Aplicaciones de Supercómputo Distribuido Usando Tecnologías de Grid

Eduardo Murrieta León

José Luis Gordillo Ruiz

Departamento de Supercómputo

DGSCA - UNAM

Contenido

- Objetivos
- Grids computacionales
- Globus
- MPICH
- Zeus - MP
- Experimentos y resultados
- Conclusiones
- Trabajo futuro
- Referencias

Objetivos

- Implementar una infraestructura de grids
  - Balancear la carga de trabajo entre las supercomputadoras
    - AlphaServer, Origin 2000, Cluster
  - Proporcionar recursos para diferentes esquemas de trabajo
    - Aplicaciones de supercómputo distribuido y de alta demanda de trabajo
  - Incorporar equipos a una infraestructura global
    - Supercomputadoras, salas de estaciones de trabajo

Objetivos (2)

- Primer paso: Infraestructura local
  - Instalación y configuración de servicios
  - Adaptación de aplicaciones
- Simulaciones de sistemas astrofísicos
  - Ejecución de Zeus a través de diferentes plataformas
    - Arquitecturas, sistemas operativos
    - Agregar la memoria de diferentes máquinas para realizar experimentos más grandes

Grids computacionales

- Una Grid es una infraestructura para el mejor aprovechamiento de recursos
  - CPUs, memoria, almacenamiento, instrumentos
  - Recursos distribuidos geográficamente
    - Redes eficientes
  - Dominios de administración independientes
    - Compartir recursos sin afectar políticas de uso
    - No existe una administración centralizada

Grids computacionales (2)

- Es útil para
  - Aprovechar recursos ociosos
  - Usar el mismo equipo bajo diferentes esquemas de trabajo
  - Realizar más simulaciones en menos tiempo
  - Realizar simulaciones más grandes
  - Usar equipo *ad hoc* para cada paso de la simulación
  - NO para realizar una simulación en un tiempo menor

Grids computacionales (3)

- Las Grids no son un proyecto nuevo
  - Globus existe desde hace 5 años
  - Metacómputo desde hace 8 años
    - Proyecto I-WAY
  - En DGSCA se ha planteado este proyecto desde 2001
    - Reunión CUDI de otoño
    - Semana de Supercómputo

Grids computacionales (4)

- Las Grids están funcionando
  - Cientos de proyectos relacionados
    - La mayoría utiliza Globus
  - TeraGrid
    - 5 sitios: Pittsburgh, Urbana-Champaign, San Diego, Argonne, Pasadena
    - 20 Teraflops, 1 PetaByte, 40 Gigabits/s
    - 88 millones de dólares

Grids computacionales (5)

- Elementos para usar una grid
  - Servicios: para manejar los recursos de la grid
    - Autenticación, comunicación, disponibilidad, creación de procesos, información
  - Herramientas: para aprovechar los servicios
    - Compiladores, bibliotecas, depuradores, analizadores
  - Aplicaciones
  - Sistemas auxiliares
    - Balanceadores de carga, contabilidad

Globus

- **"Globus es un proyecto que está desarrollando las tecnologías fundamentales para la construcción de GRIDS computacionales"**
- Sus características principales son
  - Funciona como una "bolsa de servicios"
    - No adopción de un modelo específico de programación
  - Coexiste con otros mecanismos similares
    - Soluciones comerciales y no comerciales
  - Da un valor considerable a la información y a la seguridad

Globus (2)

- Posee una arquitectura multicapas
  - Los servicios globales se construyen sobre los locales
- Es modular
  - Es posible hacer que una aplicación aproveche solamente algunos de los servicios de Globus
  - Incrementalmente se construye una aplicación totalmente adaptada a una grid
- Interfaces de “reloj de arena”
  - Una misma interfaz mapea diversos protocolos superiores a diversos protocolos inferiores

Globus (3)

- Servicios del Globus Toolkit:
  - Localización y administración de recursos, autenticación, comunicaciones, acceso a datos
    - MDS, GRAM, Nexus, GSI, GASS
  - Basados en estándares ya existentes
    - LDAP, SSL, FTP, TCP

Seguridad

- Autenticación única para todos los recursos
- No es necesario conocer todos los nombres de cuentas y *passwords* en todos los equipos de la GRID
- Interfaz única para todos los mecanismos locales de envío de trabajos: LSF, NQE, fork, etc.
- Modelo de autenticación basado en el estándar de certificados X.509
- Uso de servidores *proxy* para delegar los servicios de autenticación con la GRID

Seguridad (2)

- Requerimientos para la autenticación
  - Poseer un certificado expedido por alguna Autoridad Certificadora
  - Cada usuario requiere un certificado
    - *Subject: O=Grid, O=Globus, OU=super.unam.mx, CN=Eduardo Murrieta*
  - Cada host destinado a brindar servicios requiere su certificado
    - *Subject: O=Grid, O=Globus, CN=host/caguama.super.unam.mx*
  - Certificados adicionales para otros servicios pueden ser requeridos
    - *Subject: O=Grid, O=Globus, CN=ldap/caguama.super.unam.mx*

Seguridad (3)

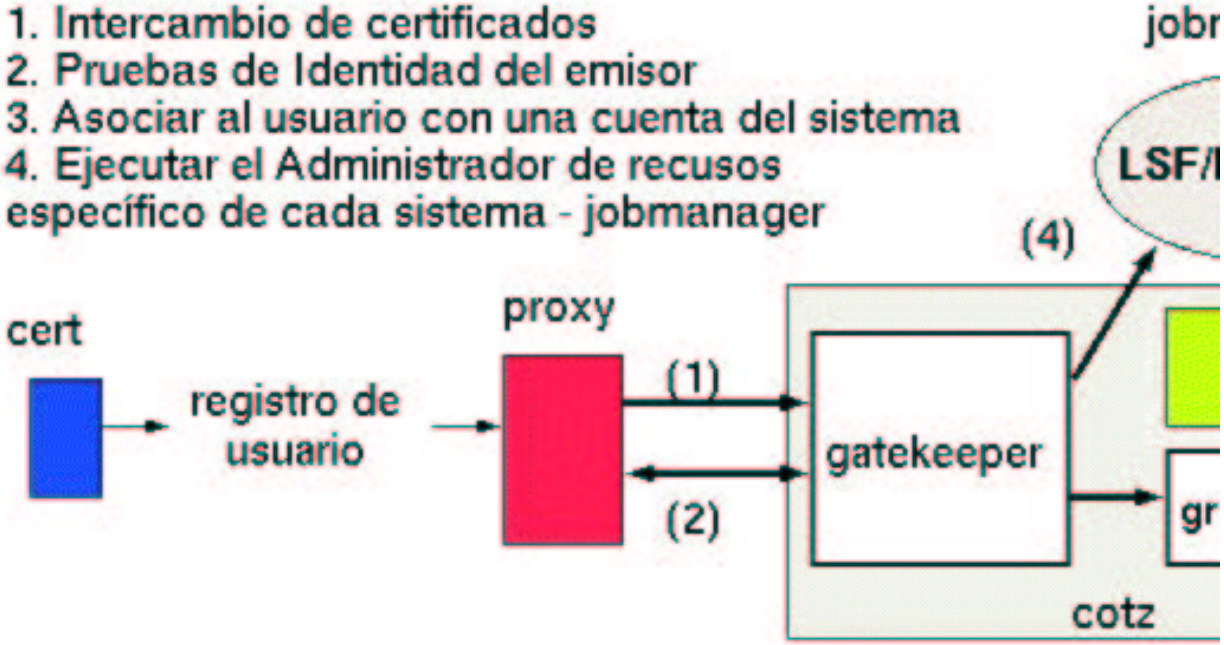
- Proceso de Registo en Globus
  - Autenticarse ante globus
    - % grid-proxy-init
    - Enter PEM pass phrase: *******
  - El proxy genera un certificado temporal firmado con la llave privada del usuario
  - Este certificado y el certificado del usuario se emplean para autenticarse con cualquier recurso de la GRID

Seguridad (4)

Ejecución de procesos:

- Para ejecutar el comando "date" en el host "cotz"
  % globus-job-run cotz /bin/date

1. Intercambio de certificados
2. Pruebas de Identidad del emisor
3. Asociar al usuario con una cuenta del sistema
4. Ejecutar el Administrador de recusos específico de cada sistema - jobmanager



Servicios de Información

- MDS provee un conjunto de herramientas y APIs para descubrir, publicar y acceder a la información referente a la estructura y estado de la GRID

- MDS emplea el protocolo LDAP para proveer una representación extendible para la información sobre los componentes de la GRID

Servicios de Información

- MDS almacena información sobre:
  - Tipo de arquitectura
  - Sistema Operativo
  - Cantidad de Memoria
  - Ancho de banda y latencia
  - Protocolos de comunicación disponibles
  - El mapeo entre direcciones IP y tecnologías de red
  - Cualquier otro tipo de información que se desee publicar

Administración de Recursos

- El Administrador de Alojamiento de Recursos de Globus (GRAM) provee el componente local para la administración de recursos.

- Una Grid construida con Globus, típicamente contiene varios GRAMs, cada uno responsable de un conjunto "local" de recursos.

Administración de Recursos (2)

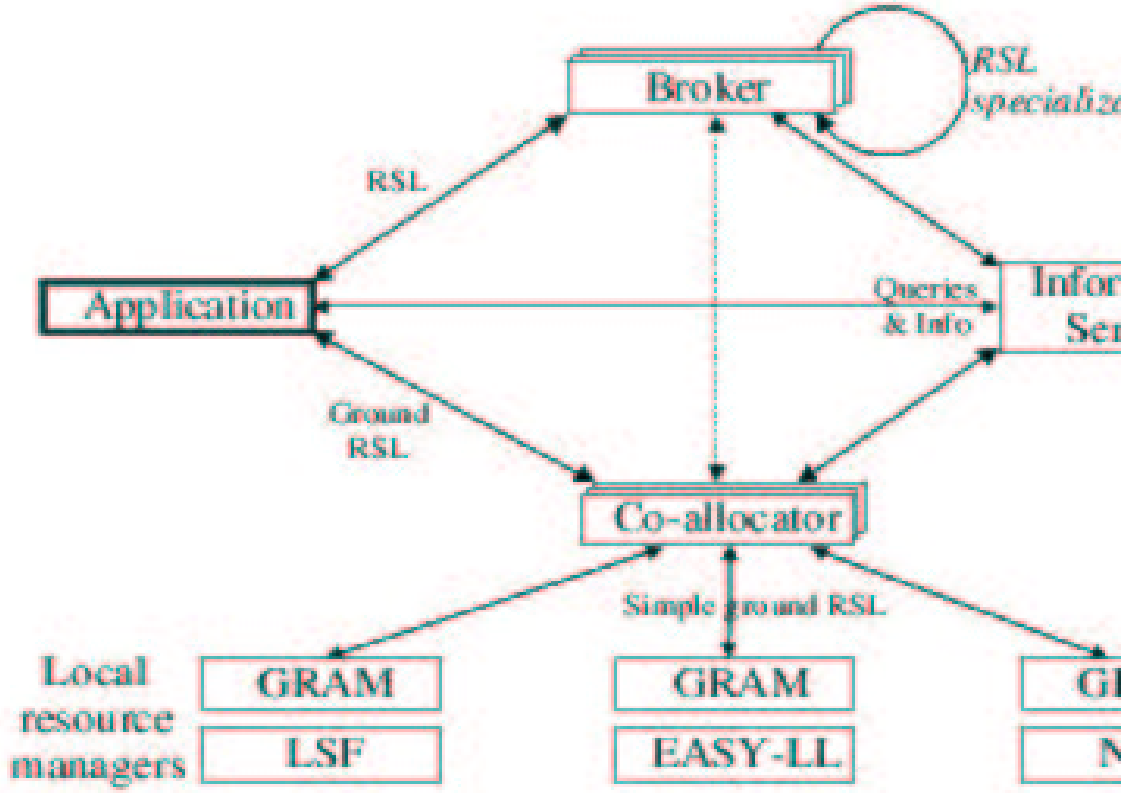
- El GRAM provee una interfaz para los sistemas de administracion de recursos locales. Esto permite que las aplicaciones y herramientas de la GRID puedan expresar sus solicitudes de recursos en términos de un API estándar.

- Con el API del GRAM, las solicitudes de recursos son expresadas en términos de un Lenguaje de Especificación de Recursos (RSL) extensible.

Administración de Recursos (3)

```
+
( &(resourceManagerContact="caguama.super.unam.mx")
  (count=1)
  (jobtype=mpi)
  (label="subjob 0")
  (environment=(GLOBUS_DUROC_SUBJOB_INDEX 0)
          (LD_LIBRARY_PATH /usr/local/globus/lib/))
  (arguments= "visual.loc" "0.1")
  (directory="/home/eml/.")
  (executable="/home/eml/./dimfract")
)
( &(resourceManagerContact="cotz.super.unam.mx")
  (count=1)
  (label="subjob 1")
  (environment=(GLOBUS_DUROC_SUBJOB_INDEX 1)
          (LD_LIBRARY_PATH /usr/local/globus/lib/))
  (arguments= "visual.loc" "0.1")
  (directory="/home/emurrieta/.")
  (executable="/home/emurrieta/./dimfract")
)
```

Administración de Recursos (4)



MPICH

- MPICH es una implementación de MPI
  - Multiplataforma
  - Arquitectura multicapas
    - Una misma API es mapeada a diferentes "dispositivos" (máquinas)
  - Contiene un dispositivo especial llamado g*lobus2*
    - Usa los métodos de comunicación provistos por el globus toolkit

MPICH (2)

- El intercambio de mensajes puede realizarse mediante diferentes protocolos
  - TCP entre procesos de diferentes máquinas
  - Protocolo nativo entre procesos de la misma máquina
    - Interfaces de “reloj de arena”
  - Las comunicaciones colectivas se realizan en 3 pasos
    - Intramáquina, LAN, WAN

Zeus - MP

- Zeus es un código de dinámica de fluidos
  - Sistemas astrofísicos
  - Configurable para diferentes *experimentos*
  - MP es la versión paralela de memoria distribuida
    - Hecha con MPI
    - Escalable

Experimentos

- Instalación y configuración de globus
  - AlphaServer, Origin 2000, PCs, Alpha DS20
- Instalación y configuración de MPICH-G2
- Ejecución de Zeus – MP a través de las diferentes arquitecturas

Experimentos (2)

- Instalación y configuración de globus
  - Instalación a partir de los códigos fuente
  - PCs: configuración estándar (gcc, 32 bits)
  - Origin 2000: configuración con compilador nativo, 32 bits y “sabor” MPI
    - Activa la comunicación multimétodo
  - AlphaServer: configuración con compilador nativo, 64 bits
    - No fue posible instalar el “sabor” MPI
  - Alpha DS20: configuración estándar (gcc, 64 bits)

Experimentos (3)

- Instalación y configuración de MPICH-G2
  - Origin 2000, PCs y Alpha DS20
    - No hay “sabor” MPI de globus en AlphaSever
  - En Origin2000, los procesos utilizan la comunicación multimétodo
  - En PCs y AlphaDS20, la comunicación se hace a través de TCP

Conclusiones

- Globus Toolkit es un producto estable
  - NO es difícil configurarlo en diferentes arquitecturas
- MPICH-G2 no lo es tanto
  - No está pensado para máquinas de memoria distribuida
    - Son las más grandes en la actualidad
- Con los servicios de globus y una herramienta adecuada, es posible convertir rápidamente una aplicación a su versión “grid”
  - Más trabajo se requiere para que la aplicación aproveche eficientemente los recursos de la grid
- Es necesario contar con herramientas de programación
  - Edición de archivos, compilación, depuración, análisis

Trabajo futuro

- Configurar MPICH-G2
  - AlphaServer + Cluster + Origin 2000 = 66 Gbytes de memoria
- Configurar Zeus para mallas de 1024^3
  - Experimentos de gran resolución
- Integrar equipos de otras instituciones
  - Cuantificar los beneficios de I2
- Integrar más aplicaciones en versiones “grid”
  - Cactus
- Explorar herramientas de programación
- Explorar balanceadores de carga

Referencias

- Globus: www.globus.org
- MPICH – G2: www3.niu.edu/mpi
- Zeus: zeus.ncsa.uiuc.edu:8080/lca_intro_zeusmp.html
- Teragrid: www.teragrid.org
- Supercómputo UNAM: www.super.unam.mx
- Semana de Supercómputo 2003:

www.super.unam.mx/semana2003