

Parecer sobre o projeto de TCC

Título: Avaliação Automatizada de Desempenho de Sistemas de Virtualização GNU/Linux e VMware para Computação de Alto Desempenho (HPC)

Autor: Willian Soares da Silva

Orientadores: Prof. Dr. Claudio Schepke/Ms. Raul Dias Leiria

Instituição: Universidade Federal Do Pampa

Acredito que esse o trabalho seja um projeto de TCC, a escrita está aceitável, pequenos erros foram abstraídos, no geral foi possível entender o trabalho.

Abstract: Revisar a frase: “Na verdade o desempenho da tecnologia de virtualização depende do tipo de aplicação que será executado.”

Introdução

- Estrutura está ok, revisar melhor o conteúdo
- Talvez um excesso de comentários sobre energia (inclusive Green500)
- Recomendo cuidado com afirmações muito “fortes”, como a que menciona o uso de virtualização automaticamente se consome menos energia?

Seção Cap2 - HPC

Problemas na explicação da Tabela 1. A impressão que tenho é que os cores das GPUs estão contabilizados no top500. Ressalva em relação a documentação da máquina chinesa. A máquina da IBM power9 ter um desempenho melhor parece ser pelo fato de ter cores mais poderosos e um conjunto melhor (Ex: rede de 100 Gbps). A máquina chinesa parece usar mais GPUs (mais cores, mas com poder inferior).

Seção 2.4 (virtualização pra HPC) não ta ok. Sem uma melhor elaboração acho melhor colocar numa única seção virtualização e hypervisors dentro. Tomar cuidado, discutir a questão de virtualização e HPC. Existem controvérsias, vai usar melhor os recursos, mas causa overhead.

Seção 2.5 Hypervisor precisa ser melhor elaborada, recomendo usar a estrutura do capítulo 3 do livro citado do buyya, Mastering cloud computing. Recomendo ser uma única seção, com hypervisor e contêineres, e no final uma discussão sobre virtualização no teu cenário (HPC, benchmarks)

(KVM e VMware ESXi) e que será usado o NAS e outro benchmark. Justificativas para a escolha. Vmware é open source? VMWare tem algum limite de recursos para as VMs?

1 único nodo físico? 1 processo em cada VM?

- Descrição incompleta, revisar o que é relevante ser apresentado. Linha “CPUs” da tabela 5 me parece se referir ao número de threads com hyperthreading. Senti falta do total de memória RAM, informação dos discos disponíveis.

Possível discussão sobre a representatividade de um experimento com um único nodo usando MPI.

Métricas de desempenho

- Média do tempo de 16 execuções e descartar os dois melhores e os dois piores tempos.
- Não recomendo essa prática. Aumentar o número de repetições (30-40) e calcular o desvio padrão. Com mais execuções a tendência é se ter um aumento na confiabilidade, pois o desvio padrão tende a diminuir.

Capítulo 4 e 5

Escolha dos virtualizadores e NAS-MPI - falta de justificativa

Capítulo 6: Correção da afirmação “A computação de alto desempenho vem recebendo crescente atenção na área acadêmica e na área comercial, pelo fato de possuir vantagens como: tolerância à falhas e aproveitamento eficiente dos recursos computacionais disponíveis.”

Em relação a contribuição do script. Tenho dificuldade de ver uma contribuição científica nisso.

- Coletando e analisando resultados cada um dedica o tempo que tiver. Conheço pessoas que fazem tudo na mão, tem quem automatiza. Tenho dificuldade de ver como generalizar um script para diferentes benchmarks e arquiteturas, fazer algo semelhante a um arcabouço. Existem benchmarks que podem ter várias bibliotecas necessárias e dependências, teria como otimizar isso? Talvez isso seja o mais demorado e mais complexo. O usual é cada pesquisador ou grupo/laboratório ter uma estratégia de testes. Um script pra rodar (atualizar o binário), um script de coleta e análise. Outra ferramenta tende a ser usada para gráficos.

Se você fizer um script automatizado, isso certamente vai te economizar tempo no teu trabalho. Será que isso pode contribuir para outras pessoas?

- Cronograma, coleta dos dados precisa de dois meses?
- Referências precisam ser revisadas. Negrito em livros está o título, em conferências está o nome do evento.