

Clusters



Trabalho elaborado por:
980368 - Sérgio Gonçalves Lima
1010949 - Nisha Sudhirkumar Chaganlal

Clusters

- O que é um Cluster?
 - Tipos de Clusters
 - Razões para usar um cluster
 - Aplicações de Clusters
 - Era da super-computação
-

O que é um Cluster?

Entende-se por cluster um conjunto de dois ou mais sistemas, habitualmente denominados *nodes*, que cooperam entre si transparecendo para o exterior (utilizador) como sendo apenas um só sistema. Assim, é possível obter soluções que disponibilizam uma enorme quantidade de recursos, aumentando a performance e fiabilidade do próprio sistema.

Existem vários tipos de cluster:

- Alta Disponibilidade (High Availability (HA) and Failover)

Estes modelos de clusters são construídos com o objectivo de proporcionar uma disponibilidade de serviços e recursos de forma ininterrupta através do uso da redundância implícita no sistema. A ideia geral consiste em, caso um nodo do cluster falhe, as aplicações ou serviços que estivessem a ser executados nesse nodo, possam ser disponibilizados através de um nodo alternativo. Estes tipos de cluster são utilizados para bases de dados de missões críticas, correio electrónico, servidores de arquivos e aplicações.

- Load Balancing

Neste modelo de clustering, existe a distribuição de tráfego ou de processamento de pedidos externos pelos vários nodos do cluster de modo a não sobrecarregar apenas um dos nodos do cluster. Todos os nodos no cluster possuem a capacidade de lidar com os pedidos efectuados. Caso um nodo falhe, os pedidos são redistribuídos entre os restantes nodos disponíveis. Este tipo de distribuição de pedidos é útil em contextos de *web hosting*.

- Combinação entre Alta disponibilidade e Load Balancing

Este tipo de clustering resulta da combinação das características dos dois tipos de cluster, aumentando assim a disponibilidade e a escalabilidade de serviços e recursos. Este tipo de configuração de cluster é frequentemente utilizado em servidores web, mail, notícias ou ftp.

- Processamento distribuído ou processamento paralelo

Com esta configuração é possível aumentar a disponibilidade e performance para as aplicações e particularmente grandes tarefas computacionais. Uma tarefa computacional destas características pode ser dividida em pequenas tarefas que são distribuídas pelos nodos circundantes para serem executadas ao invés de serem executadas por um só super computador. O tipo mais comum de cluster de processamento distribuído/paralelo é um Beowulf cluster. Estes clusters são frequentemente utilizados na área científica ou na área de análises financeiras, tarefas típicas para a exigência de alto poder de processamento.

Razões para usar um cluster

Clusters ou combinações de clusters são usadas quando é crítico que determinado conteúdo ou serviço esteja disponível e/ou seja processado o mais rápido possível. Internet Service Providers e sites de comércio electrónico requerem frequentemente uma alta taxa de disponibilidade, balanceamento de carga e escalabilidade. Clusters ganharam importância na indústria filmatográfica para a renderização de gráficos e animações de alta qualidade. Os clusters Beowulf são usados em investigação científica, engenharias e finanças. São usados porque permitem aumentar a escalabilidade, são fáceis de gerir, tem uma boa disponibilidade, permitindo grande poder de processamento a um preço baixo.

Assim, o uso de clusters visa atingir os seguintes objectivos:

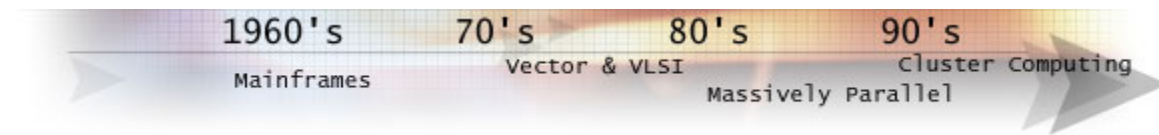
- Melhoria no desempenho – maior performance a baixo custo
- Melhoria da disponibilidade – gestão de falhas
- Single System Image (SSI) – “look and feel” de um sistema único
- Escalabilidade – hardware e aplicações

- Comunicação rápida - redes & protocolos
- Equilíbrio de carga - CPU, rede, memória, disco
- Segurança e Criptografia - clusters de clusters
- Facilidade de administração
- Facilidade de programação - APIs simples
- Aplicabilidade - suporte de aplicações cluster-aware e non-aware

Exemplos práticos de clusters em funcionamento:

- Meteorologia e Modelação de ecossistemas
- Matemática
- Física
- Biotecnologia
- Química
- Modelação financeira
- Geofísica

A Era da super-computação



Super-computadores são os sistemas mais avançados no que diz respeito à capacidade de processamento. Muitos dos “truques” que permitiram aos Super-computadores melhor performance, são agora regularmente incorporados nos computadores pessoais fazendo um PC dos dias de hoje mais poderoso que um super computador de 15 anos.

Os primeiros computadores da era moderna eram do tamanho de quartos e eram programados por cartões perfurados. A computação nos anos 60 e 70 era dominada por mainframes máquinas de grandes dimensões que suportavam milhares de utilizadores em simultâneo, acedendo ao sistema através de terminais.

O desenvolvimento das tecnologias LSI (Large-Scale Integration) e VLSI (Very-Large-Scale Integration) nos anos 70 e 80 permite que milhares de componentes possam ser integrados num simples chip de hardware aparecendo assim o microprocessador. Adicionalmente, o desenvolvimento da “vector CPU architecture” significa que uma larga quantidade de dados matemáticos possam ser tratados muito mais rápido que com outro tipo de processadores, apesar disto, a velocidade de processamento diminuía consideravelmente quando se tratavam de instruções complexas. Os Super-computadores Cray foram os mais rápidos durante os anos 70 e 80.

O número de processadores que pode ser agrupado num tradicional “Vector processor supercomputer” é limitado pela capacidade de aceder a uma larga quantidade de memória partilhada. Os sistemas paralelos foram desenhados para resolver este problema, ligando directamente os processadores a memórias locais através de uma rede, aumentando assim a escalabilidade, ficando limitado apenas pelos custos.

Porque muitos problemas executados por Super-computadores podem ser facilmente divididos em partes mais pequenas que podem ser trabalhadas em simultâneo. Tradicionalmente, os Super-computadores podem frequentemente ser substituídos por processamento em paralelo (Clusters) que são nada mais que

máquinas individuais programadas para funcionar como só um computador. Beowulf é provavelmente o tipo de processamento em paralelo mais conhecido hoje em dia. Foi desenvolvido por Donald Becker e Thomas Sterling em 1994 para a NASA. Consistia em 16 processadores 486-DX4 ligados por uma Ethernet.

Hoje em dia, os clusters Beowulf são usados para resolver problemas que requerem triliões de operações de computação executadas rápida e eficientemente. A ciência moderna cada vez mais depende da possibilidade de correr simulações em computadores que são muito complexas, demoradas ou caras para serem estudadas de outra forma. O poder de processamento de um cluster Beowulf é aplicado a complexos problema de computação tais como o genoma humano, previsão meteorológica, simulações de explosões nucleares, entre outros.

No futuro...

Simular uma simples batida de coração num modelo tridimensional do coração pode demorar vários dias a ser processado. Com os futuros avanços da tecnologia em termos de velocidade de processamentos, armazenamento de informação, velocidade de comunicação e memória, os super computadores vão com certeza aumentar exponencialmente a sua capacidade de cálculo.

Factos e vantagens

Um Cluster pode crescer de forma a ficar um sistema muito grande, sendo possível, com algum trabalho, ligar milhares de máquinas formando um cluster gigante.

Substituir uma máquina avariada dentro de um cluster é algo trivial comparando com o arranjo de uma parte de um sistema SMP .

Problemas/Desvantagens

Com algumas excepções, o hardware de rede não foi desenhado a pensar em processamento paralelo, logo, a latência é muito alta e a largura de banda relativamente baixa comparando com um SMP e um processador ligado.

Existe pouco software de suporte para utilizar clusters como um único sistema.

Bibliografia e Links úteis

<http://www.clubedohardware.com.br/super.html>
<http://www.clubedohardware.com.br/cluster.html>
<http://support.microsoft.com/default.aspx?kbid=309395>

<http://www.aspsys.com/clusters/overview/>
<http://www.aspsys.com/clusters/beowulf/disciplines/>

<http://www.windowsclusters.org/hardware.htm>
<http://www.epcc.ed.ac.uk/HPCinfo/hware-clusters.html>
<http://yara.ecn.purdue.edu/~pplinux/ppcluster.html>
http://www.penguincomputing.com/products/clusters/cluster_hardware.php

Curiosidades

<http://www.top500.org/> --> Top 500
<http://www.offmyserver.com/cgi-bin/store/cluster.html> --> Fazer um cluster
<http://www.chem.arizona.edu/theochem/beowulf/>