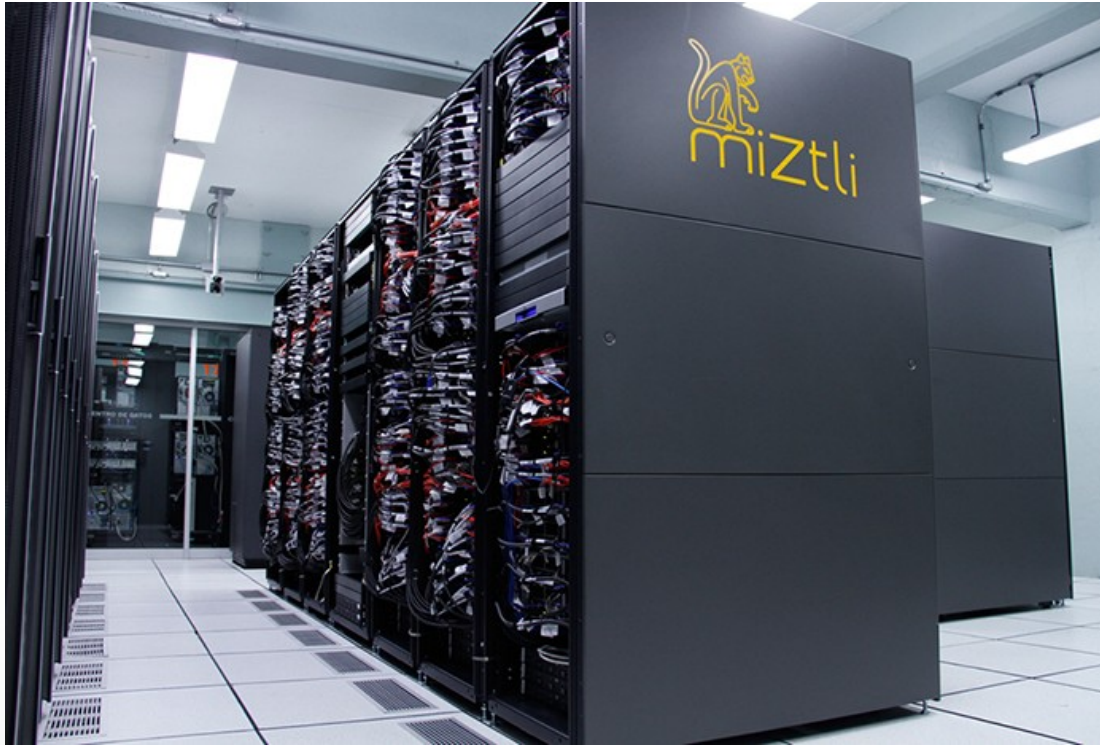
```
luis@debian:~$ whoami
luis
luis@debian:~$
```

- Comandos básicos:

man cat ls cd touch

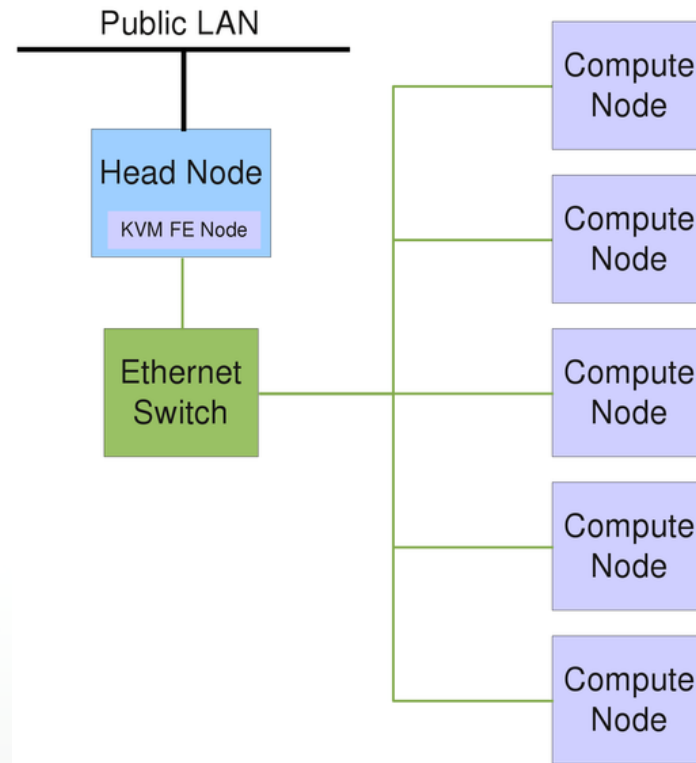
mkdir cp mv rm pwd **nano**

¿Qué es un cluster de supercómputo?



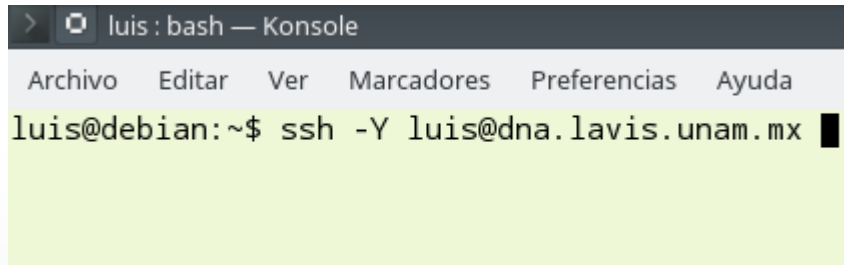
- Unión de computadoras:
- Núcleos de cómputo
- Memoria RAM
- Almacenamiento
- +
- Red de alta velocidad

Esquema de un cluster de supercómputo



¿Cómo acceder a DNA?

- Creación de cuenta de usuario
- Tener acceso a alguna red de la UNAM o acceso a VPN del LAVIS
- Acceder por ssh `ada.lavis.unam.mx`

A terminal window titled 'luis : bash — Konsole'. It has a menu bar with 'Archivo', 'Editar', 'Ver', 'Marcadores', 'Preferencias', and 'Ayuda'. The command prompt shows 'luis@debian:~\$' followed by the command 'ssh -Y luis@dna.lavis.unam.mx' and a black cursor. The terminal background is light green.

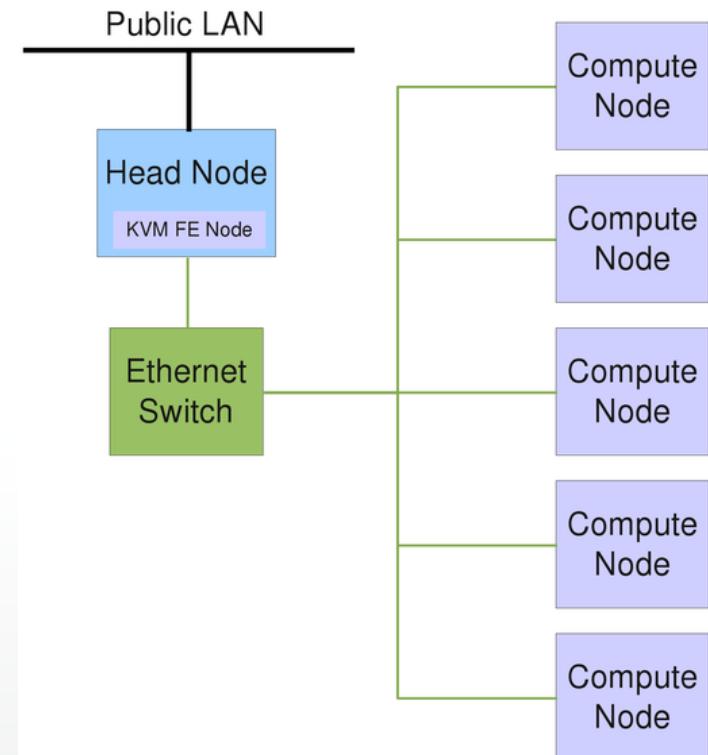
```
> luis : bash — Konsole
Archivo  Editar  Ver  Marcadores  Preferencias  Ayuda
luis@debian:~$ ssh -Y luis@dna.lavis.unam.mx
```

¿Cómo acceder a los recursos del cluster?

- Manejadores de recursos: TORQUE, PBS, LSF, etc
- ADA cuenta con:

SGE

- Núcleos de cómputo (procesadores)
- Memoria RAM





Jobs interactivos: qlogin

- Utilizar los recursos de los nodos de cómputo en lugar del nodo principal

```
(luis) dna.liigh.unam.mx — Konsole
Archivo  Editar  Ver  Bookmarks  Preferencias  Ayuda
[luis@protein ~]$ qlogin
Your job 8994 ("QLOGIN") has been submitted
waiting for interactive job to be scheduled ...
Your interactive job 8994 has been successfully scheduled.
Establishing /cm/shared/apps/sge/var/cm/qlogin_wrapper session to host compute-00-02.cm.cluster ...
Last login: Tue Dec 13 18:11:25 2016 from protein.cm.cluster
[luis@compute-00-02 ~]$
```



Elementos de SGE

- **Job:** Es el proceso a ejecutar en el cluster
- **Id:** Es un identificador numérico único que SGE le asigna a los jobs
- **Cola:** Literalmente es una fila donde los jobs toman un lugar para entrar a ejecución
- **Host:** Equipo de cómputo perteneciente a una cola



Comandos básicos de SGE

- **qstat:** Muestra los jobs en ejecución del usuario

```
[luis@protein ~]$ qstat
```

job-ID	prior	name	user	state	submit/start at	queue	slots	ja-task-ID
8985	0.50500	QLOGIN	luis	r	12/13/2016 17:54:11	all.q@compute-00-02.cm.cluster	1	

```
[luis@protein ~]$ █
```




Comandos básicos de SGE

- `qsub` + nombre de archivo: Envía un nuevo job a SGE

```
[luis@protein ~]$ qsub mi_script.sge
```

- `qdel` + id de job: Elimina un job de la cola

```
[luis@protein ~]$ qdel 8986  
luis has registered the job 8986 for deletion  
[luis@protein ~]$
```

Comandos básicos de SGE

- qquota: Se usa para verificar la cuota asignada sobre los recursos del cluster

```
[luis@protein ~]$ qquota
resource quota rule limit          filter
-----
Max_slots/1          slots=1/60    users luis
[luis@protein ~]$
```

- qghost: Muestra el uso de los recursos en los equipos de cómputo

```
[luis@protein ~]$ qghost
```

HOSTNAME	ARCH	NCPU	LOAD	MEMTOT	MEMUSE	SWAPT0	SWAPUS
global	-	-	-	-	-	-	-
compute-00-00	linux-x64	31	0.04	125.9G	3.3G	0.0	0.0
compute-00-01	linux-x64	64	11.28	251.9G	145.9G	0.0	0.0
compute-00-02	linux-x64	64	5.37	251.9G	113.9G	0.0	0.0
compute-00-03	linux-x64	64	8.83	251.9G	118.3G	0.0	0.0
compute-00-04	linux-x64	64	6.51	251.9G	60.1G	0.0	0.0
compute-00-05	linux-x64	64	-	251.9G	-	0.0	-

```
[luis@protein ~]$
```

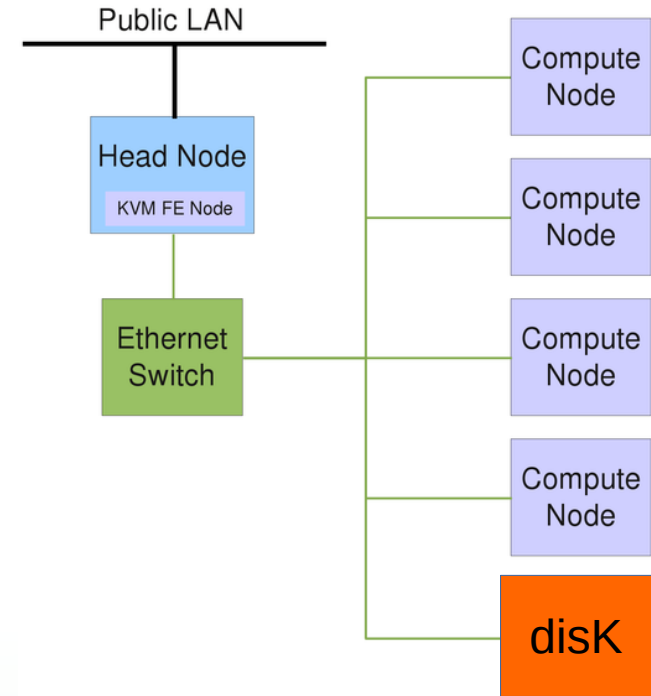


Ejecución de procesos con SGE

- Escribir las propiedades del job en un archivo de texto
- Enviar a SGE el archivo de texto desde **el nodo principal**
- El job entra a ejecución si hay recursos disponibles, o queda en espera de liberación de recursos.
- Termina el job y libera los recursos para otros jobs

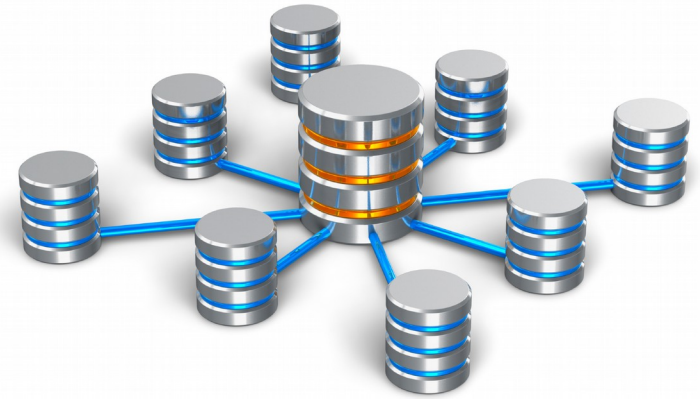
Almacenamiento en DNA

- Almacenamiento compartido:
- Directorio home
- /home/luis (máximo 1GB)
- Directorio con investigador
- /mnt/MD1200A/luis/jsgracia
- /mnt/MD1200/luis/jsgracia



Cuotas de almacenamiento

- Cuota de usuario:
- `quota -s`
- Cuota de grupo:
- `quota -gs`



Módulos de entorno (Programas)

- Los programas se instalan en un directorio compartido
- Son accesibles desde cualquier nodo dentro del cluster
- Permiten tener varias versiones de un mismo programa





Uso de módulos (qlogin)

Comando	Comportamiento
module load	Agrega un modulo a la sesión actual
module unload	Elimina un módulo de la sesión actual
module avail	Muestra todos los módulos o programas disponibles en el cluster
module list	Muestra los módulos que están activos en la sesión actual
module initadd	Carga un módulo al inicio de sesión (útil para módulos que se usan de forma frecuente)
module initrm	Borra un modulo de los inicios de sesión



Script de SGE

```
#!/bin/bash
# Use current working directory
#$ -cwd
#
# Join stdout and stderr
#$ -j y
#
# Run job through bash shell
#$ -S /bin/bash
#
# You can edit the script since this line
#
# Your job name
#$ -N Luis_job
#
# Send an email after the job has finished
#$ -m e
#$ -M laguiar@liigh.unam.mx
#
# If modules are needed, source modules environment (Do not delete the next line):
. /etc/profile.d/modules.sh
#
# Add any modules you might require:
module load bowtie2/2.2.6
#
# Write your commands in the next line
bowtie2 --sensitive -x Escherichia_coli_K12_MG1655/Escherichia_coli_K12_MG1655 -U SRX189773_FNR_ChIP.fastq -S
SRX189773_FNR_ChIP.sam
```



Para comentar...

- ¿Qué es el SO?
- ¿Cuál SO tiene tu computadora?, ¿y el cluster?
- ¿Cuál gestor de recursos tenemos en el cluster y qué recursos administra?
- ¿Cuáles programas utilizarás en el cluster?
- ¿Qué temas te gustaría conocer más a detalle?



¿Dónde buscar recursos?

- Blogs de linux
- Páginas de programas bioinformáticos
- Comando “man”
- Cursos en línea (coursera, edx)