



Evento	Salão UFRGS 2013: SIC - XXV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2013
Local	Porto Alegre - RS
Título	Caracterização do consumo energético do Hadoop MapReduce
Autor	Flavio Alles Rodrigues
Orientador	CLAUDIO FERNANDO RESIN GEYER

O GPPD (Grupo de Processamento Paralelo e Distribuído) - grupo de pesquisa ao qual o autor do trabalho está vinculado - tem como objeto de pesquisa diversos tópicos relacionados a processamento paralelo e distribuído. Este trabalho está inserido no subgrupo *MapReduce*.

MapReduce [1] é um modelo de programação que tem por objetivo abstrair os detalhes de implementação de aplicativos paralelos e, assim, facilitar o desenvolvimento de aplicações desta natureza. Desta forma, o usuário da plataforma deve preocupar-se apenas com o desenvolvimento de sua solução, que segue o modelo *MapReduce*. Logo, a plataforma *MapReduce* encarrega-se de controlar a comunicação entre nós, distribuição de tarefas e dados, falhas, dentre outros aspectos/problemas comuns em ambientes distribuídos. A implementação *MapReduce* utilizada para o desenvolvimento deste trabalho é o *Hadoop* [2] - implementação de código aberto, desenvolvida com base no artigo que apresentou o modelo à comunidade científica [1].

O objetivo do trabalho é caracterizar o consumo de energético do modelo de programação *MapReduce* (i.e. *Hadoop*) a partir de seus parâmetros de configuração. Desta forma, o trabalho também se insere em um contexto de pesquisa em «Computação Verde» [3]. Posteriormente à caracterização, espera-se inserir uma funcionalidade de predição/simulação do consumo energético do modelo no simulador MRSO (*MapReduce over Simgrid*) desenvolvido pelo mestrando do GPPD Wagner Kolberg [4]. Para tanto, a metodologia do trabalho é a seguinte:

- Realização de testes seguindo modelo de testes 2^k [5]: sete parâmetros de configuração considerados os mais relevantes - segundo extensa pesquisa bibliográfica - foram testados para determinar aqueles que têm mais influência no consumo energético do *Hadoop*. (Realizado)
- Determinação de um modelo de consumo energético conforme [6]: um modelo de consumo energético baseado em estatísticas de uso, conforme descrito no trabalho citado, está em desenvolvimento. As estatísticas de uso foram coletadas através do monitor desenvolvido no GPPD pelo aluno Pedro Dusso [7]. (Em andamento)
- Nova rodada de testes - seguindo a metodologia de testes fatoriais completos [5], avaliando somente os parâmetros mais relevantes, para melhor entender o comportamento do consumo energético sob variadas configurações. (A realizar)
- Aplicação dos resultados ao MRSO. (A realizar)

[1] Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." *Communications of the ACM* 51.1 (2008): 107-113.

[2] White, Tom. *Hadoop: The definitive guide*. O'Reilly Media, Inc., 2012.

[3] Hooper, Andy. "Green computing." *Communications of the ACM* 51.10 (2008): 1-13.

[4] Kolberg, Wagner, et al. "MRSO-A MapReduce simulator over SimGrid." *Parallel Computing* (2013).

[5] Jain, Raj. *The art of computer systems performance analysis*. Vol. 182. Chichester: John Wiley & Sons, 1991.

[6] Economou, Dimitris, et al. "Full-system power analysis and modeling for server environments." *In Proceedings of Workshop on Modeling, Benchmarking, and Simulation*. 2006.

[7] Dusso, Pedro Martins. "A monitoring system for WattDB: an energy-proportional database cluster." 68p. Trabalho de Conclusão de Curso (Graduação em Ciência da Computação). Instituto de Informática, Universidade Federal do Rio Grande do Sul, 2012. Disponível em: <http://hdl.handle.net/10183/54146>