

Utilização de hardware legado para o ensino de CAD

David Beserra*, Rubens Karman, Felipe Oliveira,
Alexandre Borba, Jean Araujo, Alberto Araújo
Unidade Acadêmica de Garanhuns
Universidade Federal Rural de Pernambuco, UFRPE
Garanhuns, Pernambuco
dw.beserra@gmail.com*

Felipe Fernandes
Pós-Graduação em Sistemas e Computação
Instituto Militar de Engenharia
Rio de Janeiro, Brasil

Abstract— Computação de Alto Desempenho (CAD) é um tema atual e que demanda por uma infraestrutura adequada para seu ensino em cursos de graduação. Entretanto, nem todas as instituições de ensino têm recursos financeiros para proporcionar estruturas computacionais dedicadas exclusivamente ao ensino de tópicos de CAD. Neste trabalho, é apresentada uma infraestrutura para o auxílio no ensino de CAD construída com a utilização de computadores destinados ao descarte. O uso de tais computadores destinados originalmente ao descarte colaborou também na transmissão de princípios de TI Verde.

Keywords—CAD; Infraestrutura; TI Verde; Reúso;

I. INTRODUÇÃO

Computação de Alto Desempenho (CAD) é um termo genérico utilizado para se referir a um conjunto de aplicações que fazem uso intensivo de um ou mais tipos de recursos computacionais e são de natureza técnico-científica [1]. Exemplos de aplicações de CAD são a resolução de sistemas lineares de grandes dimensões [2] e a simulação de estruturas físicas de larga escala, como o universo [3]. A estrutura mais comum para computação paralela são os aglomerados de computadores, mais conhecidos pelo termo em inglês *clusters*, que são utilizados para a execução de programas paralelos que utilizam a Interface de Passagem de Mensagens (MPI) [4].

Apesar de sua importância, é constatado que o ensino de tópicos relativos a computação paralela apresenta desafios, tanto em instituições de ensino superior nacionais [5], [6], quanto estrangeiras [7]. O principal desafio apontado nestes trabalhos é a falta de infraestrutura adequada para a realização de atividades práticas que, além de reduzirem as possibilidades de aprendizado, diminuem a motivação dos discentes para a participação nas atividades da disciplina [5].

Em paralelo à necessidade de proporcionar uma melhor qualidade no ensino de tópicos de CAD nos cursos de graduação em computação, existe também uma demanda global por uma maior conscientização em questões ambientais. Em computação, essa preocupação é refletida no conceito de TI Verde, que busca reduzir o impacto do uso de tecnologias da informação no meio ambiente [8]. Nesse contexto, o reaproveitamento de hardware obsoleto, destinado ao descarte por uma universidade, pode ser considerado uma política socioeducativa de TI verde.

Neste artigo, descrevemos a construção de um ambiente de *clusters* virtualizados para uso em atividades de ensino de tópicos relacionados a sistemas de hardware paralelos e sua

programação. Os *clusters* foram hospedados em hardware reciclado, que originalmente estava destinado ao descarte. O ambiente foi desenvolvido sob o nome de Projeto ReciCluster, dada a dupla função de ensinar tópicos relativos a CAD e transmitir aos envolvidos no projeto princípios de TI verde, como o reúso. Este artigo está organizado da seguinte forma, a Seção II apresenta uma seleção de trabalhos relacionados. A Seção III descreve o processo de implementação do projeto ReciCluster. A Seção IV discute os resultados obtidos. A seção V encerra o artigo com as considerações finais e propostas de trabalhos futuros.

II. TRABALHOS RELACIONADOS

Existem trabalhos na literatura onde são descritas soluções para proporcionar infraestrutura para disciplinas que contém tópicos de CAD. Em um desses trabalhos, é apresentado um *cluster* direcionado para uso em atividades educacionais e científicas [7]. O *cluster* apresentado em [7] é constituído de Máquinas Virtuais (VMs) construídas com o uso do virtualizador Kernel-based Virtual Machine (KVM). As VMs estão hospedadas nos servidores de um laboratório de ensino e o *cluster* está apto a executar aplicações que façam uso de MPI ou de MapReduce. Uma desvantagem da estrutura apresentada em [7] é que o *cluster* só pode ser utilizado pelos discentes em períodos nos quais o laboratório que o hospeda não está em utilização, como em finais de semana e feriados (períodos em que os estudantes tipicamente não estão interessados em realizar atividades acadêmicas).

Em [9], é apresentado um mecanismo que converte um laboratório de ensino em um *cluster* para uso em atividades de CAD e depois o retorna à configuração original, sem utilização de máquinas virtuais. A solução apresentada em [9] não é necessariamente focada para o ensino de tópicos de CAD, embora também possa ser utilizada para este fim. A solução descrita em [9] apresenta as mesmas limitações que a descrita em [8]. Junto a isto, em [10], é apresentado um ambiente onde um laboratório é convertido em um *cluster* e as atividades de outras disciplinas são executadas em VMs, que foram implementadas com o virtualizador VMWare. A solução apresentada em [10] elimina o problema da pouca disponibilidade para uso que ocorre nas soluções apresentadas em [8] e [9]. Todavia, nessa abordagem, os nós do *cluster* provocam queda de desempenho nas VMs de uso geral, pois não foram definidos limites para a quantidade de recursos que o *cluster* pode utilizar. As VMs de uso geral utilizaram apenas a porcentagem de recursos não utilizadas pelo *cluster*, o que

pode provocar impactos negativos no uso de tais VMs pelos estudantes de outras disciplinas.

Outra solução é apresentada em [6], onde estudantes implementam um ambiente de *cluster* dentro de uma nuvem computacional e executam algumas aplicações MPI. Contudo, os autores reiteram que a maior dificuldade enfrentada na elaboração de seu trabalho foi a indisponibilidade de infraestrutura, pois os computadores utilizados eram de um departamento alheio ao departamento onde a disciplina foi lotada.

Em [5], é implementado um *cluster* virtualizado para o ensino de programação paralela em um laboratório universitário e as VMs foram construídas com o virtualizador VirtualBox. A solução descrita em [5] não apresenta restrições quanto à disponibilidade para uso e utiliza no máximo 50% dos recursos de memória disponível nos hospedeiros. Entretanto, tendo em vista os computadores hospedeiros utilizarem o Windows 7 como SO, o sistema não apresenta escalabilidade de desempenho, conforme observado em [4]. Isto limita as possibilidades da solução descrita em [5], pois é necessária alguma escalabilidade para que discentes possam entender determinados conceitos fundamentais de CAD, como o conceito de *Speedup*. O *Speedup* é definido como o ganho de desempenho obtido por uma aplicação paralela executada em p processadores em relação a uma aplicação sequencial [11].

Um ponto em comum entre os trabalhos citados até agora nesta seção é que todos necessitam de hardware com grande capacidade de processamento e de armazenamento em memória primária para serem utilizados no ensino de CAD, e que nem sempre podem ser disponibilizados pelas universidades. Logo, uma alternativa à falta de infraestrutura para o ensino de tópicos de CAD é o reaproveitamento de computadores obsoletos destinados ao descarte. Tais computadores, embora obsoletos para utilização em outros departamentos, disciplinas e atividades em geral de uma universidade, ainda podem ser aproveitados em disciplinas de arquitetura de computadores e sistemas distribuídos.

Estima-se que 75% de todos os computadores vendidos até hoje estão obsoletos ou descartados [12]. De posse dessa informação, é correto afirmar que o reaproveitamento de hardware obsoleto, além de proporcionar infraestrutura para o ensino de CAD, ajuda na preservação do meio ambiente, uma vez que impede que mais resíduos sejam descartados no mesmo, e promove a consciência ambiental nos discentes, ratificando os conceitos de TI verde. Um exemplo do uso de hardware legado em CAD é descrito em [13]. Em [13] foi construído um *cluster* que utilizou computadores heterogêneos em capacidade de processamento e armazenamento em memória primária que estavam abarrotados no almoxarifado da universidade dos autores do trabalho.

Foi observado que nenhum dos trabalhos citados utiliza simultaneamente de computadores obsoletos e de técnicas de virtualização na implementação de uma infraestrutura para o ensino de tópicos de CAD, o que constitui em um dos diferenciais deste trabalho em relação ao que já está disponível na literatura. Na próxima seção, é descrita a arquitetura e o processo de construção de um ambiente para o ensino de CAD com hardware legado e com uso de virtualização, além de como a mesclagem dessas tecnologias pode colaborar para o aumento de oportunidades de aprendizado para os discentes.

III. ARQUITETURA E IMPLEMENTAÇÃO DO RECICLUSTER

O projeto ReciCluster foi desenvolvido no decorrer da execução de um curso com duração de um semestre letivo em Tópicos Especiais em Sistemas Distribuídos, ministrado em nível de graduação. A disciplina tem uma ementa flexível que permite que o docente que a ministre adapte o conteúdo da ementa às necessidades correntes dos discentes. Neste curso, os alunos têm contato com vários tópicos relacionados a CAD, tanto em nível de infraestrutura lógica e física, quanto em nível de programação e uso dessas infraestruturas. Em nível de infraestrutura são trabalhados os conceitos de *cluster*, *grid* e *cloud*, com forte referência às tecnologias de virtualização, por serem o conjunto de tecnologias básicas da computação em nuvem [4].

Em nível de programação, são abordados os conceitos de programação paralela, com enfoque na implementação de aplicações que utilizam MPI e que são executadas em *clusters* de computadores. São abordados também conceitos de *Speedup* e correlatos. Dentre as aplicações paralelas desenvolvidas pelos alunos para execução em *clusters* estão a implementação de algoritmos para a multiplicação de matrizes, tal qual disponível em [14] e geração de números aleatórios em paralelo [15]. Como inovação, também são expostos na disciplina alguns conceitos de TI verde.

A partir da análise de trabalhos selecionados na literatura disponível, foi possível elencar alguns requisitos que devem ser atendidos em um projeto de infraestrutura para o ensino de CAD, como o ReciCluster:

- Proporcionar disponibilidade de acesso contínuo para a submissão de aplicações pelos discentes;
- Não provocar impactos nas outras atividades de ensino para as quais um laboratório de Informática é utilizado (quando se está implementando uma estrutura para o ensino de CAD sob uma infraestrutura já existente);
- Não sofrer impactos em seu desempenho pelo uso em paralelo da infraestrutura na qual está hospedada;
- Se possível, permitir que a construção e manutenção da infraestrutura de ensino seja conduzida pelos discentes;
- Se possível, reutilizar material ocioso, como computadores retirados do serviço por obsolescência.

A partir desses requisitos foi elaborada a arquitetura do ReciCluster, que é exibida simplificada na Figura 1. O ReciCluster é constituído por uma série de *clusters* de VMs instalados sobre uma infraestrutura hospedeira. Todos os *clusters* são implementados utilizando o virtualizador *VirtualBox*, que foi escolhido por não exigir tecnologia de virtualização assistida por hardware nos servidores, ao contrário de outras opções, como KVM e Xen [4]. Todos os *clusters* possuem um nó em cada servidor hospedeiro, exceto quando especificado pelo professor da disciplina. Esta decisão visa impedir que hajam diferenças significativas na velocidade de comunicação entre máquinas virtuais de um mesmo *cluster* pois, no caso de duas VMs de um mesmo *cluster* compartilharem um mesmo servidor, a comunicação entre elas seria mais rápida, podendo levar os alunos a erros na avaliação de desempenho e outras incoerências.

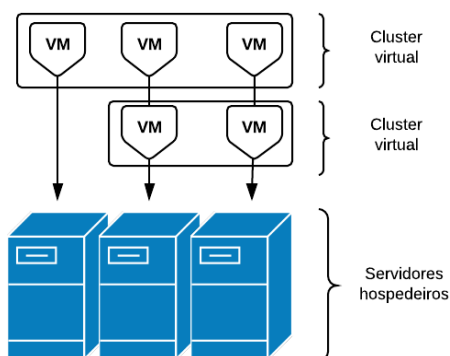


Fig. 1. Arquitetura do ReciCluster

Outra decisão de arquitetura foi a configuração individual de cada VM. Como o hardware do ambiente é limitado, cada VM utiliza no máximo 20% da capacidade de memória disponível e cada servidor instancia no máximo 4 VMs simultaneamente. Medida similar foi adotada para a divisão dos recursos de CPU, com cada VM também utilizando 20% do tempo de CPU. Esse tipo de controle é permitido pelo *VirtualBox*.

Dessa forma, cada servidor do ReciCluster aloca no máximo 80% dos recursos de processamento e memória para as VMs, com os 20% restantes sendo utilizados exclusivamente pelo servidor hospedeiro. Para aumentar a simplicidade na administração do ambiente, foi utilizado o mesmo SO, tanto nos servidores hospedeiros, quanto nas VMs. Dada a limitação do hardware utilizado, optou-se por utilizar um SO que demande por poucos recursos de processamento e memória para seu funcionamento, mas que, ao mesmo tempo fosse fácil de utilizar e de instalar pacotes de software com uso de ferramentas como o *apt-get*. Assim, optou-se por utilizar o SO *CrunchBang* Linux, distribuição Linux baseada no *Debian* Linux que necessita de apenas 128 MB de memória e um processador com frequência de operação de 1 GHz para que possa ser utilizado. Assim como o *Debian* e similares, o *CrunchBanb* também permite a instalação de pacotes de software com o uso de *apt-get*.

O processo de coleta de computadores aproveitáveis entre os que estavam destinados ao descarte foi realizado em conjunto com o Núcleo de Tecnologia da Informação (NTI) da universidade. O NTI cedeu o controle sobre os computadores destinados ao descarte e forneceu bancos de memória novos, oriundos de excedentes nos estoques para manutenção. Os computadores selecionados passaram por um processo de limpeza antes de serem utilizados. Observou-se heterogeneidade entre os computadores disponíveis, com a presença de computadores com processadores Intel *Pentium 4* e AMD *Sempron*.

Como decisão de projeto, optou-se por utilizar como servidores hospedeiros computadores com configurações similares, para garantir homogeneidade no desempenho das máquinas virtuais e aplicações. Neste sentido, optou-se por utilizar os computadores com processador *Pentium 4*, disponíveis em maior quantidade no estoque de descarte e mais modernos que os *Sempron*. Os computadores utilizados dispõem de um processador *Pentium 4*, com frequência de operação de 3.0 GHz, e memória principal de 2 GB, com

frequência de operação de 533 MHz. Os computadores também dispõem de discos rígidos com capacidade de 80 GB. O ambiente já montado pode ser visualizado na Figura 2.



Fig. 2. Ambiente ReciCluster montado

IV. DISCUSSÃO

Neste trabalho foi implementado um ambiente para hospedagem de *clusters* virtuais destinados ao auxílio no ensino de tópicos de CAD. Do ponto de vista do ensino de arquitetura de computadores, tal ambiente virtualizado possibilita a simulação de diversos tipos de arquiteturas de processadores, com suas respectivas instruções. Assim, os alunos podem testar como um determinado programa se comporta em uma arquitetura diferente, o que os auxilia a adaptar seu código para obter maior desempenho em uma arquitetura específica, o que é uma atividade típica de CAD, e tal possibilidade não foi mencionada nos trabalhos relacionados analisados. De um ponto de vista do ensino de programação paralela, o ReciCluster possibilita um ambiente no qual os discentes podem testar o desempenho de aplicações MPI, podendo observar como o ganho (ou não) de desempenho de uma aplicação paralela à medida que a quantidade de elementos de processamento disponíveis aumenta.

O processo de construção do ReciCluster também colaborou para reforçar a ideia do trabalho organizado em equipes como ferramenta eficiente para resolução de problemas, o que é necessário em qualquer tipo de organização humana, similarmente a [7] e [5]. Foi notada também uma maior motivação dos discentes para participar do processo de

aprendizagem ao trabalhar com algo construído por eles mesmos.

O ReciCluster também permite que ao menos 4 grupos de alunos sempre tenham suas aplicações sendo executadas simultaneamente, sem interferir nos programas e testes de outros alunos e com hardware lógico de desempenho similar. Outra possibilidade educacional do ReciCluster é proporcionar experiência prática com tecnologias de virtualização, como em [5], [6], [7] e [10]. Por sua vez, o uso de técnicas de virtualização na implementação do ambiente proporciona ao ambiente flexibilidade em configuração por meio de e um mecanismo de redundância, que garante uma maior confiabilidade contra interrupções de serviço ocasionadas por falhas de hardware (possibilidade que deve ser levada em consideração quando se está utilizando hardware legado e outrora destinado ao descarte). Quanto ao uso de hardware legado no projeto, além de permitir a criação de uma infraestrutura para o ensino de CAD, serviu também para demonstrar a importância da reutilização de recursos computacionais de TI de material de descarte como forma de diminuir os danos causados pela humanidade ao meio ambiente.

V. CONSIDERAÇÕES FINAIS

Neste trabalho, foi implementado um ambiente funcional para o ensino de tópicos de CAD com o uso de computadores destinados ao descarte pela universidade. O projeto contou com auxílio do corpo de técnicos de TI da universidade e dos alunos da disciplina. O ambiente ReciCluster permite que os alunos submetam aplicações e tenham contato com tecnologias de computação paralela, além de participarem da administração e manutenção do ambiente. Como trabalhos futuros pretende-se integrar os computadores com processador *Sempron* ao ambiente, porém com um tipo de VM de configuração padronizada que garanta desempenho igual independente do hospedeiro. Também pretende-se desenvolver parcerias com fornecedores locais de materiais de informática para recolhimento de computadores extras que possam ser utilizados em *clusters* educacionais. Será analisado também o consumo energético de computadores obsoletos em relação a computadores atuais, levando em consideração os princípios de TI verde, para atestar se também nesse aspecto é viável o reuso dos mesmos.

AGRADECIMENTOS

Os autores agradecem ao Núcleo de Tecnologia da Informação da UAG/UFRPE pelo auxílio na coleta e restauração dos computadores utilizados neste trabalho e a todos os discentes comprometidos com o projeto.

REFERÊNCIAS

- [1] J. Simons and J. Buell. "Virtualizing high performance computing." *ACM SIGOPS Operating Systems Review* 44.4 (2010): 136-145.
- [2] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, S. Bird, D. Nelson and L. Hernquist. Properties of galaxies reproduced by a hydrodynamic simulation. *Nature*, 509 (7499), pp. 177-182, 2014.
- [3] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529-551, April 1955. (*references*)
- [4] D. Beserra, R. Karman, K. Camboin, J. Araujo, A. Borba, and A. Araújo, "Análise do Desempenho de Sistemas Operacionais Hospedeiros de Clusters Virtualizados com o VirtualBox". In XII Workshop em Clouds, Grids e Aplicações (WCGA-SBRC), pp. 3-15, 2014.
- [5] D. Beserra, M França, C. Melo, Y. Sousa, S. Souto, E. Sousa, M. Andrade and A. Araújo, "Ambiente Virtualizado para Ensino de Programação Paralela e Computação em Cluster". In: XXXIII Congresso da Sociedade Brasileira de Computação / XXI Workshop sobre Educação em Computação, Maceió, 2013.
- [6] E. Coutinho and G. Paillard, "Aplicações de Computação de Alto Desempenho e Computação em Nuvem na Disciplina de Desenvolvimento de Aplicações Distribuídas", *International Journal of Computer Architecture Education (IJCAE)*, v.2, m.1, dezembro, pp 32-36, SBC, 2013.g
- [7] E. Johnson, P. Garrity, T. Yates, and R. Brown, "Performance of a Virtual Cluster in a General-Purpose Teaching Laboratory". In *Cluster Computing (CLUSTER)*, 2011 IEEE International Conference on Cluster Computing, pp. 600-604, IEEE, 2011.
- [8] G. Lunardi, R. Simões, and R. Frio "Green IT: an analysis of the main benefits and practices used by organizations". In *Revista Eletrônica de Administração (READ)*, pp. 1-30. 2014.
- [9] S. Wang, C. Li, H. Chang, Y. Chen and C. Hsu, "Developing an Automated Mechanism for Cluster Computing in Computerized Classroom", In *Procedia Computer Science*, Vol. 4, pp. 2068-2075, ISSN 1877-0509, <http://dx.doi.org/10.1016/j.procs.2011.04.226>, 2011.
- [10] N. Bergman, J. Funston and P. Gilfeather-Crowley, "Low-cost compute clusters in virtualized lab environments". In *Journal of Computing Sciences in Colleges*, 25(1), 159-166, ACM, 2009.
- [11] J. Gustafson, "Reevaluating Amdahl's law". *Communications of the ACM*, 31(5), 532-533, 1988.
- [12] A. Calvão, D. Rose, D. Ribeiro, M. Dalmeida, R. Almeida, and R. Lima, "O lixo computacional na sociedade contemporânea". In *ENINED- Encontro Nacional de Informática e Educação*, Cascavel, pp. 262-269, 2009.
- [13] D. Pinheiro, A. Medeiros, L. Barbosa, J. Medeiros and J. Neto, "Reutilização de equipamento legado na criação de um cluster", In *Anais da VI Escola Potiguar de Computação e suas Aplicações*. Mossoró, pp. 145-150, 2013.
- [14] D. Giménez, J.Mantas, and A. Vidal, "Introducción a la programación paralela". *Paraninfo Cengage Learning*, 2008.
- [15] Foster, I. (1995). *Designing and building parallel programs* (pp. 83-135). Addison Wesley Publishing Company.