

DEFINIÇÃO DE AMBIENTE COMPUTACIONAL DE ALTO DESEMPENHO PARA MINERAÇÃO DE INFORMAÇÃO EM BANCO DE DADOS ASTRONÔMICOS

Autor: Murilo Moritz Parize

Orientador: Marcelo Massocco

Os grandes avanços na observação espacial, através de telescópicos terrestres e espaciais tem gerado uma grande quantidade e variedade de dados. A heterogeneidade dos dados vem da particularidade construtiva de cada sistema de observação, que tornam sua análise muitas vezes restrita a um determinado conjunto de características para cada galáxias analisada. A união desses dados produz uma análise mais detalhista, pois permite obter uma visão mais ampla das características do ponto que se deseja estudar e essa ideia está sendo um importante mecanismo de complementação das informações entre os projetos de observação e uma rica base dados para os pesquisadores. Porém, há problema na hora de integrar esses dados principalmente no que se refere aos recursos computacionais, a grande quantidade de interação entre as diversas base gera uma demanda de processamento e transações com os dispositivos de armazenamento necessita de sistemas computacionais dimensionados e otimizados especificamente para esse tipo de atividade. Ao definir esses ambientes de computação de alto desempenho se faz necessário uma serie de testes para obtenção de dados concretos que servirão para a analise de qual método seria o mais eficiente. Dentre os métodos para otimização de consultas a banco de dados, dois componentes são os principais pontos de gargalo, a memória e o processador. No que se refere a memória, o disco rígido, responsável pelo armazenamento de grande quantidade de dados pode ter seu desempenho melhorado através de RAID (redundant array of independent disks) que consiste numa solução de baixo custo e que provê melhoria considerável do desempenho, pois utiliza disco rígido espelhados e também uso de dispositivos de memoria flash, como pen drives e SSD's. No processamento, são disponíveis duas técnicas tradicionais, o multiprocessadores e o multicomputadores o primeiro fazendo um paralelismo de instruções nos processadores e o segundo utilizando-se de um cluster ou nuvem. E nessa área ainda, uma filosofia de processamento que vem tomando peso é a que utiliza a placa de vídeo, devido ao acesso ao banco precisa serem feitos inúmeros cálculos, a GPU atuaria na busca desses dados, pois possui enorme capacidade para cálculos de ponto flutuante. Ao final do projeto espera-se apresentar e quantificar as melhorias no processo de mineração de dados astronômicos proposto pelo observatório municipal. Os indicadores de resultados serão baseados na velocidade de resposta das consultas feitas ao banco de dados verificando qual dos métodos de computação de alto desempenho tem a melhor aplicação.