

Computação de Alto Desempenho
(*High Performance Computing*)



Paulo Trezentos
(Paulo.Trezentos@iscte.pt)
ISCTE / ADETTI
24/6/2003

HPC ?



Introdução

Mainframes

Clusters

Grids

O que é HPC ?

- A tecnologia empregue na resolução de problemas que:
 - requerem poder computacional significativo
 - necessitam de aceder ou processar grandes quantidades de informação
 - necessitam de operar interactivamente através de uma rede geograficamente dispersa

Definição II

HPC compreende uma colecção de sistemas de hardware, ferramentas de software e linguagens de programação que tornam possíveis a resolução de problemas anteriormente impraticáveis a um preço apropriado

EPPC - University of Edinburgh

Necessidade

- Evolução HPC
 - Mainframes
 - Cluster beowulf
 - Grid computing



Terminologia

- Comunicação
 - Largura de Banda
 - Latência
- Segurança / disponibilidade
 - Redundância (RAID ou Equipamento)
 - Tolerância a Falhas (*fault tolerance*)
 - Backups
 - Load balancing

Introdução

- Mainframes
- Clusters
- Grids

Mainframes

- Mainframe é um termo utilizado para grandes computadores, tipicamente produzidos por companhias como IBM ou Cray para aplicações computacionalmente muito exigentes.
- Centralizado
- Obsoletos rapidamente
- Dependência de um fabricante
- Bom I/O e fiabilidade

Mainframes

- Sun UltraEnterprise 10000



- IBM Z-series

Z-series (Z800 / Z900)

- Suceder ao S/390
- Solução chave-na-mão, suporte caro
- Empresas financeiras, EDP, ...
- Custo
 - Licenciamento software: zOS



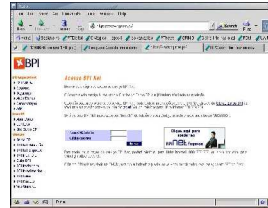
Características interessantes

- Virtualização
 - várias imagens dentro da mesma máquina
 - LPAR
- Clusterização
 - vários S/390 num cluster (Parallel Sysplex Cluster technology)
 - load balancing, tolerância a falhas e partilha de recursos (robots de tapes e discos)
 - Através de equipamento adicional: CF (*Coupling Facilities*)
 - 1 GB/s (por fibra)

Características interessantes

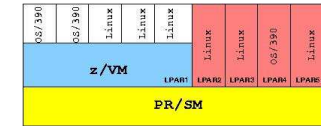
• Segurança

- implementação por hardware com algoritmos RSA, DES e triplo Des
- 2 co-processadores criptográficos CMOS (de fábrica)
- possibilidade de adicionar 2 placas adicionais
- 4300 transacções por segundo



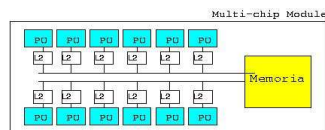
Arquitetura I

- Processor Resource / System Manager (partilha de canais físicos às LPARs)
- Limitados a 15 LPARs
- Se for instalado o Z/VM podemos ter até 15.000 imagens
- Virtualização: aproveitamento do tempo *idle*
- Redução do TCO



Arquitetura II

- Arquitectura 64 bits
- MCM (MultiChip Module)
- PU – Processor Unit: x 12 ou x 20
 - CP
 - SAP
 - ICF
 - IFL
- Ciclo de relógio de 1.3 ns



Introdução Mainframes Clusters Grids

Beowulf clusters

- Utilização de PCs vulgares
- Interligados através de FastEthernet
- Software:
 - Linux
 - Ferramentas GNU (gcc, gdb, ...)
 - Comunicação: PVM e LAM (MPI ou MPICH)
 - Job scheduler



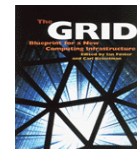
NoW- Network of Workstations

- Alpha ou Sparc
- Univ. Coimbra
 - Centopeia
- Rendering: Titanic

Introdução
Mainframes
Clusters
Grids

Grids I

- Conceito apresentado por
 - Ian Foster, Carl Kesselman
 - *The Grid: Blueprint for a New Computing Infrastructure* (1998)
- Infra-estrutura de hardware e software
- Permite- nos acesso a grandes capacidades computacionais
- ...de forma confiável, consistente, económica e persistente



Grids II

- *Conceito antigo com uma dinâmica nova*
- Ideia:
 - Usufruir e capacidade de computação (Storage / CPU)
 - Sem ter de se preocupar de onde vem, como é mantida, etc...
 - ...como na Rede Eléctrica Nacional
- Então e ...
 - Napster / Gnutella
 - SETI
 - Web

Grid checklist

- ✓ recursos coordenados que não estão submetidos a controlo centralizado
- ✓ usar protocolos e interfaces standards, abertos e genéricos
- ✓ fornecem qualidade de serviço não- trivial

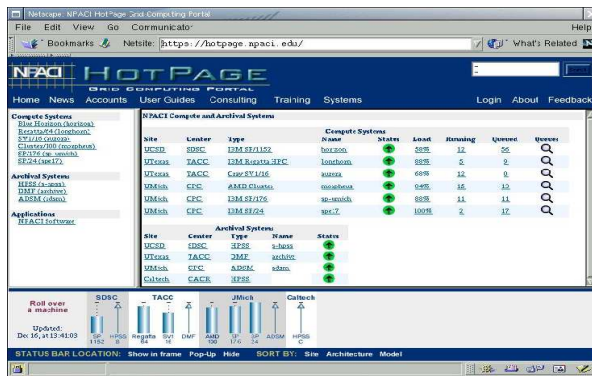
Actores activos (Standards)

- **Global Grid Forum (GGF)**
 - Interfaces e protocolos Grid
 - Grupos de trabalho:
 - Applications (APME)
 - Architecture (ARCH)
 - Data
 - Information Systems and Performance
 - Peer-to-Peer
 - Scheduling and Resource Management
 - Security
 - 7th GGF meeting: Março, Tóquio

Actores activos (Implementação)

- Globus (www.globus.org)
 - Argonne National Laboratory
 - University of Chicago
 - University of Southern California
- Globus Toolkit versão 2.0
- Próxima geração (versão 3.0):
 - Open Grid Services Architecture (OGSA)
 - Grid + WebServices
 - IBM
 - OGSA security architecture (draft)

NPACI (<https://hotpage.npaci.edu/>)



ISCTE / MGSJ

Computação de Alto Desempenho

25

TeraGrid (<http://www.teragrid.org/>)



ISCTE / MGSJ

Computação de Alto Desempenho

26

Centopeia

- Dep. Física da UC
 - 24 DEC Alpha 500au (processor 21164 at 500 MHz) with total RAM of (8x192,8x384,8x768) MB.
 - 13 COMPAQ XP 1000 (processor 21264A at 667 MHz) with total RAM (12x512 MB,1x1 GB)
 - 12 Intel P IV a 2.2 GHz com 1GB de memoria RDRAM
- PBS (job dispatcher)

ISCTE / MGSJ

Computação de Alto Desempenho

27

GridPT

- <http://adetti.gridpt.org>
 - ADETTI
 - FCUL
 - IM
 - I.Astronomia
 - FEUP
 - IPCB
- Protótipo: <http://aragorn.adetti.iscte.pt>



ISCTE / MGSJ

Computação de Alto Desempenho

28

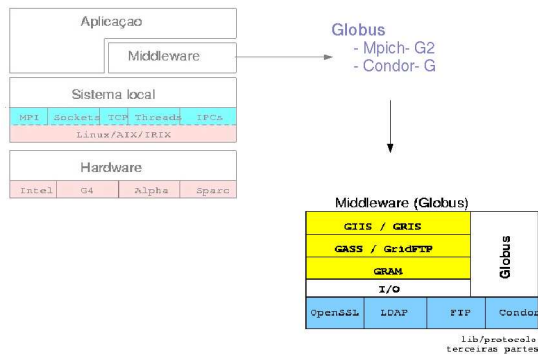
Tipos de Utilização

- Submeter um programa paralelizado num dos nós, a partir do qual é feito o 'spawn' para os restantes
- Monitorizar a tarefa
- Copiar os binários entre maquinas
- Sistemas de transferência de ficheiros (FTP) com QoS elevadas
- Sistemas de informação centralizados
- Segurança uniforme (single sign-on,...)

Criação de uma Grid (*shopping list*)

- Hardware
 - Cluster de PCs / Alphas
 - Rede FastEthernet / Gigabit / Myrinet
- Software
 - Linux
 - PVM ou LAM (MPI/MPICH G2)
 - Globus
 - Middleware
 - Condor-G
 - MPICH G2)

Arquitetura Globus



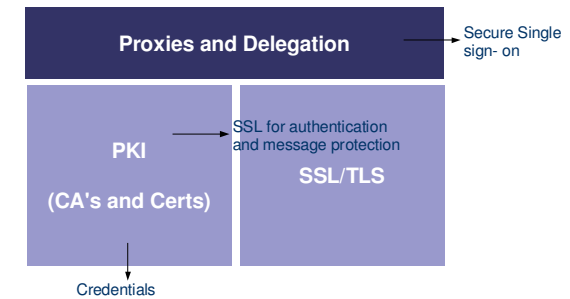
Globus II

- Middleware que suporta as Grids
- Componentes:
 - Resource management → GRAM
 - Data management → GASS, GridFTP
 - Information Services → GIIS, GRIS
 - Security → GSI, GSS-API

Requisitos de segurança

- single sign- on
- interface transparente com ferramentas existentes: FTP, SSH e browsers
- API's standardizadas para que programadores externos possam aceder ao nível de segurança
- interoperabilidade com soluções segurança existentes
- administradores dos sites devem ter controlo sobre os seus recursos

Arquitectura Segurança



Aplicação típica

- dados distribuídos (HEP) por várias máquinas
- Lançamos um *master* que
 - identifica em que máquinas é que está a informação que precisa (MetaData catalog)
 - lança slaves (apenas) nas máquinas que tiverem CPU superior a 1Ghz e *load* menor que 0.3
- clientes depois de filtrarem devolvem ao master
- Master mostra ao cliente *via* browser

Exemplos

Lançar um trabalho

```
rc = globus_gram_client_job_request(rm_contact,
    specification,
    job_state_mask,
    callback_contact,
    &job_contact);
```

Copiar Ficheiro

```
source_url_fd = globus_gass_open(source_url, O_RDONLY);
while ((n = read(source_url_fd, buf, BUF_SIZE)) > 0)
    write(destination_url_fd, buf, n);
globus_gass_close(source_url_fd);
```

Conclusões

- Conceito simples e antigo
- Resultado do aumento da largura de banda
- Largamente apoiado pelo mundo empresarial
- Atrasados em Portugal
 - Iniciativa Nacional para a Computação em Rede
- Aplicações facilmente adaptáveis

Obrigado.

Perguntas ?

<http://paulo.trezentos.gul.pt/artigos>