

# O USO DE CLUSTER DE COMPUTADORES PARA DESENVOLVIMENTO DE APLICAÇÕES E WEB-SERVICES GEOESPACIAIS – UMA INTRODUÇÃO.

Carlos Eduardo Mota<sup>1</sup>, Sandro Braga Maia<sup>1</sup> e Irineu Capeletti<sup>1</sup>

<sup>1</sup> CEDES/CPRM-RJ

**RESUMO:** O uso da Internet vem crescendo e as empresas vêm direcionando seus sistemas de informações geográficas para a web em busca de visibilidade global, redução de custos e maior automação. Além disso, com o crescente aumento da quantidade de informação geoespacial, aliado ao aumento da resolução e a necessidade de disponibilizar esta informação para o usuário, seja dentro da própria empresa ou não, espera-se que os atuais sistemas centralizados sejam substituídos por uma estrutura de serviços distribuídos e dinâmicos. Para que pudesse solucionar essa grande quantidade de requisições, foram criados os “supercomputadores” que possuem alto custo, não compensando sua implementação na maioria das empresas. Por outro lado, temos como alternativa viável a utilização de cluster de computadores. Um cluster é um arranjo computacional de processamento paralelo ou distribuído que é formado por um conjunto de computadores independentes interconectados por rede, a trabalhar cooperativamente como um único e integrado recurso computacional. Como características fundamentais, inclui-se a elevação da confiança, distribuição de carga e desempenho. Os custos de implementação costumam ser menores do que a aquisição/manutenção de supercomputadores e, algumas vezes, o processamento é até mais eficiente. A CPRM, através do Centro de Desenvolvimento Tecnológico (CEDES), dispõe de um cluster de computadores pessoais para fins de desenvolvimento de tecnologias de processamento paralelo e distribuído para aplicações diversas na área de geociências, principalmente na área de GIS, Sensoriamento Remoto e Geofísica. Existem diversos tipos de clusters, mas os tipos que foram utilizados no experimento são o HPC – *High Performance Computing* e LB – *Load Balancer*. O equipamento utilizado constitui-se de 8 processadores Intel Core 2 Quad (45,328 Gflops - teórico) com dois discos de 500 GB e 4GB de RAM, um switch de 1Gbps e cabos de rede CAT-6 formando uma rede privada. O arranjo das máquinas é do tipo *beowulf* (computador/nó mestre e nós escravos). Uma nona máquina, de menor capacidade (AMD Sempron) serviu como servidor de arquivos e foi montado via NFS nos outros nós. Em outras palavras, o arranjo dispõe de 32 núcleos, a totalizar 362 Gflops. O sistema operacional utilizado foi o Scientific Linux 6.1, cuja distribuição inclui as bibliotecas de processamento paralelo (OpenMPI, PVM, ParallelPython) e programas de gerenciamento de carga Red Hat Piranha LVS). Para testar o arranjo, alguns programas e bibliotecas geoespaciais, como o GEOS, GDAL e MapServer, foram compilados e instalados em todos os nós. Para os ensaios de computação distribuída, o seguinte teste foi utilizado: a conversão em lotes de uma série de grids SRTM de padrão ASCII para Geotiff e para os ensaios de balanceamento de carga, foi utilizada web-services OGC (WMS), que eram conectados por uma quantidade crescente de clientes. Os resultados preliminares confirmam, em caráter geral, as vantagens que um cluster pode oferecer, e isto representa a abertura de novas frentes de pesquisa, principalmente na área de desenvolvimento de algoritmos mais robustos e adaptados ao tipo de computação principalmente na área de modelagens numéricas, classificação e operação de imagens de satélite e monitoramento *real-time*.

**PALAVRAS CHAVE:** CLUSTER, SERVIDORES GIS, ALTO DESEMPENHO