

A UTILIZAÇÃO DE UM FRAMEWORK PARA CRIAÇÃO DE CLUSTERS COMPUTACIONAIS NO PROCESSAMENTO DE ALGORITMOS DE DATA MINING

Ramon Venson¹, Paulo João Martins¹, Priscyla Waleska Targino de Azevedo Simões^{1,2}

¹ Grupo de Pesquisa em Tecnologia da Informação e Comunicação na Saúde (UNACET/UNESC)

² Laboratório de Epidemiologia (PPGCS/UNESC)
ramon.venson@gmail.com, <pjm, pri>@unesc.net

Palavras-Chave: *Sistemas Distribuídos, JPPF, Clusters Computacionais, Data Mining, Benchmark.*

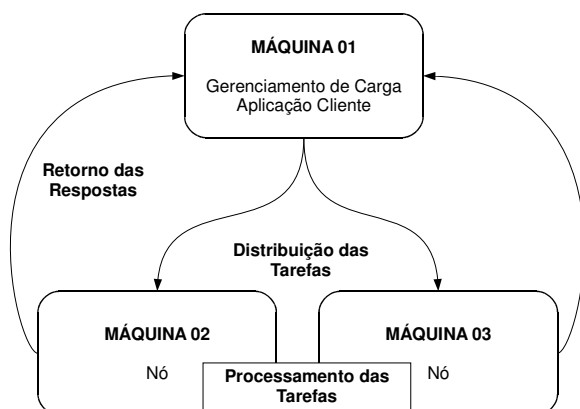
INTRODUÇÃO

Dentre as formas de ampliar o poder de processamento de um ambiente computacional, pode-se citar o agregamento de computadores. Estes, também chamados de Cluster Computacionais, caracterizam-se por utilizar máquinas dispersas fisicamente na realização de tarefas complexas e que exijam um alto tempo computacional para serem realizadas. (Dantas, 2005). Com isso, foram desenvolvidas técnicas de processamento paralelo e distribuído com a finalidade de diminuir o tempo computacional gasto, realizando a divisão do processamento entre diversas máquinas de um cluster. Nesse contexto, realizou-se um estudo e experimentos relacionados à aplicabilidade e desempenho do JPPF (COHEN, 2013), um framework na linguagem java para a construção de ambientes distribuídos, juntamente a um software de Data Mining.

METODOLOGIA

Os microcomputadores utilizados nos experimentos possuíam hardware equivalentes e os softwares foram configurados de forma a manter ambientes com um alto grau de homogeneidade. Também foram bloqueados os processos do sistema operacional que poderiam causar algum tipo de interferência ou anomalia na coleta de tempos durante a execução dos testes. Foi criado um ambiente com 3 máquinas, com funções determinadas conforme a figura 1.

Figura 01 – Funcionamento dos Ambientes



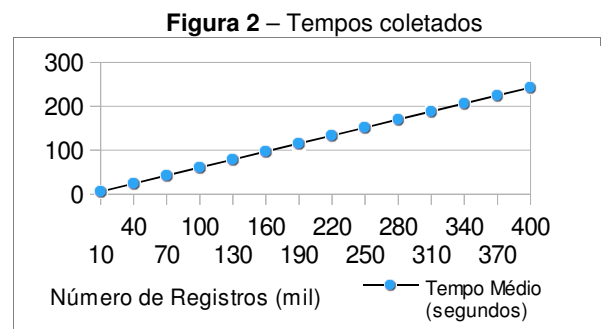
Fonte: Elaborada pelo autor.

Para os testes foram realizadas execuções de algoritmos de clusterização de dados com bases de dados de diferentes tamanhos. Cada execução teve seu tempo

coletado pela própria aplicação cliente e posteriormente analisado.

RESULTADOS E DISCUSSÃO

Através da análise efetuada da média dos tempos de processamento referentes a cada base de dados, pode-se perceber um aumento aproximadamente linear do tempo de processamento em relação ao tamanho das bases.



Fonte: Elaborado pelo autor.

A distribuição da carga de processamento entre as máquinas nós foi realizada de maneira proporcional, com cada máquina recebendo aproximadamente a mesma quantidade de tarefas.

CONCLUSÃO

A utilização da ferramenta JPPF mostrou-se viável e de fácil implementação em aplicações de grande complexidade onde existe a necessidade de um alto poder de processamento. Através do framework oferecido pelo JPPF, foi possível criar um ambiente com capacidade computacional expandida utilizando-se de infraestrutura de hardware comum, caso das máquinas utilizadas no processo.

AGRADECIMENTOS

Programa Institucional de Bolsas de Iniciação Científica (PIBIC), Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC), Financiadora de Estudos e Projetos (FINEP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e Universidade do Extremo Sul Catarinense (UNESC).

REFERÊNCIAS

COHEN, Laurent. JPPF.ORG. Disponível em: <<http://www.jppf.org>>. Acesso em: 15 ago. 2013.
DANTAS, Mário. Computação distribuída de alto desempenho. Rio de Janeiro: Axcel Books, 2005.