

Título: Análise de “traces” e “logs” de execução para prevenção de falhas de Aplicações em Grid

Descrição: Ambientes reais de grid são altamente instáveis e dinâmicos e falhas podem ocorrer em qualquer nível, seja de usuário, de sistema ou falha física. A infraestrutura de grid do projeto EELA-2 tem um repositório de execuções de aplicações que estiveram rodando desde Janeiro de 2006 até o presente momento.

O objetivo deste trabalho é utilizar técnicas de aprendizagem de máquina como as utilizadas no Weka ou outras ferramentas para gerar um classificador que seja capaz de detectar pontos de falha antes de uma tarefa ser colocada para executar. Desta forma, o classificador poderia guiar estratégias de escalonamento para uma melhor escolha de recursos a serem utilizados para executar a tarefa.

Plano de trabalho:

1. Estudo dos logs de execução das aplicações
2. Estudo do sistema Weka e outras ferramentas
3. Preparação dos logs
4. Definição dos testes e parâmetros
5. Análise de resultados
6. Escrita do relatório de trabalho

Bibliografia:

1. Diego Carvalho, Pedro Rausch Bello, Alexandre Duarte and Inês Dutra. Mining the EELA-2 infrastructure. First EELA-2 Conference, 2009, Feb, Bogotá, Colombia
2. Nitesh V. Chawla, Douglas Thain, Ryan Lichtenwalter, David A. Cieslak. 2008. Data Mining on the Grid for the Grid, 22nd IEEE International Symposium on Parallel and Distributed Processing, IPDPS 2008, Miami, Florida USA, April 14-18, <http://dx.doi.org/10.1109/IPDPS.2008.4536427>.
3. D. Cieslak, D. Thain, and N. Chawla. Short Paper: Troubleshooting Distributed Systems via Data Mining. High Performance Distributed Computing, 2006 15th IEEE International Symposium on, pages 309--312, June 19-23 2006.
4. L. Shrinivas, and J. F. Naughton. 2008. Issues in applying data mining to grid job failure detection and diagnosis. In Proceedings of the 17th international Symposium on High Performance Distributed Computing (Boston, MA, USA, June 23 - 27, 2008). HPDC '08. ACM, New York, NY, 239-240. DOI= <http://doi.acm.org/10.1145/1383422.1383464>.
5. Ian H. Witten and Eibe Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
6. D. Zeinalipour-Yazti, K. Neocleous, C. Georgiou, M. D. Dikaiakos, FailRank: Towards a Unified Grid Failure Monitoring and Ranking System Making Grids Work, Edited by Marco Danelutto, P. Fragopoulou and V. Getov, Springer Press, ISBN-13 :978-0-387-78447-2, 2008.
7. D. Zeinalipour-Yazti, K. Neocleous, C. Georgiou, and M. D. Dikaiakos. 2008. Identifying Failures in Grids through Monitoring and Ranking. In Proceedings of the 2008 Seventh IEEE international Symposium on Network Computing and Applications - Volume 00 (July 10 - 12, 2008). NCA. IEEE Computer Society, Washington, DC, 291-298. DOI= <http://dx.doi.org/10.1109/NCA.2008.10>