

Software libre en el Máster en Computación de Altas Prestaciones

Patricia González
patricia.gonzalez@udc.es



Índice

Introducción

Máster en Computación de Altas Prestaciones

Contenidos del máster y software libre

Desarrollo de la docencia y software libre

Resumen

Índice

Introducción

Máster en Computación de Altas Prestaciones

Contenidos del máster y software libre

Desarrollo de la docencia y software libre

Resumen

Computación de altas prestaciones. ¿Qué es?

High Performance Computing (HPC)

Uso de supercomputadores, clústers y redes para resolver problemas complejos

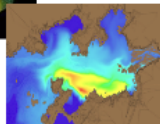
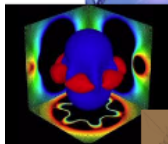
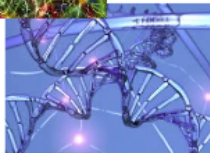
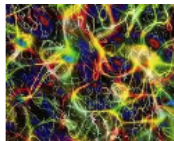
Incluye:

- ▶ Programación paralela
- ▶ Técnicas de optimización
- ▶ Computación distribuida
- ▶ Administración de sistemas, redes y seguridad

Computación de altas prestaciones. ¿Qué es?

HPC en investigación

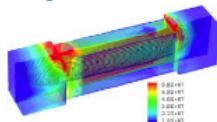
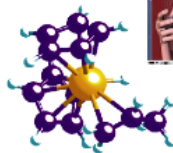
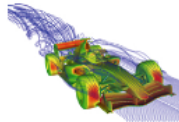
- ▶ Física de altas energías, astrofísica, física de materiales, ...
- ▶ Predicción meteorológica y cambio climático
- ▶ Química computacional
- ▶ Bioinformática, genómica
- ▶ Nanotecnología



Computación de altas prestaciones. ¿Qué es?

HPC en la industria

- ▶ Transporte
- ▶ Industria farmacéutica
- ▶ Industria de materiales y procesos
- ▶ Dispositivos electrónicos
- ▶ Render 3D y visualización
- ▶ Construcciones virtuales
- ▶ ...



Computación de altas prestaciones. ¿Qué es?

Procesamiento de datos y servicios en la *nube*

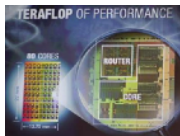
- ▶ Manejo de ingentes cantidades de información: Google, Facebook, Spotify, ...
- ▶ Centros de procesamiento de datos de pequeñas y medianas empresas
- ▶ Sistemas de almacenamiento de altas prestaciones
- ▶ Infraestructuras como servicios
- ▶ ...



Computación de altas prestaciones. ¿Qué es?

HPC en casa

- ▶ La programación secuencial no sirve:
 - ▶ 2, 4, 8 cores y más
 - ▶ AMD Opteron 6100: 12 cores
 - ▶ Intel Polaris: 80 cores
 - ▶ Nvidia Gforce GTX 480: 480 procesadores streaming



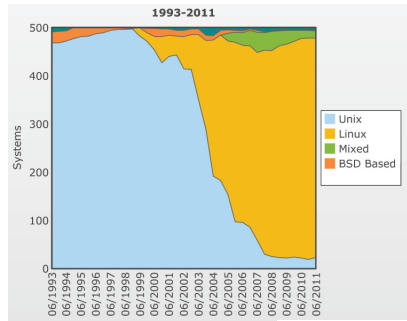
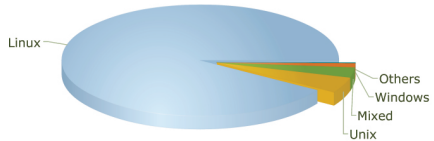
Computación de altas prestaciones y Software Libre

Expansión del SL

- ▶ Linux como plataforma de aplicación
- ▶ Internet como canal de distribución

Top 500

- ▶ El 91 % de los sistemas de la lista Top 500 (06/2011) tiene SO Linux



Índice

Introducción

Máster en Computación de Altas Prestaciones

Contenidos del máster y software libre

Desarrollo de la docencia y software libre

Resumen

Máster Oficial Interuniversitario en Computación de Altas Prestaciones

Organizadores:

- ▶ Universidade de Santiago de Compostela (USC)
- ▶ Universidade da Coruña (UDC)
- ▶ Centro de Supercomputación de Galicia (CESGA)

Objetivos:

- ▶ Formación de profesionales y/o investigadores en el campo del HPC
- ▶ Incorporación a centros tecnológicos o de proceso de datos
- ▶ Trabajo en empresas del sector HPC: HP, Bull, ...
- ▶ Realización de una tesis de doctorado

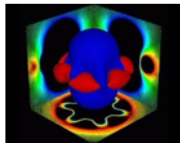
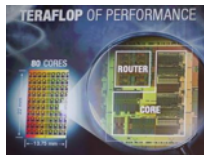
Estructura del Máster

Un curso académico (60 ECTS)

- ▶ 9 materias obligatorias
- ▶ 2 materias optativas
- ▶ PFM

Contenidos:

- ▶ Arquitectura de computadores
- ▶ Programación de aplicaciones HPC
- ▶ Administración de instalaciones HPC
- ▶ Gestión de proyectos HPC



Índice

Introducción

Máster en Computación de Altas Prestaciones

Contenidos del máster y software libre

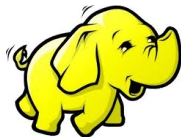
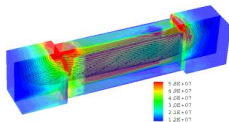
Desarrollo de la docencia y software libre

Resumen

Programación de aplicaciones

Materias

- ▶ Software de base para supercomputación
- ▶ Programación paralela
- ▶ Técnicas de optimización y paralelización
- ▶ Computación científica y simulación de altas prestaciones
- ▶ Depuración de programas y análisis de rendimiento
- ▶ Computación en sistemas distribuidos



Porcentaje

Más de un 50 % de
los contenidos del
máster

Programación de aplicaciones

Software libre usado

- ▶ Construcción de código:
 - ▶ autoconf, automake, libtool (<http://www.gnu.org>)
- ▶ Control de versiones:
 - ▶ GIT(<http://git-scm.com>)
 - ▶ SVN(<http://subversion.apache.org>)
- ▶ Depuración y análisis de rendimiento:
 - ▶ GDB(<http://www.gnu.org>)
 - ▶ libCWD(<http://libcwd.sourceforge.net/>)
 - ▶ Electric Fence/DUMA (<http://duma.sourceforge.net>)
 - ▶ Valgrind (<http://valgrind.org>)
 - ▶ MARMOT (<http://www.hlrs.de/organization/av/amt/projects/marmot>)
 - ▶ PAPI (<http://icl.cs.utk.edu/papi>)
 - ▶ TAU (<http://www.cs.uoregon.edu/Research/tau/home.php>)
- ▶ Programación en sistemas distribuidos:
 - ▶ hadoop (<http://hadoop.apache.org>)

Programación de aplicaciones

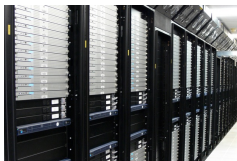
Software libre usado

- ▶ Compiladores:
 - ▶ gcc, gfortran (<http://gcc.gnu.org>)
- ▶ Programación en memoria compartida:
 - ▶ OpenMP (<http://openmp.org>)
- ▶ Programación con pase de mensajes:
 - ▶ OpenMPI (<http://www.open-mpi.org>)
 - ▶ MPICH2 (<http://www.mcs.anl.gov/research/projects/mpich2>)
- ▶ Programas y librerías matemáticas:
 - ▶ Octave (<http://www.gnu.org/software/octave>)
 - ▶ ATLAS (<http://math-atlas.sourceforge.net>)
 - ▶ scaLAPACK (<http://www.netlib.org/scalapack>)
 - ▶ PETSc (<http://www.mcs.anl.gov/petsc/petsc-as>)
 - ▶ HPC-Netlib (<http://www.nhse.org/hpc-netlib>)

Administración de instalaciones

Materias

- ▶ Administración de sistemas I
- ▶ Administración de sistemas II
- ▶ Administración de instalaciones de supercomputación
- ▶ Computación en sistemas distribuidos



Porcentaje

Un 25 % de los contenidos del máster

Administración de instalaciones

Software libre usado

- ▶ Sistemas Operativos:
 - ▶ Debian (<http://www.debian.org>)
 - ▶ CentOS (<http://www.centos.org>)
- ▶ Virtualización:
 - ▶ VirtualBox (<http://www.virtualbox.org>)
 - ▶ Xen (<http://www.cl.cam.ac.uk/research/srg/netos/xen>)
- ▶ Instalación y administración de clusters:
 - ▶ Rocks cluster (<http://www.rocksclusters.org/wordpress>)
- ▶ Gestión de backups:
 - ▶ Amanda (<http://www.amanda.org>)
- ▶ Administración de recursos y usuarios:
 - ▶ OpenLDAP (<http://www.openldap.org>)

Administración de instalaciones

Software libre usado

- ▶ Gestor de colas:
 - ▶ Open Grid Engine (<http://gridscheduler.sourceforge.net/>)
- ▶ Drivers de red (InfiniBand y 10 Gbit Ethernet):
 - ▶ OFED (<http://www.openfabrics.org>)
 - ▶ Open-MX (<http://open-mx.gforge.inria.fr/>)
- ▶ Middleware:
 - ▶ Globus (<http://www.globus.org/>)

Índice

Introducción

Máster en Computación de Altas Prestaciones

Contenidos del máster y software libre

Desarrollo de la docencia y software libre

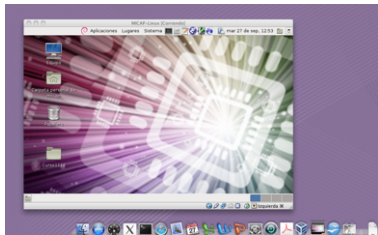
Resumen

Entorno de prácticas

- ▶ Cada alumno acude a las clases con su propio portátil
- ▶ Se usan máquinas virtuales, con distintos objetivos:
 - ▶ unificar el entorno (que todo el mundo esté usando las mismas versiones de las herramientas, por ejemplo)
 - ▶ permitir usar herramientas de simulación en SO diferentes del anfitrión
 - ▶ evitar cometer fallos catastróficos en los equipos (caso de las asignaturas de administración)

Software libre usado

- ▶ VirtualBox
(<http://www.virtualbox.org>)
- ▶ MV con Debian
(<http://www.debian.org/>)



Clases presenciales

La mayoría de las clases se llevan a cabo de forma presencial en dos aulas, una en Santiago de Compostela y otra en A Coruña:

- ▶ En la USC: sala de juntas del Departamento de Electrónica y Computación (DEC)
- ▶ En la UDC: aula 2.2B en la Facultad de Informática (FIC)
- ▶ En el CESGA: aula de formación

Imprescindible

- ▶ Herramientas colaborativas
- ▶ Herramientas que permitan multiconferencia

Plataforma docente virtual

AULA CESGA (<http://aula.cesga.es>)

Basada en Chamilo (<http://www.chamilo.org/>)

- ▶ Compartir documentos
- ▶ Notificaciones
- ▶ Agenda compartida
- ▶ Foros privados por asignatura
- ▶ ...

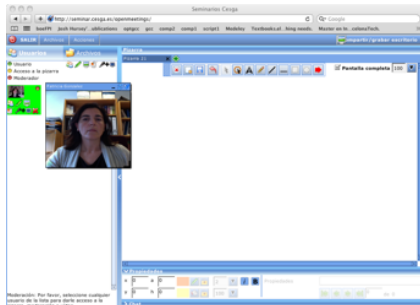


Plataforma de multiconferencia

SEMINAR CESGA (<http://seminar.cesga.es>)

Basada en openmeetings (<http://code.google.com/p/openmeetings/>)

- En horario de clase: para compartir el escritorio, tanto profesores como alumnos
- En horario extraoficial: para impartir tutorías, para reuniones de trabajo en grupo, ...



Otras herramientas

exeLearning (<http://exelearning.org/wiki>)

- ▶ Asistencia para publicación de contenidos docentes en web
- ▶ Creación de unidades didácticas que se importan en Aula Cesga

Computación Grid - HPCM1009 - Aula Cesga

Curso actual

24%

Computación Grid

Construir Vista básica Mostrar

Volgar o listado

Prácticas de Grid ✓

- Preparación de la VM
- Creación de la VM
- Configuración de la VM ✓
- Preparación de Globus ✓
- Instalación de Globus
- Configuración de Globus ✓
- Configuración de la seguridad (I) ✓
- Configuración de la seguridad (II)**
- Configuración de GridFTP
- Configuración de GRAM
- Configuración de las autorizaciones
- Uso de Globus a nivel usuario
- Seguridad
- Transferencia de ficheros
- Ejecución de trabajos

Configuración de la seguridad (II)

Contenido

INSTALACIÓN DE CERTIFICADOS

- Una vez firmadas las solicitudes de certificado por la SimpleCA, es necesario instalarlas en la VM para poder utilizarlas. El primer paso será conseguir los certificados ya firmados. **El profesor le indicará que debe hacer para conseguirlos.** En un entorno de producción lo habitual sería descargarlos desde la página web de la autoridad certificadora siguiendo los pasos indicados por esta y que previamente le habrá comunicado al usuario.
- Mueva estos certificados (de nombres `nombre=apellidos1.hostcert_signed.pem` y `nombre=apellidos1.usercert_signed.pem`) al directorio `$HOME/.globus` del usuario `globus`.

Este paso no es realmente necesario. Se hace para que todos los alumnos tengan inicialmente todos los ficheros relacionados con las credenciales (clave privada y certificado) de `host` y usuario almacenados en el mismo directorio de partida.

- Instale las credenciales de `host`.

```
-- renombramos las credenciales al directorio
/etc/grid-security (como root)
# cd /etc/grid-security
# mv
/home/globus/.globus/nombre=apellidos1.host*

# ls *.pem
nombre=apellidos1.hostkey.pem
nombre=apellidos1.hostcert.pem
nombre=apellidos1.hostcert_request.pem
nombre=apellidos1.hostcert_signed.pem

-- renombramos la clave privada y el
```

Otras herramientas

- ▶ Ofimática:
 - ▶ LibreOffice (<http://www.libreoffice.org/>)
- ▶ Procesamiento de textos (científicos):
 - ▶ LaTeX (<http://www.latex-project.org>)
- ▶ Editar imágenes:
 - ▶ Gimp (<http://www.gimp.org>)
- ▶ Generar gráficas:
 - ▶ Gnuplot (<http://www.gnuplot.info>)
 - ▶ Octave (<http://www.gnu.org/software/octave>)
- ▶ Gestión de proyectos:
 - ▶ OpenProj (<http://sourceforge.net/projects/openproj>)
- ▶ Compartir ficheros:
 - ▶ Mollify (<http://www.mollify.org>)
 - ▶ sparkleshare (<http://sparkleshare.org>)

Índice

Introducción

Máster en Computación de Altas Prestaciones

Contenidos del máster y software libre

Desarrollo de la docencia y software libre

Resumen

Resumen

Uso de software libre en el máster

- ▶ Para el desarrollo de la docencia:
 - ▶ plataformas virtuales
 - ▶ multiconferencia
 - ▶ generación de contenidos docentes
 - ▶ ...
- ▶ En los contenidos impartidos:
 - ▶ Programación de aplicaciones (compitiendo con software propietario por motivos de rendimiento)
 - ▶ Administración de sistemas (en torno al 80 % del software que se usa es SL)

Resumen

Puntos a favor

- ✓ Libertad
- ✓ Disponibilidad
- ✓ Portabilidad
- ✓ Persistencia en el tiempo

Resumen

Puntos a favor

- ✓ Libertad
- ✓ Disponibilidad
- ✓ Portabilidad
- ✓ Persistencia en el tiempo

Puntos en contra

- ✗ Rendimiento ¿?
- ✗ Soporte ¿?

Resumen

Puntos a favor

- ✓ Libertad
- ✓ Disponibilidad
- ✓ Portabilidad
- ✓ Persistencia en el tiempo

Puntos en contra

- ✗ Rendimiento ¿?
 - ✓ El rendimiento de algunas soluciones libres puede ser mayor
- ✗ Soporte ¿?

Resumen

Puntos a favor

- ✓ Libertad
- ✓ Disponibilidad
- ✓ Portabilidad
- ✓ Persistencia en el tiempo

Puntos en contra

- ✗ Rendimiento ¿?
 - ✓ El rendimiento de algunas soluciones libres puede ser mayor
- ✗ Soporte ¿?
 - ✓ Muchas compañías dan soporte de SL
 - ✓ Comunidad que resuelve problemas en horas-días-semanas

Más información sobre el máster

Coordinación del máster:

- ▶ UDC: Patricia González Gómez (patricia.gonzalez@udc.es)
- ▶ USC: Tomás Fernández Pena (tf.pena@usc.es)

Web del máster:

- ▶ <http://gac.udc.es/master>
- ▶ <http://ac.usc.es/master>

Twitter:

- ▶ http://twitter.com/#!/master_icap