

Clusters: Uma alternativa viável para Processamento Distribuído

Mateus Victorio Zagonel¹, Cristian Cleder Machado²

¹Aluno do Curso de Ciência da Computação – Universidade Regional Integrada (URI)
Caixa Postal 709 – 98.400-000 – Frederico Westphalen – RS – Brasil

²Professor do Curso de Ciência da Computação – Universidade Regional Integrada
(URI)
Caixa Postal 709 – 98.400-000 – Frederico Westphalen – RS – Brasil.

inf16755@uri.edu.br, cristian@cristian.com.br

Abstract. *This paper addresses the Cluster architecture, which is used for running programs that require high processing power or high data availability. This architecture was created in order to be an alternative to supercomputers, because it has a simpler architecture and better cost x benefit when it comes to parallel processing (concurrent) and distributed. I will explain in this paper a brief definition of Cluster, the used architecture and Cluster types based on the application.*

Resumo. *O presente artigo trata da arquitetura Cluster, que é utilizada para execuções que demandam de grande poder de processamento ou grande disponibilidade de dados. Esta arquitetura foi criada com o objetivo de ser uma alternativa viável em relação aos supercomputadores, pois apresenta uma arquitetura mais simples e melhor custo x benefício em se tratando de processamento paralelo (simultâneo) e distribuído. Explicarei no decorrer do artigo uma breve definição de Cluster, bem como a arquitetura utilizada e os tipos de Cluster de acordo com a aplicação.*

1. Introdução

Com a evolução da computação na década de 80 e devido aos supercomputadores terem um preço nada acessível, surgiram os clusters que correspondem a duas ou mais máquinas interconectadas pela rede e coordenadas por um software chamado Middleware.

Os cientistas da época não se conformavam com a ideia de ter que investir em um supercomputador específico (que era absurdamente caro), então pensaram na seguinte questão: Por que não agrupar os computadores disponíveis através de um software e coordenar o processamento? A ideia deu tão certo que hoje a maioria dos quinhentos maiores computadores do mundo utiliza a arquitetura de Cluster.

Atualmente os Clusters não são utilizados apenas para grande processamento, mas para outras tarefas como alta disponibilidade de dados e balanceamento de carga que tratarei com mais detalhes no título quatro do presente artigo.

2. Definição

Um cluster corresponde a duas ou mais máquinas interconectadas por meio de uma rede local (LAN), trabalhando em conjunto para realizar determinada tarefa de forma que o usuário “enxergue” o sistema como um todo. (PITANGA, 2003)

Não devemos confundir um Cluster com um Supercomputador, pois um cluster pode ser formado com máquinas comuns de diferentes configurações, projetadas para

uma infinidade de tarefas e interconectadas pela rede (fracamente acoplados). Por outro lado um supercomputador é formado por um conjunto de núcleos idênticos, dispostos em um mesmo circuito (fortemente acoplados) se comunicando por meio de um barramento e memória compartilhada, implementados para uma tarefa específica que necessite de grande poder de processamento. (ZEM e BRITO, 2006);

3. Arquitetura de Clusters

Para melhor entendimento utilizarei a figura 1, abaixo, em que há um desenho com uma arquitetura de um cluster geral, seguida de uma breve descrição de cada componente da mesma.

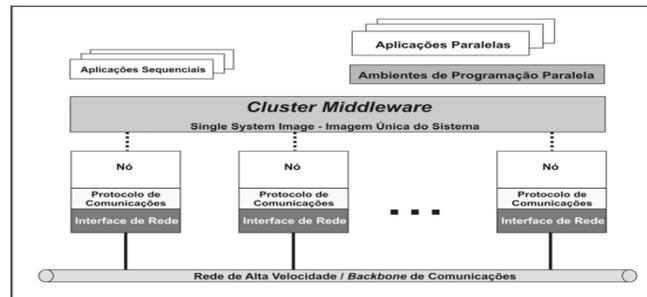


Figura 11 - Exemplo da Arquitetura de um Cluster

*BackBone ou Rede de Comunicação: Corresponde a forma de comunicação empregada, podendo ser uma rede Ethernet ou Gigabit, por exemplo.

*Interface de Rede: Correspondem aos adaptadores de rede.

*Protocolo de Comunicações: São as regras que serão utilizadas para que ocorra a comunicação.

*Nó ou Máquina: Qualquer máquina que faça parte do Cluster.

*Middleware: Software executado no Sistema Operacional responsável por “esconder” a arquitetura que é utilizada no Cluster e mostrar ao usuário o sistema como se funcionasse com apenas um computador.

*Ambiente de Programação Paralela: Ambiente para criação de novas aplicações paralelas que utilizem o poder de processamento de um Cluster.

*Aplicativos Paralelos: Programas paralelos que serão executados em cada nó do cluster obtendo o paralelismo.

*Aplicativos Sequenciais: Programas sequenciais que não utilizarão o poder de processamento paralelo de um Cluster. (ZEM e BRITO, 2006);

4. Tipos de Clusters

Neste título serão tratados três tipos de cluster, cada um dos três tipos de Cluster é projetado para atender a uma necessidade específica, variando a arquitetura de acordo com a aplicação.

4.1. Cluster de Alto Desempenho

Esse tipo de Cluster é utilizado para tarefas que necessitem de grande processamento, ou seja, programas com granularidade grossa. Os clusters de alto desempenho seguem o princípio de dividir uma grande tarefa em inúmeras tarefas menores e são utilizados para computação científica, análises financeiras, pesquisas climáticas e simulações. (ZEM e BRITO, 2006; SALLES ET AL, 2009);

Sua arquitetura, conforme nos mostra a Figura 2 abaixo, utiliza um nó controlador que envia as tarefas para os demais nós Back-End de forma que a execução seja paralelizada. (SALLES ET AL, 2009);

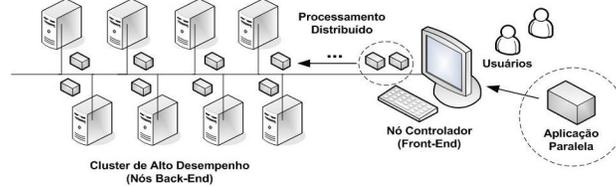


Figura 12 - Exemplo de um Cluster de Alto Desempenho

4.2. Cluster de Alta Disponibilidade

Os Clusters de Alta Disponibilidade tem a função de garantir que um serviço esteja ativo em mais de uma máquina. Para que isso seja possível, utilizam replicações dos dados ou serviços. (PITANGA, 2003);

Conforme nos mostra a figura 3 abaixo, imagine, por exemplo, um usuário acessando um banco de dados por meio da rede pública, sua requisição é enviada para o servidor ttisrv1, porém esse servidor está em manutenção. Para que o usuário tenha sua solicitação atendida à mesma é redirecionada para o servidor ttisrv2 que possui exatamente as mesmas informações que o servidor ttisrv1 e atenderá a requisição. Mas como o servidor ttisrv2 sabe que o servidor ttisrv1 está em manutenção? Existe uma técnica denominada heartbeat, em que através de uma rede privada, os integrantes do cluster trocam mensagens sinalizando que estão ativos. (ZEM e BRITO, 2006);

Poderiam ser usados inúmeros servidores e não apenas dois, o número de servidores depende da necessidade/importância do serviço.

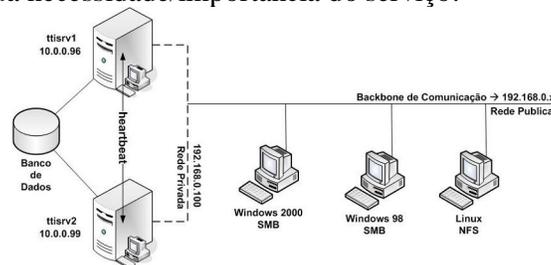


Figura 3 - Exemplo de Cluster de Alta Disponibilidade

4.3. Cluster de Balanceamento de Carga

Tem por função distribuir as solicitações entre os nós do Cluster. Diferentemente dos Clusters de alta disponibilidade, os cluster de balanceamento de carga são apenas responsáveis por distribuir as requisições, portanto não possuem “cópias” dos dados em todos os servidores, pois funcionam como redirecionadores de solicitações. (PRADO e SILVA 2010 apud BATISTA 2007);

São utilizados em serviços de Comércio Eletrônico, Provedores de Internet e sistemas que possuem grandes quantidades de acessos/requisições e necessitam de processamento em tempo real. (PITANGA, 2003);

Os integrantes do Cluster de Balanceamento de Carga, assim como nos Clusters de Alta disponibilidade, devem possuir comunicação para que uma requisição não seja enviada para um servidor que está inativo. (PITANGA, 2003);

5. Conclusão

Os Clusters correspondem a uma alternativa viável para supercomputação e para formação de sistema distribuídos, pois projetar um cluster não necessita de grande

complexidade, em comparação com um supercomputador. A grande dificuldade na área de clusters consiste na criação de softwares que gerenciem e que utilizem a capacidade máxima de cada nó da estrutura, pois os programas devem ser paralelizados dinamicamente e o Middleware deve distribuir as tarefas de forma inteligente de acordo com o processamento de cada nó.

A área de computação em Cluster está em expansão principalmente para sistemas de grande porte. As grandes empresas têm duas opções atualmente, ou constroem seu próprio cluster, ou contratam um serviço de Cloud Computing (Computação em Nuvem) que curiosamente na maioria dos casos é formada por Clusters. Cabe aos administradores realizarem um estudo se é mais viável criar uma estrutura de Cluster e contratar um administrador competente ou contratar um serviço de Cloud Computing, de certa forma “abrindo mão” do gerenciamento de suas informações.

Referências

- Prado, C. L. e Silva, J. M. A. (2010). Sobre Clusters de Computadores. Em *Aplicação de Cluster Beowulf em Instituições de Ensino*, páginas 23-41. Publicado na Internet em Portallivre.
- Zem, J. L. e Brito, S. H. B. (2006). Sobre Arquitetura de Clusters. Em *Monitoramento Distribuído de Clusters e Grids Computacionais utilizando o Ganglia*, páginas 1-6. Publicado na I Jornada Científica da Universidade Brasileira de Tecnologia, Recife (PE).
- Salles, D. ET AL (2009). Sobre Clusters de Alta Performance. Em *Cluster HPC – High Performance Computing*, 1-5. Publicado na FCT - Faculdade de Ciências e Tecnologia.
- Pitanga, M. (2003) “Computação em Cluster”, <http://www.clubedohardware.com.br/artigos/153>, Setembro.