Projeto de um Cluster Didático para Programação Paralela e Distribuída (Parte I)

Paulo Shirley
Univ. Aberta, pos@uab.pt

Resumo

O aumento do poder de cálculo permite resolver problemas mais rapidamente ou problemas de maiores dimensões antes inacessíveis. Este aumento tem um grande impacto a todos os níveis, sejam eles de desenvolvimento, investigação ou de prestação de serviços. A Computação de Alto Desempenho (HPC - High Performance Computing) implementada através de um conjunto de computadores pessoais dedicados ligados por uma rede privada, vulgarmente denominado Beowulf cluster, surge como uma solução viável e relativamente económica para o acesso a um maior poder computacional. Este trabalho descreve a conceção, planeamento e implementação de um cluster experimental que replica o ambiente e operacionalidade de clusters de maiores dimensões, permitindo o contato com este tipo de tecnologia e a sua utilização para fins didáticos em programação paralela e distribuída, entre outros.

Palavras-chave: cluster, HPC, programação paralela, MPI, cálculo científico

Title: Design of a Didactic Cluster for Parallel and Distributed Programming (Part I)

Abstract

Increasing computing power allows one to solve problems faster or solve larger previously inaccessible problems. This increase has a major impact at all levels, be it development, research or service delivery. High Performance Computing (HPC) implemented through a set of dedicated personal computers connected by a private network, called a Beowulf cluster, emerges as a viable and relatively inexpensive solution for access to increased computing power. This work describes the design, planning and implementation of an experimental cluster that replicates the environment and operation mode of larger clusters, allowing the contact with this type of technology and its use for didactic purposes in parallel and distributed programming, among others.

Keywords: cluster, HPC, parallel programming, MPI, scientific computing

1. Introdução

O poder computacional do computador pessoal (PC) tem vindo sempre a aumentar desde o seu aparecimento e assim continuará previsivelmente no futuro próximo. No entanto, um PC pode não ser suficiente para certos tipos e complexidade sempre crescente de problemas que se pretendem resolver, criando ou elevando barreiras aos resultados que se conseguem alcançar. Uma solução é utilizar não um, mas muitos PC trabalhando em paralelo e em modo coordenado para resolver o problema em questão. É nesta situação que a tecnologia dos clusters (aglomerados) de computadores podem intervir, expandindo e multiplicando as capacidades de computação do computador individual.

Clusters de computadores podem ser construídos com objetivos específicos, possivelmente sobrepostos, tais como:

- Alta Disponibilidade (High Availability). O objetivo é a tolerância a falhas.
- Alto Débito (High Troughput). O objetivo é calcular a solução de um conjunto de N problemas no menor tempo possível, não sendo importante o tempo gasto na solução de um problema particular.
- Alto Desempenho (High Performance). O objetivo é calcular a solução de um único problema no menor tempo possível.

Neste trabalho o foco é na Computação de Alto desempenho ou HPC (High Performance Computing), onde o objetivo principal é obter a máxima velocidade de resolução, ou seja, que N computadores resolvam um problema tão perto quanto possível do ótimo teórico de 1/N do tempo que leva um único computador a resolver o mesmo problema. Por outras palavras, isto significa resolver um problema N vezes mais rápido ou resolver um problema N vezes maior no mesmo espaço de tempo.

Este aumento de capacidade computacional tem um grande impacto a todos os níveis, sejam eles de desenvolvimento, investigação ou de prestação de serviços. O recurso a clusters como plataformas de computação paralela já é considerada uma solução padrão por cientistas e engenheiros [Quinn 2004] para a resolução de problemas tão diversificados como por exemplo a evolução de galáxias, modelação e previsão do clima, desenho de circuitos integrados, projeto de aviões e dinâmica de moléculas. Os clusters têm também um papel importante no método científico moderno, onde a simulação numérica substitui a experimentação física, dado esta poder ser demasiado cara, demorada, não-ética ou impossível de realizar.

Dada a atual importância e potenciais aplicações da tecnologia dos clusters, este trabalho descreve um projeto que consistiu na conceção, planeamento e implementação de um cluster experimental para computação de alto desempenho que replica o ambiente e operacionalidade de clusters de maiores dimensões, contribuindo para um acesso mais fácil a este tipo de tecnologia e proporcionando um instrumento que pode ser utilizado para fins didáticos em programação paralela e distribuída, entre outros.

Montar um cluster experimental pode ser economicamente acessível se estiverem disponíveis alguns computadores pessoais para o efeito. Todo o software utilizado é baseado no sistema operativo Linux e é gratuito. Em conjunto com um comutador de rede

(switch), cabos, algum tempo livre e conhecimentos de informática a nível de redes, é relativamente fácil construir um cluster seguindo as ideias e indicações dadas neste artigo.

Este artigo está estruturado em duas partes, sendo esta a primeira parte e constituída por 5 secções. Na secção 2 descreve-se a arquitetura geral do cluster, referindo-se o ponto de vista do utilizador e o ponto de vista da administração. Na secção 3 descrevem-se os componentes de Hardware. Na secção 4 descreve-se um primeiro conjunto de componentes de Software correspondentes aos serviços básicos de operação do cluster e na secção 5 são apresentadas as conclusões. Os restantes componentes de Software, tais como os ficheiros de configuração automática dos nós, o gestor de recursos e a biblioteca padrão MPI para programação paralela distribuída, assim como exemplos de utilização do cluster, serão abordados na segunda parte deste artigo.

2. Arquitetura e funcionamento geral

O tipo de cluster considerado neste trabalho é constituído por elementos considerados comuns ou de baixo custo, tais como computadores pessoais (PC), comutadores ethernet, cabos ethernet e software gratuito baseado no sistema operativo Linux. Este tipo de cluster é usualmente denominado de Beowulf Cluster [Sloan 2005] [wikipedia, Beowulf cluster], dado ter sido o nome dado ao primeiro cluster montado deste género.

O sistema operativo escolhido foi a distribuição Linux CentOS (Community ENTreprise Operating System), que é uma distribuição derivada a partir de código fonte livremente cedido ao público do Red Hat Enterprise Linux (RHEL), conhecida por ser robusta e estável. Foi escolhida a versão CentOS 6 por esta estar disponível quer para computadores de 64 bits quer para computadores de 32 bits, um fator importante se apenas estiverem disponíveis computadores de 32 bits ou a quantidade de memória RAM disponível for reduzida. A arquitetura geral do cluster é simplesmente um conjunto de computadores pessoais ligados por uma rede ethernet privada mas pode ter variantes conforme a função desempenhada por cada computador que o compõe. A figura 1 mostra um diagrama da arquitetura utilizada neste trabalho.

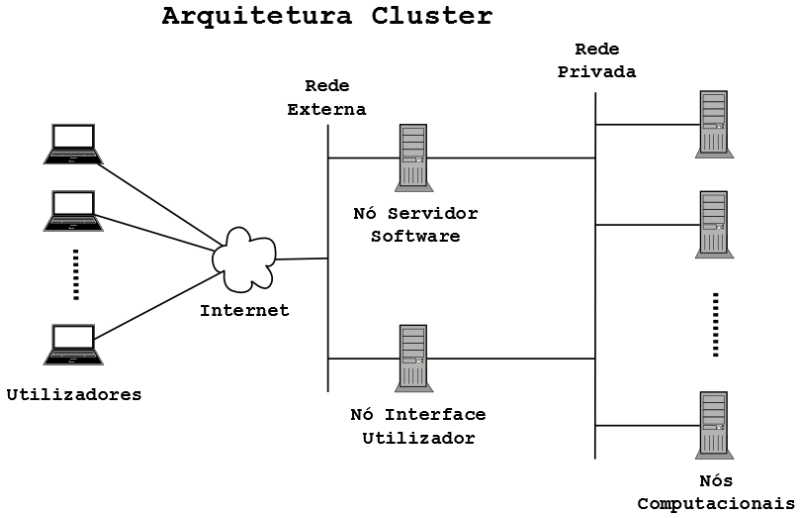

Figura 1. Arquitetura proposta para o cluster.

Nesta arquitetura identificam-se três tipos de nós:

1) Os nós Computacionais. Estes são os computadores dedicados exclusivamente a efetuar os cálculos necessários à resolução dos problemas.

2) O nó Interface com o Utilizador. Este é o computador que efetua a ligação dos utilizadores com o cluster, ou mais propriamente, com os nós computacionais. Nele os utilizadores fazem login, editam ficheiros, importam e exportam dados relacionados com os problemas que pretendem resolver, compilam programas e executam-nos. A execução de programas não é efetuada diretamente pelos utilizadores mas sim por intermédio de um gestor de recursos [Slurm] que gere os nós computacionais. Um utilizador que pretenda executar um programa submete um trabalho (job) ao gestor de recursos que o coloca numa fila de espera e efetua o seu processamento em lote (batch) de acordo com critérios de prioridade definidos. Um trabalho consiste num conjunto de definições tais como programa a executar, dados de entrada, número de processadores a utilizar, etc, coligidas num ficheiro.

3) O Servidor de Software. Este computador contém os ficheiros de instalação de todo o software necessário ao cluster (sistema operativo, serviços, gestão do cluster, software do utilizador tais como compiladores, bibliotecas, utilitários). Deve de um modo prático poder instalar um nó computacional rapidamente assim como o nó de Interface com o Utilizador. É por defeito o computador através do qual é feita a administração e manutenção de todos os computadores do cluster e o único que dispõe de consola para acesso local, ou seja, monitor, teclado e rato. A sua interferência na atividade computacional do cluster pretende-se que seja mínima, razão pela qual também tem uma segunda placa de rede, permitindo um acesso externo direto ao cluster por parte do administrador, muito útil quando não é necessária a presença no local.

3. Componentes de Hardware

Para os componentes de hardware do cluster deverão ser reunidos os seguintes elementos:

- Um computador para o Servidor de Software, com a maior capacidade de disco possível e dois portos de rede;
- Um computador para o nó de Interface com o Utilizador, com dois portos de rede, um dos quais com suporte de arranque pela rede por PXE (Preboot Execution Environment) [intel 1999];
- Um conjunto de N computadores com porto de rede com suporte de PXE, onde N é o número de nós computacionais;
- Um monitor, teclado e rato para a consola do Servidor de Software;
- Um ou mais comutadores ethernet que permitam ligar em rede os N+2 computadores;
- N+4 cabos ethernet;
- Espaço físico para a colocação dos computadores, com instalação elétrica e ventilação/refrigeração adequadas ao número de computadores utilizados;

- Acesso à Internet a partir da rede externa;
- Um CD ou uma Pen USB com uma versão live do sistema operativo Linux CentOS 6;
- Ferramentas várias.

Antes da montagem dos computadores em rede, cada computador deve:

1) Ser atribuído e identificado com um número de série 1,2,3,... para sua fácil identificação.

2) Ser previamente aberto, inspecionado e observado, limpo o seu interior com especial atenção ao dissipador do processador e instalada a segunda placa de rede se caso disso. O computador que irá ser o nó Interface com o Utilizador deve ter a maior capacidade de disco, se possível até mais do que um disco, dado que irá guardar os dados dos utilizadores.

3) Ser ligado individualmente, analisado e configurado o BIOS, tendo em conta os seguintes aspetos:

- Acertar o relógio de hardware;
- Desligar equipamento desnecessário como portas série e paralelas;
- Desligar o erro por falta de teclado;
- Registar o endereço MAC, dimensão do disco rígido, memória RAM e nº de núcleos físicos (cores) do processador.
- Provisoriamente, configurar o arranque pelo leitor de CD ou USB;

4) Ser efetuado o arranque do computador com um CD ou uma Pen USB com uma versão live do CentOS 6, seguido dos seguintes passos:

- Verificar o sucesso do arranque do sistema operativo. Se o arranque for bem sucedido, o computador deverá estar apto a integrar o cluster;
- Executar o utilitário Applications/System Tools/Disk Utility. Este utilitário permitirá confirmar a dimensão do disco rígido e testá-lo se suportar a tecnologia SMART;
- Executar o utilitário Applications/System Tools/Terminal para obter uma linha de comandos e executar o comando "ifconfig -a". Este comando irá listar os dados associados aos portos de rede. Registar os nomes lógicos (eth0, eth1, ...) e os respetivos endereços MAC. Esta informação é muito importante no caso dos nós Interface com o Utilizador e Servidor de Software para se saber como o sistema operativo identifica cada placa de rede. Admite-se que eth0 corresponderá ao porto de rede da placa mãe (motherboard) e que eth1 à placa de rede extra instalada.

5) Exceto para o computador que vai ser o Servidor de Software, voltar ao BIOS e configurar o arranque pela rede (PXE boot) em primeiro lugar e pelo disco rígido em segundo lugar.

No final destes passos deverá obter-se uma tabela onde para cada computador conste o número de série atribuído, número de núcleos do processador, capacidade disco, RAM,

MAC e correspondentes nomes ethx dos portos de rede, após o que se poderá proceder à montagem física em rede dos computadores como ilustrado na figura 1.

4. Componentes de Software

Para a instalação dos nós computacionais e de interface com o utilizador adotou-se uma estratégia automatizada baseada no arranque do nó pela rede e definição da configuração da instalação por um ficheiro. O arranque pela rede é feito utilizando o ambiente PXE (Preboot Execution Environment) [intel 1999] sendo o servidor de software configurado como um servidor de arranque PXE (PXE boot server) [redhat 2014,2013]. Para a definição da configuração da instalação o CentOS permite criar um único ficheiro, denominado ficheiro Kickstart [redhat 2014], que contém as respostas a todas as perguntas que normalmente seriam feitas durante uma instalação manual, automatizando assim o processo.

Sendo a instalação dos nós computacionais e de interface com o utilizador automatizadas, o processo de instalação do cluster consiste essencialmente na instalação e configuração do servidor de software. Uma vez feita, o passo seguinte é ligar os nós e esperar que as instalações ocorram. O nó interface com o utilizador, dada a sua natureza particular, precisará ainda de uma intervenção final manual.

A instalação do servidor de software é dividida em 10 passos como indicado na tabela 1 e cada passo é explicado nas seções seguintes. Admite-se como exemplo que é utilizada a versão 6.9 de 32 bits do CentOS e que todos os comandos indicados são executados pelo utilizador root. Atribui-se a título exemplificativo, o nome de “juno” ao cluster. Esse será também o nome (hostname) do nó interface com o utilizador, dado que é o computador onde os utilizadores fazem login. Ao servidor de software atribui-se o nome “juno2”.

Tabela 1. Passos para a instalação do Servidor de Software.

1- Instalação do sistema operativo
2- Estrutura de diretorias para o servidor FTP
3- Configuração do servidor FTP
4- Configuração do cliente NFS
5- Configuração do servidor PXE (DHCP + TFTP)
6- Configuração do servidor NTP
7- Compilação e configuração do gestor de recursos Slurm
8- Compilação e configuração da biblioteca MPI
9- Configuração do Firewall
10- Ficheiros de configuração dos nós

A rede privada do cluster é definida pelos seguintes parâmetros (que podem ser usados na prática) que suportam um cluster com até cerca de 250 nós.
Domínio: cluster.local
Rede: 172.17.0.0/23 ou 172.17.0.0/255.255.254.0
Servidor de Software: IP fixo 172.17.0.254

Nó Interface com o Utilizador: IP fixo 172.17.0.253
Nó Computacional com número de série i: O IP é atribuído em função do MAC pelo servidor DHCP. Segue-se uma estratégia em que o nó é configurado para o IP 172.17.0.i se se pretender que o nó arranque do sistema operativo do seu próprio disco ou para o IP 172.17.1.i se se pretender que ocorra uma instalação do nó por PXE. É assim possível pela simples configuração do serviço DHCP definir se um nó arranca normalmente ou se é para ser instalado. Tendo em conta as definições dadas para a rede privada, a figura 2 mostra o ficheiro hosts a utilizar em todos os computadores do cluster.

```
# /etc/hosts p/ rede do cluster
127.0.0.1   localhost localhost.localdomain localhost4
localhost4.localdomain4
::1         localhost localhost.localdomain localhost6
localhost6.localdomain6

# Nó Interface com Utilizador
172.17.0.253           juno.cluster.local        juno
# Servidor de Software
172.17.0.254           juno2.cluster.local       juno2
# nós computacionais
172.17.0.1             node1.cluster.local       node1
172.17.0.2             node2.cluster.local       node2
...
172.17.0.10            node10.cluster.local      node10
```

Figura 2. Ficheiro hosts.

A rede externa do cluster é definida pelos seguintes parâmetros (a título exemplificativo, substituir por valores apropriados).
Rede: 192.168.5.0/24 ou 192.168.5.0/255.255.255.0
Servidor de Software: IP fixo 192.168.5.24
Nó Interface com o Utilizador: IP fixo 192.168.5.23
Gateway: 192.168.5.254
DNS: 192.168.5.10

4.1. Instalação do Sistema Operativo

A instalação do sistema operativo no software server é feita manualmente, sendo a distribuição Linux CentOS 6 composta por dois DVD que serão ambos necessários. Sugere-se o seguinte procedimento:
(i) Num PC com gravador de DVD, descarregar (www.centos.org) as imagens (.iso) dos dois DVD da última versão do CentOS 6 diretamente para um disco USB externo e verificar os checksums.
(ii) Gravar fisicamente o DVD1.
(iii) A partir do DVD1 efetuar uma instalação tipo Software Development Workstation + Customização. Escolher opção sincronizar tempo pela rede (ativa serviço NTP). Configurar portos de rede conforme seção anterior e considerando o porto eth0 ligado à rede externa e o porto eth1 ligado à rede privada. Acrescentar os itens Base/{Console Internet tools,

Network tools}, Language/Português, Servers/{FTP, NFS, Net Infrastructure, System admin tools}. Escolher Virtualization/OFF.

Depois dos passos anteriores, reiniciar o sistema, seguir as instruções de primeiro arranque do sistema operativo e verificar o bom acesso à Internet pelo porto eth0. Criar o ficheiro hosts (figura 2) e executar como utilizador root os comandos da figura 3,

```
yum -y update
yum -y install epel-release
yum -y install yum-utils createrepo ntfs-3g
cp hosts /etc/hosts
```

Figura 3. Comandos a executar após a instalação do Sistema Operativo.

onde yum é o comando do CentOS para manipular pacotes (packages) de software em repositórios. O primeiro comando efetua uma atualização do sistema operativo, o segundo acrescenta um repositório de software adicional (EPEL - Extra Packages for Enterprise Linux) e o terceiro instala utilitários para manipular repositórios e para aceder a sistemas de ficheiros NTFS.

4.2. Estrutura de Diretorias para o Servidor FTP

Os nós obtêm todos os ficheiros que necessitam do servidor de software por FTP. Os ficheiros dividem-se em duas categorias: ficheiros de configuração e repositórios locais de software. Um repositório em CentOS é uma diretoria que contém os pacotes de software em formato .rpm, normalmente numa subdiretoria Packages, e meta-informação sobre o software disponível numa subdiretoria Repodata. Localmente é necessário replicar os repositórios oficiais “base”, “extras” e “updates”, sendo ainda criado um repositório adicional “others” para guardar software específico que não se encontra nos outros (por exemplo pacotes obtidos por compilação e pacotes fornecidos por terceiros). A figura 4 mostra a estrutura principal de diretorias para o serviço FTP. As subdiretorias Packages e Repodata dos repositórios serão criadas aquando da cópia do conteúdo dos repositórios.

```
/var/ftp/pub/centos6-32/cfg
/var/ftp/pub/centos6-32/repos/base
/var/ftp/pub/centos6-32/repos/extras
/var/ftp/pub/centos6-32/repos/others
/var/ftp/pub/centos6-32/repos/updates
```

Figura 4. Estrutura de diretorias para serviço FTP.

O conteúdo do repositório base é obtido por cópia dos dois DVD da distribuição, mais concretamente, a partir das imagens .iso gravadas anteriormente no disco USB externo que deverá agora ser ligado ao servidor de software. Os repositórios extra e updates são obtidos por sincronização com os repositórios na Internet. A figura 5 mostra os comandos necessários para a criação dos repositórios locais base, extras e updates.

```
DIRFTP=/var/ftp/pub/centos6-32
```

```
# paths dvds .iso no disco USB externo
ISO1=/media/hd-usb/isos/CentOS-6.9-i386-bin-DVD1.iso
ISO2=/media/hd-usb/isos/CentOS-6.9-i386-bin-DVD2.iso
# repo base
mkdir $DIRFTP/DVD
mount -o loop -t iso9660 $ISO1  $DIRFTP/DVD
echo 'copy DVD1...'
rsync -avuq $DIRFTP/DVD/ $DIRFTP/repos/base
umount $DIRFTP/DVD
mount -o loop -t iso9660 $ISO2  $DIRFTP/DVD
echo 'copy DVD2...'
rsync -avuq $DIRFTP/DVD/Packages/ $DIRFTP/repos/base/Packages
umount $DIRFTP/DVD
rmdir $DIRFTP/DVD
# repo extras
echo 'sync repo extras...'
reposync --repoid=extras -n -m -q -p $DIRFTP/repos
createrepo -q $DIRFTP/repos/extras
# repo updates
echo 'sync repo updates...'
reposync --repoid=updates -n -m -q -p $DIRFTP/repos
createrepo -q $DIRFTP/repos/updates
```

Figura 5. Comandos para criação repositórios locais: “base” a partir dos DVD; “extras” e “updates” a partir da Internet.

4.3. Configuração do Servidor FTP

O pacote de software correspondente ao servidor FTP tem o nome "vsftpd" (very secure ftp). O seu ficheiro de configuração encontra-se em "/etc/vsftpd/vsftpd.conf" e a configuração por defeito é apropriada, sendo apenas necessário indicar a diretoria raiz para utilizadores anónimos. A figura 6 mostra os comandos e ações a efetuar para configurar o serviço FTP.

```
# instalar servidor e cliente ftp
yum -y install vsftpd ftp
# editar vsftpd.conf e acrescentar a linha
# "anon_root=/var/ftp/pub"
vim /etc/vsftpd/vsftpd.conf
# configurar serviço para iniciar no arranque do PC
chkconfig vsftpd on
# iniciar serviço agora
service vsftpd start
```

Figura 6. Comandos e ações para configurar o serviço FTP.

Após a instalação o servidor pode ser testado com o comando "ftp localhost" e verificado se a estrutura de diretorias criadas estão acessíveis.

4.4. Configuração do Cliente NFS

Os dados dos utilizadores do cluster são guardados num disco do nó interface com o utilizador na diretoria /data e são exportados por NFS para todo o cluster. Por questões de administração é vantajoso que o servidor de software também tenha acesso a essa diretoria, para o que basta acrescentar uma linha ao ficheiro /etc/fstab como indicado na figura 7.

```
# editar /etc/fstab e acrescentar a linha
# "172.17.0.253:/data   /data    nfs   rw,hard,intr   0 0"
vim /etc/fstab
```

Figura 7. Comandos para configurar o cliente NFS.

4.5. Configuração do servidor PXE (DHCP + TFTP)

O servidor PXE é obtido pela configuração de dois serviços, DHCP e TFTP. O servidor DHCP é responsável por reconhecer um pedido de um cliente PXE, atribuir-lhe um IP e indicar o servidor TFTP do qual obtém os ficheiros de arranque.

4.5.1. Configuração do Servidor TFTP

O pacote de software correspondente ao servidor TFTP tem o nome "tftp-server" e um cliente simples que pode ser usado para testes tem o nome "tftp". O serviço tftp é de fato controlado por outro serviço, xinetd, pelo que ambos devem ser instalados e configurados. A configuração do servidor tftp é feita no ficheiro /etc/xinet.d/tftp no qual para ativar o serviço se deve definir "disable=no". Por defeito o ficheiro define a diretoria raiz do serviço como /var/lib/tftpboot. Para o serviço tftp é necessário construir uma estrutura de duas subdiretorias na diretoria raiz, ./pxelinux e ./pxelinux/pxelinux.cfg. A figura 8 mostra os comandos e ações iniciais a efetuar para configurar o serviço TFTP.

```
# instalar servidor e cliente tftp
yum -y install xinetd
yum -y install tftp-server tftp
# editar /etc/xinet.d/tftp e definir "disable=no"
vim /etc/xinetd.d/tftp
# criar estrutura diretorias
mkdir -p /var/lib/tftpboot/pxelinux/pxelinux.cfg
# configurar serviços para iniciar no arranque do PC
chkconfig xinetd on
chkconfig tftp on
```

Figura 8. Comandos iniciais para configurar o servidor TFTP.

Para o serviço tftp é necessário construir uma estrutura de duas diretorias,

```
/var/lib/tftpboot/pxelinux
/var/lib/tftpboot/pxelinux/pxelinux.cfg
```

A diretoria /var/lib/tftpboot/pxelinux contém os seguintes ficheiros com a função e proveniência descritas,

- pxelinux.0: Programa de "network bootstap". É o programa que inicia todo o processo de descarregar por tftp todos os outros ficheiros aqui descritos. É obtido do projeto [pxelinux].
- vesamenu.c32: Programa capaz de gerar um menu com opções de execução definidas por um ficheiro que se encontra na subdiretoria ./pxelinux.cfg. É obtido do projeto [pxelinux].
- splash.jpg: Imagem de fundo do CentOS para o menu. É obtido da cópia do DVD na diretoria /var/ftp/pub/centos6-32/repos/base/isolinux.
- vmlinuz e initrd.img: Kernel Linux e imagem inicial. Obtidos da cópia do DVD na diretoria /var/ftp/pub/centos6-32/repos/base/images/pxeboot.

A diretoria /var/lib/tftpboot/pxelinux/pxelinux.cfg contém ficheiros que definem menus com opções de execução. De especial interesse são a opção de arranque do kernel linux da diretoria anterior com determinados argumentos e a opção de arranque do disco local. O ficheiro de menus utilizado depende do IP do cliente e assim é possível definir em função do IP diferentes cursos de acção. O nome do ficheiro do menu pode ser um IP completo ou uma sua truncatura, permitindo definir um conjunto de IP constituído por todas as possibilidades da parte omitida. Como referido anteriormente, os nós do cluster que arrancam do disco local têm IP do tipo 172.17.0.i e podem ser representados pelo ficheiro de menu de nome AC1100 que corresponde à parte fixa do IP em hexadecimal. Igualmente os nós do cluster que vão ser instalados têm IP do tipo 172.17.1.i e podem ser representados pelo ficheiro de nome AC1101. Finalmente o nó Interface com o Utilizador tem IP 172.17.0.253 ou 172.17.1.253 e é representado pelo ficheiro de nome AC1100 ou AC1101FD.

A figura 9a mostra o ficheiro de menu AC1101 para instalação de um nó computacional e a figura 9b o ficheiro de menu AC1100 para arranque do disco local. A diferença entre eles é a opção por defeito (linha menu default) que é ativada após o "timeout" de 5 segundos. Para a instalação do nó é efetuado o arranque com o kernel Linux tendo como argumento a localização de um ficheiro kickstart de nome ksnode.cfg que contém toda a informação para a instalação automatizada do nó. É possível incluir mais opções no menu, como por exemplo o arranque em modo recuperação (rescue) ou teste de memória RAM, opções vulgarmente oferecidas nos CD de arranque (live boot CD) e que não são mostradas aqui.

Para o nó Interface com o Utilizador, os ficheiros de menu AC1100FD e AC1101FD são semelhantes mas contêm o nome de outro ficheiro kickstart, ksnode-ui.cfg, dado que a instalação/configuração deste nó é diferente dos nós computacionais.

```
# Cluster - Menu para instalação de nó computacional
default vesamenu.c32
prompt  0                   # Sem prompt
timeout 50                  # Esperar 5 segundos
menu background splash.jpg  # Imagem de fundo

label kickstart
```

```
   menu label ^Kickstart Install
   menu default
   kernel vmlinuz
   append initrd=initrd.img ks=ftp://172.17.0.254/centos6-
32/cfg/ksnode.cfg
label local
   menu label Boot from ^Local drive
   localboot 0
```

(a)

```
# Cluster - Menu para arranque local
default vesamenu.c32
prompt  0                   # Sem prompt
timeout 50                  # Esperar 5 segundos
menu background splash.jpg  # Imagem de fundo

label kickstart
   menu label ^Kickstart Install
   kernel vmlinuz
   append initrd=initrd.img ks=ftp://172.17.0.254/centos6-
32/cfg/ksnode.cfg
label local
   menu label Boot from ^Local drive
   menu default
   localboot 0
```

(b)

Figura 9. Ficheiros de menu para arranque por PXE: **(a)** AC1101 para instalação do nó; **(b)** AC1100 para arranque do disco local.

4.5.2. Configuração do Servidor DHCP

O pacote de software correspondente ao servidor DHCP tem o nome "dhcp". A configuração do servidor dhcp é feita no ficheiro /etc/dhcp/dhcpd.conf no qual essencialmente se definem as redes e IP a atribuir, e no ficheiro /etc/sysconfig/dhcpd onde se podem definir algumas opções, uma das quais é o porto de rede onde o serviço é oferecido. No caso do cluster o serviço DHCP é só para a rede privada e estando o Servidor de Software ligado à mesma pelo porto eth1, define-se a opção DHCPDARGS=eth1. A figura 10 mostra os comandos e ações iniciais a efetuar para configurar o serviço DHCP.

```
# instalar servidor dhcp
yum -y install dhcp
# editar /etc/sysconfig/dhcpd e definir "DHCPDARGS=eth1"
vim /etc/sysconfig/dhcpd
# configurar serviço para iniciar no arranque do PC
chkconfig dhcp on
```

Figura 10. Comandos para configurar o servidor DHCP.

O ficheiro /etc/dhcp/dhcpd.conf é dado como exemplo para a rede privada 172.17.0.0/23 na figura 10. Na primeira parte são definidas algumas opções relacionadas com o PXE [redhat

2014], seguidas da especificação da rede e da identificação dos clientes PXE, indicação do IP do servidor tftp e programa de arranque pela rede (network bootstrap program). Na segunda parte, cada nó do cluster é identificado pelo seu MAC e feita uma correspondência 1:1 com um IP fixo atribuído. Para os nós computacionais, também é atribuído o hostname.

No exemplo dado, todos os nós estão configurados para efetuar um arranque local do seu disco (IP tipo 172.17.0.i) enquanto o nó "node2" está configurado para efetuar a instalação (IP tipo 172.17.1.i). Deste modo torna-se extremamente simples alterar entre arranque local e instalação de um nó, bastando para o efeito alterar entre 0 e 1 o terceiro campo do IP do nó no ficheiro dhcpd.conf seguido da execução do comando "service dhcp restart".

```
# Ficheiro DHCP para clusters com arranque por PXE
# Rede privada 172.17.0.0/23 ou 172.17.0.0/255.255.254.0
# Servidor de Software:     172.17.0.254
# Nó Interface Utilizador:  172.17.0.253

# pxe options
option space pxelinux;
option pxelinux.magic code 208 = string;
option pxelinux.configfile code 209 = text;
option pxelinux.pathprefix code 210 = text;
option pxelinux.reboottime code 211 = unsigned integer 32;

subnet 172.17.0.0 netmask 255.255.254.0 {
   # PXE boot
   class "pxeclients" {
      match if substring (option vendor-class-identifier, 0, 9) =
"PXEClient";
      next-server 172.17.0.254;
      filename "pxelinux/pxelinux.0";
   }
}

group {
default-lease-time 31536000;
max-lease-time     63072000;
# nodes 1-10
host node1 {
   hardware ethernet 00:0F:1F:EB:0F:34; fixed-address 172.17.0.1;
   option host-name "node1";}
host node2 {
   hardware ethernet 00:0D:56:DB:53:24; fixed-address 172.17.1.2;
   option host-name "node2";}
...
host node10 {
   hardware ethernet 00:0D:56:DB:61:AB; fixed-address 172.17.0.10;
   option host-name "node10";}
# Nó Interface Utilizador
host juno {
   hardware ethernet 00:19:DB:2E:09:41;
```

```
   fixed-address 172.17.0.253;}
}
```

Figura 11. Ficheiro dhcpd.conf para a rede privada. Exemplo para 10 nós computacionais.

4.6. Configuração do Servidor NTP

A sincronização dos relógios dos vários computadores que compõem o cluster é um requisito para o seu bom funcionamento, não só pela coerência dos ficheiros de registos (logs) mas também uma exigência do software, como o gestor de recursos.

Os servidores NTP funcionam com uma estrutura hierárquica por níveis designados stratum, com o relógio de referência no nível 0 (nível de topo) e os servidores propriamente ditos nos níveis seguintes 1,2,...,etc. Cada computador pode atuar simultaneamente como cliente dos níveis superiores e servidor dos níveis inferiores. O nível pode ser visto como uma distância ao relógio de referência.

No cluster, o servidor NTP local é o Servidor de Software, sendo os nós computacionais e o nó de Interface com o Utilizador seus clientes, ou seja, acertam o seu relógio pelo do Servidor de Software. Por outro lado, o Servidor de Software é cliente de servidores na Internet disponibilizados para o efeito pela própria distribuição Linux CentOS.

O pacote de software correspondente ao servidor NTP tem o nome "ntp" e a sua configuração é feita no ficheiro /etc/ntp.conf. Este pacote já deve ter sido instalado aquando da instalação do sistema operativo e o serviço ntpd a funcionar como cliente de servidores na Internet. Para configurar o serviço também como servidor NTP da rede local, basta editar o ficheiro /etc/ntp.conf e introduzir junto ao comentário "# Hosts on local network" a linha "restrict 172.17.0.0 mask 255.255.254.0 nomodify nopeer noquery" onde os últimos termos se destinam a não permitir que os clientes inquiram ou alterem o estado do serviço ntpd.

5. Conclusões

Nesta primeira parte do artigo foi apresentada uma perspetiva geral de um cluster (agregado) de computadores pessoais ligados entre si por uma rede privada ethernet e da sua instalação e operação.

As suas características mais apelativas são o seu relativo baixo custo global, facilidade de obtenção dos seus componentes de hardware e software, grande potencial de aumento da capacidade computacional e a criação de uma plataforma de programação paralela e distribuída. Por estas razões, este tipo de tecnologia tem vindo cada vez mais a ser utilizada pelo que importa tomar contacto com ela e difundir a sua aprendizagem .

Referências

Granjal J. (2013), "Gestão de Sistemas e Redes em Linux", FCA

Intel (1999), "Preboot Execution Environment (PXE) Specification Version 2.1", www.pix.net/software/pxeboot/archive/pxespec.pdf [2008-09-29]

Quinn Michael J. (2004), "Parallel Programming in C with MPI and OpenMP", McGraw-Hill Higher Education.

Redhat (2011), "Red Hat Enterprise Linux 6 Managing Confined Services"

Redhat (2012), "Red Hat Enterprise Linux 6 Developer Guide"

Redhat (2013), "Red Hat Enterprise Linux 6 Deployment Guide"

Redhat (2014), "Red Hat Enterprise Linux 6 Installation Guide"

Sloan Joseph D. (2005), "High Performance LINUX Clusters", O'Reilly Media.

Wikipedia, "Beowulf cluster", https://en.wikipedia.org/wiki/Beowulf_cluster [2017-11-17]

Paulo Shirley, Professor Auxiliar no Departamento de Ciências e Tecnologia (DCeT), Secção de Informática, Física e Tecnologia (SIFT). Coordenador da Licenciatura em Informática no triénio 2014–2016 e Vice-Coordenador do Mestrado Tecnologias e Sistemas Informáticos Web no biénio 2014–2015. Licenciado em Engenharia Electrotécnica e de Computadores em 1988 pelo IST-UTL. Obteve os graus de Mestre (perfil de Controlo e Robótica) e de Doutor em Eng. Electrotécnica e de Computadores, pelo IST-UTL em 1993 e 2003 respetivamente. Tem como áreas de interesse, a intersecção da Informática (Computer Science) com a área do Controlo Automático, nomeadamente a área da "Computação de Alto Desempenho" aplicada a problemas de otimização.