



SISTEMAS INFORMATICOS EUROPEOS S.L.

LADONoS HPC S.O. PROYECT

LADONoS 7

v8

HPC ENVIRONMENT



[WWW.SIE.ES](http://WWW.SIE.ES)  
[WWW.LADONOS.ORG](http://WWW.LADONOS.ORG)

@HPCSIE  
@LADON\_OS





# POTENTE, ESTABLE Y LIBRE



LadonOS está basado en distribuciones CentOS. Una variable de código libre de Red Hat. Al utilizar dicha distribución el sistema ofrece una perfecta armonía entre fiabilidad, seguridad y eficiencia. Optimizado para ofrecer un entorno de total estabilidad en Centos 7.X o RedHat 7.x para entornos que deseen sistemas con soporte oficial.

*(wiki)CentOS (Community ENTERprise Operating System) es una bifurcación a nivel binario de la distribución Linux Red Hat Enterprise Linux RHEL, compilado por voluntarios a partir del código fuente publicado por Red Hat.*

Podemos utilizar un gran número de drivers propietarios tales como Infiniband, Intel PHI, GPUS y compiladores CUDA. Ofrece un abanico prácticamente ilimitado de librerías y compiladores.

Todo ello desarrollado íntegramente en software GNU, sin capas propietarias ni de terceros. Lo que permite a LadonOS ser totalmente personalizado. El código de desarrollo es plenamente libre y podrá ser modificado en función de las necesidades del sistema a instalar. Por lo que cada LadonOS instalado se personaliza para el HPC destinado. LadonOS siempre se dará **LLAVE EN MANO**, plenamente configurado.

LadonOS está pensado para hacer de su entorno HPC un centro sencillo de utilizar, con todos los elementos bajo control y gestionado desde un nodo principal o "frontend", el cual se encargará de administrar el resto de nodos.



## ENTORNO DE RED

LadonOS utiliza una red gigabyte o 10Base T para el control y gestión de nodos, así como los servicios de los mismos. Una vez configurado el entorno de servidor, éste instalará software en los nodos a través del sistema de **LadonOSDeploy**

Así mismo, en dicho sistema se incluye una red adicional dedicada al entorno IPMI (Vlan). Con dicha red se obtiene información sobre eventos de nodos, sensores y es posible el uso de un KVM Over Lan, como si estuviéramos delante del nodo afectado. Esta tecnología mejora el orden del cableado, siendo prescindible un entorno KVM físico. La red IPMI puede instalarse por separado si el usuario lo desea, o bien, entornos bounding o fault tolerance con redundancia LAN que mejora el rendimiento.

Es plenamente compatible con redes Infiniband - omniopath, pudiendo gestionar un entorno de cálculo paralelo o archivos distribuidos en varios nodos con un excelente rendimiento.

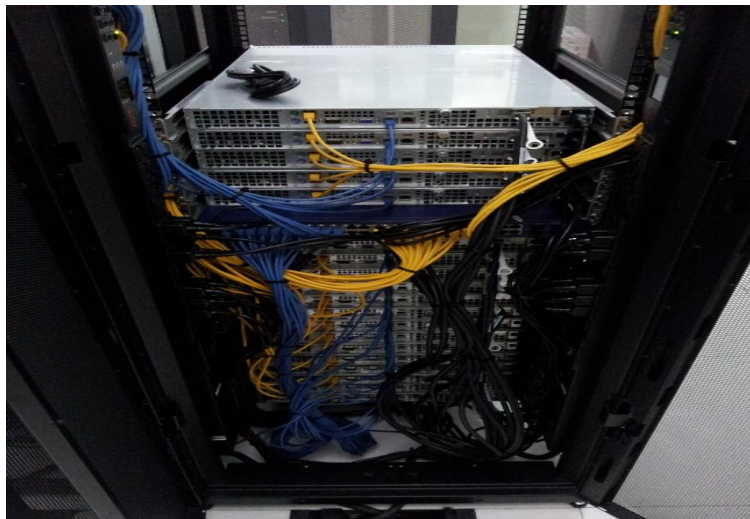


Redes Infiniband / Omnipath

Lan Gestión, PXE e IPMI



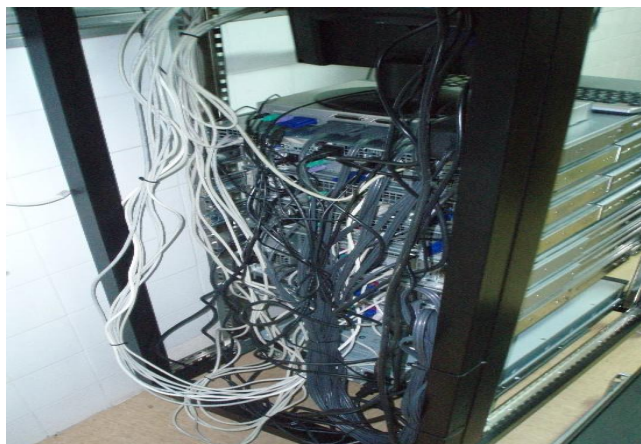
# EJEMPLOS RED LADONos



Ejemplo de LadonOS instalado en la UAB. Con conectividad IPMI dedicada para gestión (cableado amarillo), Lan de gestión de OS e instalación PXE (cableado azul) y sistema Infiniband 40gb/s (cableado negro).

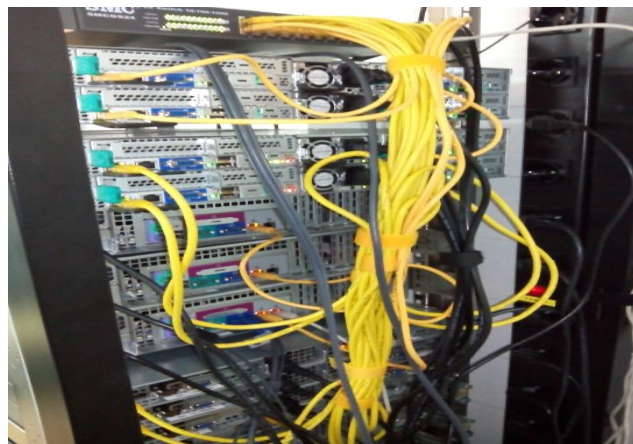


En las imágenes inferiores se puede apreciar la diferencia del mismo HPC usando KVM e IPMI. *Universidad de Alcalá actualizado a LadonOS con IPMI en Enero 2015.*



SISTEMA LAN E IPMI 

 SISTEMA LAN Y KVM



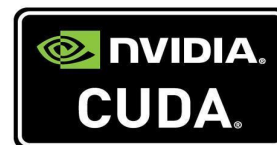
# ENTORNOS GPU

LadonOS es totalmente compatible con entornos de cálculo en GPU.

Ha sido verificado en un sinfín de entornos de producción y actualmente está en plena productividad en diversas instituciones de renombre internacional. La compatibilidad con CUDA y Nvidia-SMI es total, permitiendo estaciones híbridas de CPU+GPU / CPU+PHI / CPU+GPU+PHI



En la imagen de la izquierda se encuentra el HPC del IRB de Barcelona instalado con LadonOS. 15 nodos cálculo en entorno híbrido con 4 GTX Titan Black cada nodo y CUDA 5.5. Dispone de una totalidad de 173280 cores de GPU y 300 cores de CPU.



*LadonOS es plenamente con arquitectura CUDA8*





## ENTORNOS DUALES

El último escenario ha sido instalado en diciembre de 2016. El entorno HPC “Galatea” de la UDG se aproxima a los 2PTflops de rendimiento gracias a sus 22 nodos y sus 176 GTX 1080 especialmente adaptadas a sistemas HPC. Este sistema incluye el sistema LadonOS HA, CLUES, y el sistema de archivos BeeGFS.



Así mismo a nivel de SW. Dispone de automatización IT basado en Ansible, repositorio global local y EasyBuild 3.0.2





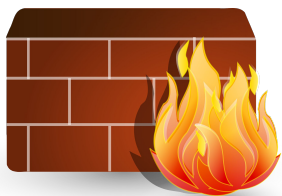
# GESTIÓN Y SEGURIDAD

LadonOS está pensado para entornos de plena producción 24x7. Por este motivo dispone de una serie de herramientas de control, gestión y monitorización para el control de los componentes.

La seguridad ha sido especialmente cuidada: el servidor hace de pasarela web al resto de nodos para la correcta actualización de parches y seguridad. Dispone de servicios de firewall (iptables o firewalld) y entornos de seguridad tales como denyhosts o Fail2ban para evitar ataques de terceros.

El servidor o servidores hacen de pasarela WAN >> LAN (O varias redes). Lo que permite que independientemente del número de nodos. El entorno solo necesite una dirección IP para el acceso de usuarios.

Igualmente el entorno de red es soportado con sistemas Bond y VLAN soportando diferentes protocolos de seguridad en redes.





# USUARIOS Y GRUPOS

LadonOS permite la integración de diversos entornos de usuarios, tales como NIS, Open-Ldap y 389 Directory.

Por defecto se incluye un dominio NIS encargado del manejo de usuarios, grupos y hosts. Con mínimo mantenimiento y excelentes resultados.

Así mismo, ofrecemos la instalación de FreeIPA Server basado en 389 Directory (Ldap) con importantes elementos de control, gestión y monitorización. Las principales ventajas que ofrece son las siguientes:

- Sencillo manejo de usuarios y grupos con políticas dedicadas.
- Interface WEB para manejo de todos los servicios.
- Integra servicios de DNS para la gestión de Hosts.
- Integra servicio de certificado Dogtag.
- Dispone de cliente de fácil instalación para nodos basado en SSSD.
- Servicios MIT Kerberos y servidor NTP.
- Integración con Active Directory.

The screenshot shows the freeIPA web interface. The top navigation bar includes 'Identity', 'Policy', 'Authentication', 'Network Services', and 'IPA Server'. Below this, a secondary navigation bar shows 'Users', 'User Groups' (highlighted), 'Hosts', 'Host Groups', 'Netgroups', 'Services', and 'Automen'. The main content area is titled 'User Groups' and contains a search box and a table of user groups.



## User Groups

Search

<input type="checkbox"/>	Group name	GID	Description
<input type="checkbox"/>	admins	147400000	Account administrators group
<input type="checkbox"/>	editors	147400002	Limited admins who can edit oth
<input type="checkbox"/>	ipausers		Default group for all users
<input type="checkbox"/>	trust admins		Trusts administrators group

Showing 1 to 4 of 4 entries.

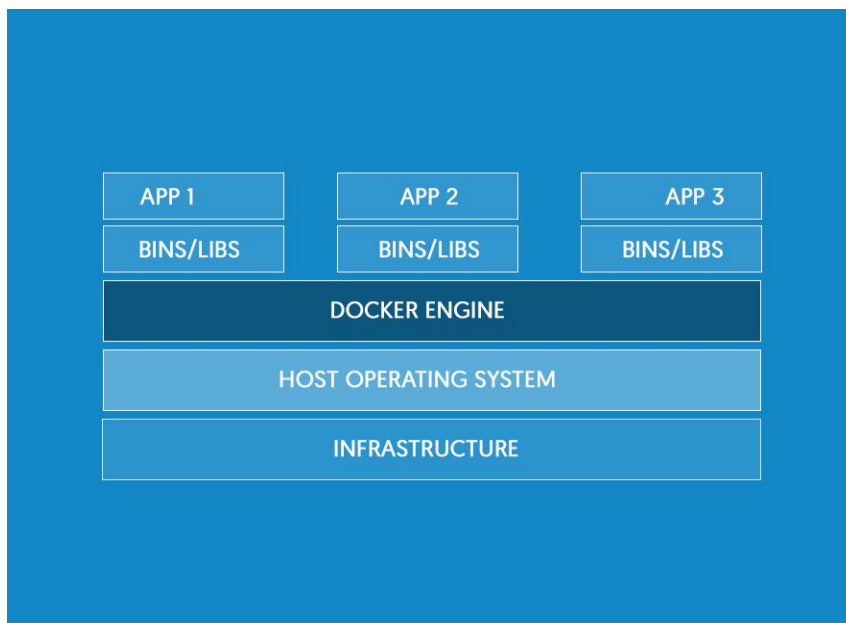


# CONTAINERS



LadonOS v8 permite instalar docker en cualquiera de sus servidores y nodos a través de repositorios oficiales.

Su instalación en el entorno lleva pocos minutos y ofrece un enorme abanico de posibilidades. Permitiendo crear diversos containers en función de determinadas necesidades de aplicaciones o servicios.





# REDUNDANCIA



Para la estructura de servidores redundados se implantará la solución LADONoS 7LHA. La cual dispone de una arquitectura en alta disponibilidad en plataformas de virtualización XEN. Gracias a esta arquitectura, los servicios serán segmentados en diferentes máquinas virtuales minimizando impacto de servicios y optimizando los recursos. Sus características principales son:

- Sistema Citrix Xen Server7 (Últimos parches y actualizaciones aplicadas)
- Entorno POOL, con servidor Maestro – Esclavo.
- VM instaladas en nodo esclavo, en caso de fallo del mismo automáticamente migran al servidor maestro.
- Sistema DRBD-ISCSI entre los nodos montado por red Bond dedicada o 10BT optimizando el clonado de disco en tiempo real.
- Posibilidad de migrar nodo esclavo a maestro en caso de error del nodo maestro, permitiendo la continuidad de la producción.
- Creación de snapshots de máquinas virtuales, exportación, migración y clonado. Así como posibilidad de exportar ficheros a entornos de backup.
- Posibilidad de herramienta de gestión Windows / MAC y OpenXenManager



# REDUNDANCIA

Si en un futuro se realiza una actualización de HW la migración es muy sencilla y rápida. Lo que evita la posibilidad de un largo corte de productividad.

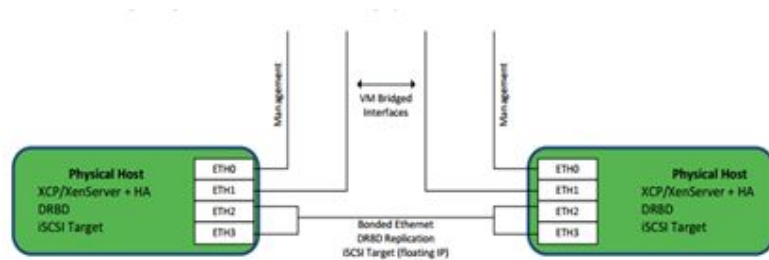
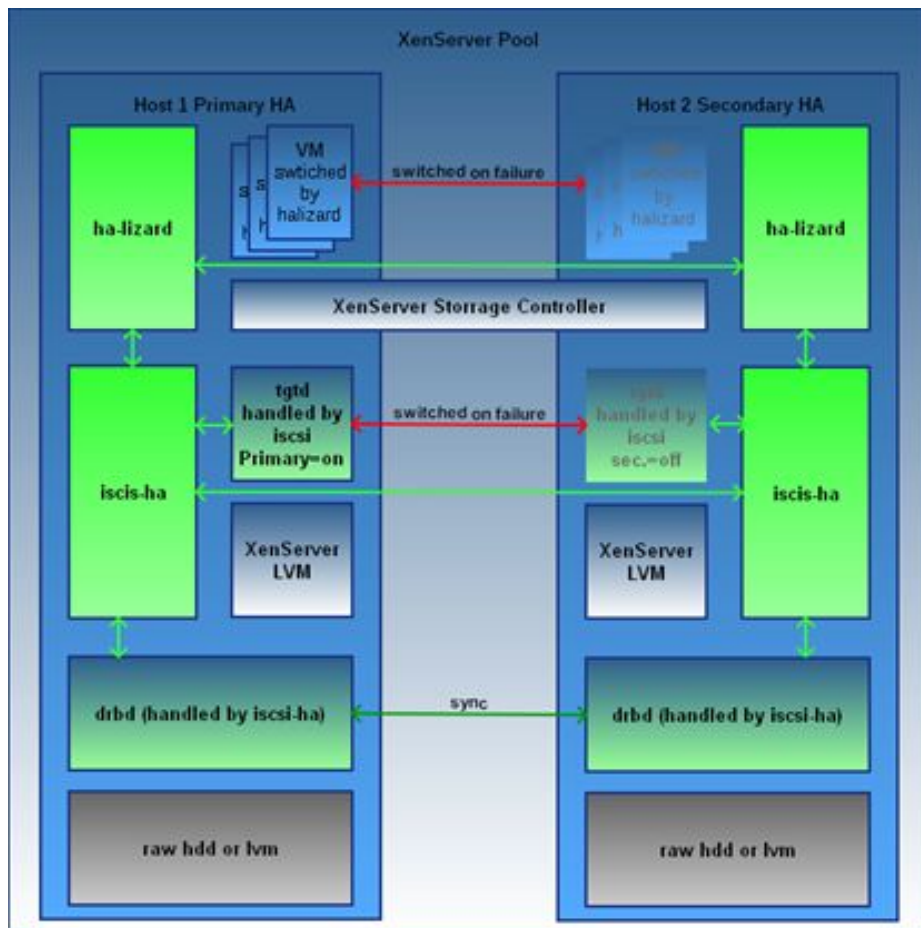
- Es posible disponer igualmente de un entorno de almacenamiento externo vía NFS o iscsi. Dicho almacenamiento se puede utilizar para programar backups de las VM con una fecha determinada o como almacenamiento principal (Requiere un consumo elevado de red).
- Permite exportación de backups a través de snapshots.
- Ofrece consolas GUI y CLI
- Permite hasta 4 puntos de redundancia. Físicos y lógicos, permitiendo la instalación de servidores virtuales Maestros – Esclavos en los diferentes servidores. Por ejemplo
 

o	Servidor		Virtual		1
	o	Servidor	Usuarios	Freelpa	Server
		o	Principal		de
					Slurm
	o	Servidor	Virtual		2
		o	Usuarios	Freelpa	Server
		o	Secundario		Secundario
					Slurm

Para la instalación del entorno se instalarán en los servidores Citrix Xen Server 7 con sistema HA-Lizard. El cual implanta un sistema DRBD-ISCI entre los dos servidores por una red dedicada a tal efecto para su sincronización en tiempo real. Igualmente se puede disponer de otro almacenamiento exterior para poder disponer de backups de las máquinas virtuales o bien usar dicho almacenamiento con otros fines a convenir.



# REDUNDANCIA



TODAY DEMO!



# CONTROL Y MONITORIZACIÓN

Diversas utilidades de control y monitorización son instaladas por defectos en los entornos HPC LadonOS:

- Monitorización Ganglia: Muestra la carga del sistema global y por nodo. Muy sencilla y funcional:
- Check\_MK. Gran utilidad basada en Nagios con numerosos elementos de control y monitorización. Con envío de alertas y servicios SNMP

Así mismo LadonOS es compatible con diversos sistemas: Nagios, Zabbix, Icinga, Cacti, etc..



The screenshot displays the Check\_MK web interface. The main overview includes:

- Host Statistics:** Up: 3, Down: 0, Unreachable: 0, In Downtime: 0, Total: 3.
- Service Statistics:** OK: 97, In Downtime: 0, On Down host: 0, Warning: 2, Unknown: 2, Critical: 5, Total: 108.
- Host Problems (unhandled):** A table with columns for state, host, icons, age, and status detail.
- Service Problems (unhandled):** A table with columns for state, host, service, icons, and status detail. It lists several critical issues, such as 'CUPS Queue tax' and 'CUPS Queue HP\_Color\_LaserJet\_4700'.
- Events of recent 4 hours:** A log of recent events, including interface status changes and OMD health performance.

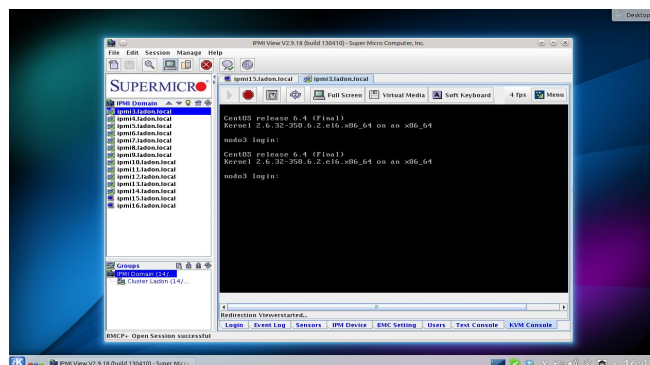


## CONTROL IPMI

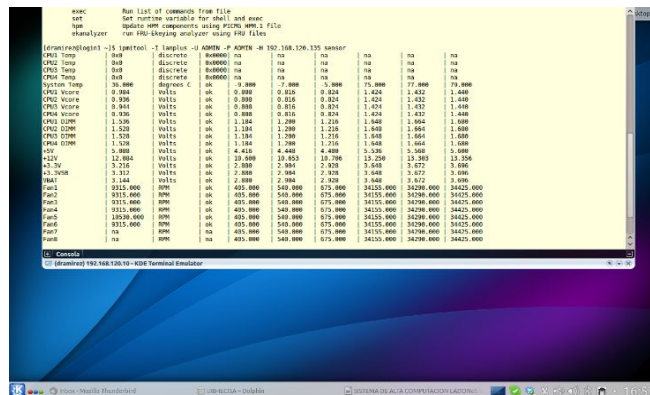
LadonOS ha sido totalmente desarrollado para el uso de tecnología IPMI. El entorno de instalación PXE, la conectividad de nodos, la lectura de sensores, la carga de imágenes.... Todo ha sido configurado para el uso remoto con tecnología KVMOverLAN o SOL (Serial Over Lan). El nodo maestro es el encargado de la gestión IPMI del resto de nodos, siendo este independiente, con una conexión dedicada para el manejo del mismo.

IPMI permite la lectura de sensores, apagado y encendido de máquina, visor de eventos y configuraciones de BIOS desde el ordenador remoto. Sin necesidad de desplazamientos al CPD o servidor físico.

# KVM



# SOL





# GESTIÓN CENTRALIZADA

El servidor de LadonOS, será el encargado del manejo en conjunto del sistema HPC. Dispone de los servicios necesarios para el correcto funcionamiento del entorno. Siempre apostamos por soluciones sencillas y fiables, que cumplan todas las funciones necesarias.

Además, en entornos críticos, se ofrece la posibilidad de hacer instalaciones en diversos servidores con tecnología HA, servicios maestro-esclavo, réplicas o entornos de virtualización para servicios dedicados por VM, con posibilidad de migrar, snapshot, HA, etc...

El servidor dispone de las siguientes soluciones para la gestión del entorno HPC entre otras.

Directorio de Usuarios	Directorio de HOSTS
Directorios compartidos NFS	Directorios compartidos GFS
Routing y NAT Nodos	Servidor WEB
Servicio de Gestor de Colas	Servicio de DMZ Lan
Servicio de IPMI (Nodos)	Servicio de PXE
Servicio FirewallID	Servicio Backup
Librerías X11	Servicio Log
Servicio OFED IB	Servicio Update

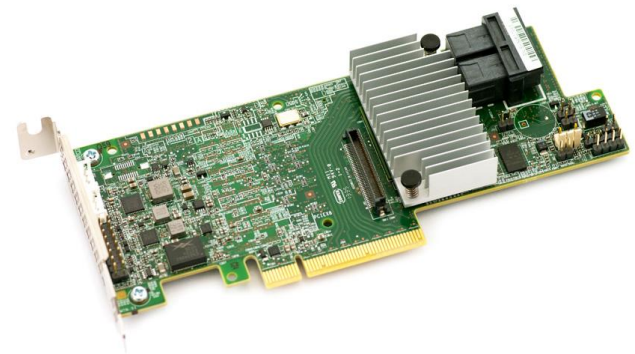
# GESTIÓN DE FICHEROS

LadonOS dispone de diversos directorios exportados por NFS, dichos directorios son utilizados para la instalación y compilación de programas en el entorno HPC, de este modo el resto de nodos podrá disponer de ellos a la hora de ejecutar programas y cargar librerías. El entorno usado es NFS o GlusterFS (tanto por lan, 10G o Infiniband) dada su estabilidad.

En entornos de cálculo en Paralelo y sistema de ficheros distribuido de alto rendimiento se instala la solución BeeGFS.

[http://www.beegfs.com/docs/Introduction to BeeGFS by ThinkParQ.pdf](http://www.beegfs.com/docs/Introduction%20to%20BeeGFS%20by%20ThinkParQ.pdf)

[http://www.beegfs.com/docs/BeeGFS\\_Flyer.pdf](http://www.beegfs.com/docs/BeeGFS_Flyer.pdf)



**SIE ES PARTNER OFICIAL DE BEEGFS Y OFRECE EL SOPORTE DE 1er Y 2º Nivel en todo el territorio nacional.**



# BEEGFS

BeeGFS es el sistema de ficheros de alto rendimiento del Centro de Computación de Fraunhofer. La arquitectura distribuida de metadatos BeeGFS ha sido diseñado para proporcionar la escalabilidad y la flexibilidad que se requiere para ejecutar aplicaciones HPC más exigentes de hoy en día.

## **Sistema Distribuido de Almacenamiento y Metadatos**

La división de sistema de almacenamiento y metadatos evita importantes cuellos de botella. Igualmente el sistema Striping permite que varios servidores puedan heredar dichos roles, aumentando el performing y los IOPS, Los grandes sistemas se benefician enormemente de estos sistemas gracias a los múltiples servidores de metadatos.

## **Tecnología HPC**

BeeGFS no requiere parches del Kernell, los componentes son fácilmente instalables gracias a sus herramientas de gestión. Igualmente permite añadir más clientes y servidores en el sistema HPC siempre que se desee. Así mismo el rendimiento es excepcional dado que dispone de protocolo nativo RDMA. En caso de no disponer de infiniband, el rendimiento en 10G es sobresaliente.

## **BeeGFS On Demand**

En cálculos puntuales, es posible generar un entorno dedicado sumando particiones de diversos nodos para obtener un sistema distribuido temporal para cálculos específicos. Todo ello se realiza de un modo inmediato, siendo de un manejo extremadamente sencillo para el usuario..

## **Cliente y servidores en cualquier máquina**

A diferencia de otros sistemas como Lustre, BeeGFS no requiere hardware específico, incluso en pequeños entornos los servidores puede efectuar funciones de cliente en pequeños entornos de HPC. Procurando Storage y Metadatos en el mismo servidor, ahorrando importantes costes de HW.

## **Gran aumento de coherencia**

Comparado con el sistema NFS los cambios son inmediatamente visibles, lo que garantiza un aumento de coherencia y concurrencia.



# SLURM WORKLOAD MANAGER



Slurm es un gestor de colas actual y extremadamente potente, diseñado para un total control y optimización de recursos de los entornos HPC.

Slurm ofrece entre otras opciones.

- Escalabilidad: Está diseñado para operar en un cluster heterogéneo con hasta decenas de millones de procesadores.
- Rendimiento: Se puede ejecutar 500 trabajos simples por segundo (dependiendo de la configuración del hardware y del sistema).
- Libre y Open Source: Su código fuente está disponible libremente bajo la Licencia Pública General de GNU.
- Portabilidad: Slurm es compatible con un amplio entorno de lenguajes.
- Administración de energía: Cada trabajo puede especificar su frecuencia de la CPU y la potencia deseada por el uso de trabajos. Los recursos que no sean usados pueden ser apagados hasta su requerimiento.
- Tolerancia a fallos.
- Integra componentes MPI.
- Mejora notablemente la estructura HW, aprovechando el 100% de recursos.
- Trabajos modificables bajo demanda. Permite asignar mayor número de recursos a un trabajo “en caliente”
- Soporte profesional: SLURM dispone del soporte profesional ofrecido por SchedMD

Partition	Default	Part State	Time Limit	Node Count	Node State	NodeList
admin	no	up	infinite	82		n[3-84]
develop	yes	up	00:31:00	2	idle	n[1-2]
long_term	no	up	5-01:00:00	82		n[3-84]
parallel	no	up	23:30:00	82		n[3-84]
performance	no	up	infinite	82		n[3-84]
serial	no	up	23:30:00	82		n[3-84]
testing	no	up	infinite	82		n[3-84]

# TOMORROW DEMO



# LIBRERÍAS Y COMPILADORES

Dependiendo del uso, LadonOS será instalado y compilado en función de los programas o deseos de los usuarios finales. Sin embargo se entregan siempre una serie de elementos comunes plenamente configurados para su inmediata puesta en marcha. Todo el entorno HPC queda en óptimo funcionamiento en el instante de la instalación.

LadonOS incluye entre otros:

<b>GCC , C++ y Boost ++</b>	<b>Atlas, Lapack, Scalapack</b>
<b>HDF5</b>	<b>NetCDF</b>
<b>GSL</b>	<b>FFTW</b>
<b>ATLAS</b>	<b>OpenMP</b>
<b>Python</b>	<b>GnuPlot</b>
<b>OpenMPI</b>	<b>Hwloc</b>
<b>Valgrind</b>	<b>StdC</b>
<b>Glibc</b>	<b>QT</b>

Aún así, si se requiere mayor o menor número de librerías, éstas serán configuradas. Gracias a los directorios de exportación y a los sources compartidos, se pueden propagar en conjunto HPC sin problemas y con una gran escalabilidad



# EASYBUILD



LadonOS es compatible con el desarrollo EasyBuild. El cual dispone de un gigantesco catálogo de software fácilmente instalable y configurable. Lo que permite a través de un entorno de módulos compilar e instalar diferentes versiones de software

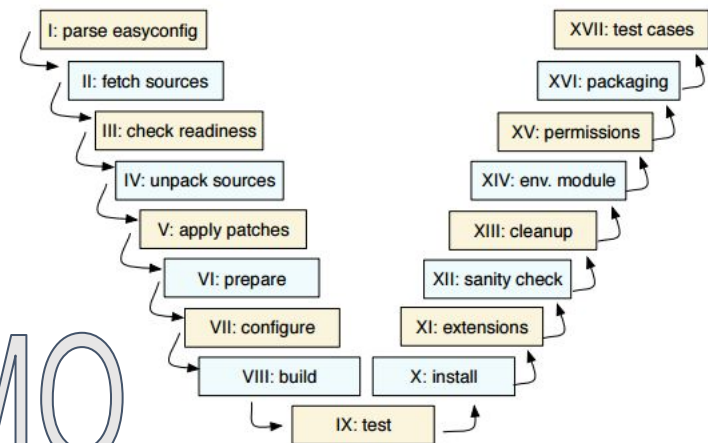
Numerosos programas han sido instalados en perfectas condiciones en los entornos LadonOS de SIE. Citamos algunos de ellos:

CP2K, GAMESS-US, GROMACS, NAMD, NWChem, OpenFOAM, PETSc, Quantum, ESPRESSO, WRF, WPS, . . .

La totalidad supera más de 800 paquetes de software

Más información en:

<https://hpcugent.github.io/easybuild/>



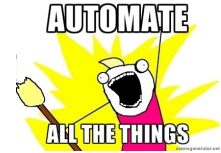
# TOMORROW DEMO



## LADONos DEPLOY



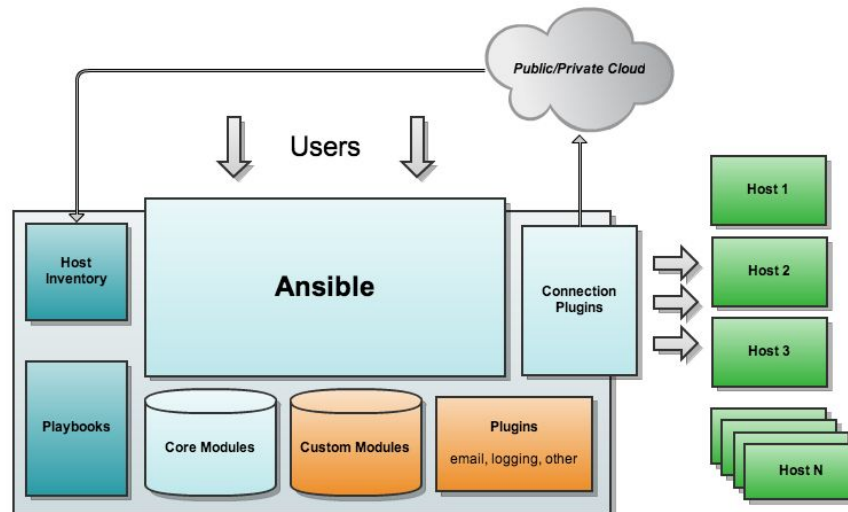
## TOMORROW DEMO



LadonOS dispone de un sistema de automatización IT basado en ANSIBLE. A través de los PLAYBOOKS o GALAXIAS (Imagen de configuración de entorno). Su funcionamiento es similar al clonado de imágenes. Creando un entorno de configuración inicial para seguidamente aplicar de modos simultáneo en los nodos incorporados. O bien actualizar todo el sistema a un escenario común..

Gracias a esta utilidad todos los nodos pueden tener los mismos componentes de configuración, programas, librerías, servicios, etc... Sin necesidad de ir nodo por nodo reconfigurando. Lanzando actualizaciones y propagaciones en cuestión de segundos. Pudiendo seleccionar un elevado número de operaciones de implementación

A diferencia de los sistema de clonado. Ansible se encarga de mantener el cluster homogéneo simplemente actualizando un fichero en función de nuestras necesidades. Creando un PlayBook maestro (Imagen de configuración y servicios) que será implementada en todo el entorno HPC





# AUTOMATIZACIÓN (BETA)



## FOREMAN

En la nueva versión v8 se incluye opcionalmente “The Foreman” como sistema de instalación a través de red. Siendo por defecto el sistema KickStart.

The Foreman es un entorno que une Kickstart + Ansible. Lo que permite disponer de los nodos plenamente configurados una vez instalados. Así mismo los repositorios Centos se convierten en locales mejorando considerablemente la velocidad de instalación.

Dispone de importantes utilidades para la configuración de red, tales como:

- Sistema de detección automático de MAC y asignación
- Configuración secuencial de diversas redes. Incluyendo IPMI
- Integrado con ANSIBLE para poder finalizar un nodos plenamente configurado con el playbook correspondiente.
- Gestión WEB , control y monitorización de nodos.

