

## **GRADE COMPUTACIONAL COMO ALTERNATIVA PARA O PROCESSAMENTO DE ALTO DESEMPENHO EM PESQUISA TECNOLÓGICA**

K. F. M. Lucena

Coordenação de Ciências Gerenciais – CEFET-PB

Av. 1º de Maio, 720 - CEP: 58.015-430 - João Pessoa -PB

E-mail: kennedyflavio@yahoo.com.br

E. C. M. N. Machado; I. S. Lacerda; P. R. V. Lucena; C. O. Galvão

Universidade Federal de Campina Grande

Av. Aprígio Veloso, 882 - Campina Grande -PB

E-mail: erica@lsd.ufcg.edu.br

### **RESUMO**

É cada vez maior o número de aplicações computacionais, em todas as áreas do conhecimento, e que exigem continuamente maior esforço no que se refere ao desempenho computacional nos processamentos de dados e da informação. Algumas soluções adotadas como supercomputadores e os *clusters* de processadores ainda apresentam limitações de uso devido ao elevado custo dos mesmos. Uma alternativa mais racional e acessível é o processamento em grade de computadores. Consiste na formação de cooperativas de computadores em rede, localizados em qualquer parte do planeta e, portanto, além de solucionarem os problemas de processamento, otimizam o uso dessas máquinas reduzindo sua ociosidade. O CEFET-PB participa de uma grade computacional de aplicações técnico-científicas, a Cooperativa de Processamento de Alto Desempenho – COPAD, coordenada pela UFCG. A COPAD é integrada à grade OurGrid, que conta atualmente com aproximadamente 500 computadores. Uma das linhas de pesquisa do CEFET-PB, que está usando a COPAD, é a de redes hidráulicas, com aplicações a irrigação, transporte de petróleo e abastecimento de água. Nestas aplicações é freqüentemente necessário realizar simulações intensivas de cenários com objetivos de dimensionamento otimizado, análise de operações dos sistemas e reabilitação de redes de abastecimento. No CEFET-PB, tem-se conseguido ganhos de desempenho bastante relevantes nestas aplicações, que são apresentados neste trabalho.

**PALAVRAS-CHAVE:** processamento de alto desempenho, grade computacional, hidráulica.

## 1. INTRODUÇÃO

Em muitos problemas de engenharia é comum o uso de modelos computacionais de simulação e/ou de otimização com finalidades de análise, projeto e planejamento e que demandam muitas vezes um elevado custo de processamento. No caso de simulações hidráulicas pode-se fazer uso de estruturas de alto desempenho como *clusters* computacionais. No entanto, os custos para aquisição de *clusters* ainda é o principal aspecto restritivo e que pode se tornar inviável economicamente para a maioria das empresas. A utilização de grades computacionais surge como uma alternativa aos *clusters*.

As grades computacionais são uma nova plataforma para execução de aplicações paralelas que apresentam como principal atrativo em relação às alternativas tradicionais a possibilidade de alocar uma enorme quantidade de recursos a uma aplicação paralela e fazê-lo a um custo muito menor, Cirne (2002).

Segundo Foster (2004) apud Py (2005) o termo grid denota "uma proposta de infraestrutura de software e hardware para a integração de recursos computacionais, dados e pessoas geograficamente dispersas de modo a formar um ambiente colaborativo de trabalho. Para Krauter (2002) apud Py (2005) é um sistema computacional de rede que pode escalar ambientes do tamanho da internet com máquinas distribuídas através de múltiplas organizações e domínios administrativos.

De um modo geral, as grades computacionais se diferem de outras estruturas de processamento paralelo por serem mais distribuídas, diversas e complexas. Os aspectos principais que diferenciam as grades dos *clusters* computacionais são, segundo (Galvão et al., 2004): máquinas heterogêneas, alta dispersão geográfica (podem ter escala mundial), múltiplos domínios administrativos (podem congrega várias instituições), controle distribuído (tipicamente, não há uma entidade central que tenha o poder sobre o todo).

Segundo Cirne et al. (2006), a maioria das soluções atuais em grades computacionais tem aplicações voltadas para uma pequena quantidade de grandes laboratórios que trabalham juntos em um mesmo problema. Mas, a situação que se encontra na maioria dos grupos de pesquisa – compartilhada também pelas empresas concessionárias de serviços de abastecimento de água – é a de pequenos laboratórios, geralmente focados em pesquisas ou aplicações específicas e que não podem dispor de grandes estruturas computacionais.

Esse tipo de aplicação é muito comum na área de análise de redes de abastecimento de água, onde cada tarefa geralmente é definida como a simulação de um cenário específico (ou um conjunto de cenários) da rede.

Neste trabalho apresenta-se uma aplicação de grade computacional com o objetivo de avaliar a performance computacional, em relação ao tempo de processamento, para simulação do escoamento em uma rede de abastecimento de água, dando continuidade aos estudos de Galvão et al. (2006).

## 2. METODOLOGIA

### 2.1. Aplicação de modelo de simulação hidráulica

Dentre os diversos modelos computacionais para simulação hidráulica o EPANET (Rossman, 2000) é o mais difundido e amplamente utilizado por pesquisadores, engenheiros hidráulicos e companhias de abastecimento de água. É um sistema computacional que permite executar simulações estáticas e dinâmicas do comportamento hidráulico e de qualidade da água de sistemas de distribuição em pressão. São inúmeras as aplicações do EPANET, podendo-se citar aplicações de análise, dimensionamento, otimização e simulação em redes de abastecimento urbano e para irrigação, como: López-Ibáñez et al. (2005), Pedrosa Filho et al. (2005), Carrijo (2004), Lucena (2005) e Cheung (2004).

Segundo Galvão et al. (2006) o EPANET dispõe de um toolkit de programação que facilita sua interação com outros modelos, consiste em um conjunto com mais de 50 funções, integradas em uma Dynamic Link Library (DLL), que podem ser incorporadas em aplicações de 32 bits para MS® Windows®, escritas em linguagens C/C++, Delphi™ Pascal, Visual Basic™ ou outro tipo de linguagem que permita a chamada de funções Windows DLL.

Neste trabalho faz-se a aplicação do EPANET em grade computacional OurGrid na simulação de uma rede de abastecimento de água. Foram realizadas 10, 100, 500, 2500, 5000 e 10000 simulações de uma mesma rede para avaliar a performance computacional, ou seja, o ganho na execução da grade e as implicações advindas dos experimentos.

A rede hidráulica utilizada nas simulações foi um sistema adutor, parte do sistema de abastecimento da cidade de Campina Grande (Barbosa et al., 2003), com suprimento de água realizado a partir do reservatório R-9 (M1 neste trabalho) da CAGEPA (Companhia de Abastecimento). É composto por três estações de bombeamento (Estações

Elevatórias 3, 4 e 9 da CAGEPA, denominadas respectivamente, de B1, B2 e B3), três reservatórios de distribuição (Reservatórios R-13, R-14 e R-22 da CAGEPA, denominados respectivamente, neste trabalho, de R1, R2 e R3) e nós de distribuição, conforme apresentado na Figura 1, Galvão et al. (2006).

Figura 1. Rede hidráulica simulada (Galvão et al. 2006).

As simulações foram definidas para um período de 120 horas, com intervalos de tempo de 1 hora. Em um computador com boa capacidade de processamento (1.024 RAM, HD 160GB, INTEL® PENTIUM® 4 HT, CPU 3.00GHz, Sistema operacional Linux Debian Sarge) o tempo de simulação para uma execução é de aproximadamente 20 segundos. Assim, um problema que requeira 10000 simulações no EPANET de redes com características semelhantes à utilizada e com computador de desempenho equivalente, por exemplo, consome aproximadamente 55 horas para fornecer os resultados de todas as simulações, Galvão et al. (2006).

## 2.2. A grade computacional

Neste estudo foi utilizada a grade computacional denominada OurGrid (Cirne et al., 2006; <http://www.ourgrid.org>) que é um sistema de grades computacionais cooperativas com software aberto e disponível para uso, em que os laboratórios doam seus recursos computacionais para a grade quando eles estão localmente ociosos e obtêm recursos remotos da grade quando necessário (o CEFET-PB é membro da cooperativa COPAD). Atualmente o OurGrid suporta apenas aplicações do tipo **Bag-of-Tasks** (BoT), as quais são aplicações paralelas cujas tarefas são independentes, podendo ser executadas em qualquer ordem. Esse tipo de aplicação é muito comum na área de análise de redes de abastecimento de água, onde cada tarefa geralmente é definida como a simulação de um cenário específico (ou um conjunto de cenários) da rede.

O OurGrid é constituído de três componentes principais: MyGrid, Peers e User Agents. O MyGrid é o componente responsável pelo controle e execução da aplicação, contém todos os dados de entrada e coleta os resultados obtidos em cada máquina (grid machine) que executa as aplicações Galvão et al. (2006). Deve ser instalado em uma máquina com sistema operacional Linux, a qual é denominada de home machine (Figura 2).

Neste trabalho foi utilizada a versão 3.3.1 do OurGrid cujos componentes estão disponíveis para download no site oficial do sistema, com descrição detalhada dos requerimentos necessários e procedimentos para instalação e utilização do sistema no Quick Start Guide - OurGrid 3.3.1 (2006) e no OurGrid 3.3.1 Online Manual (2006).

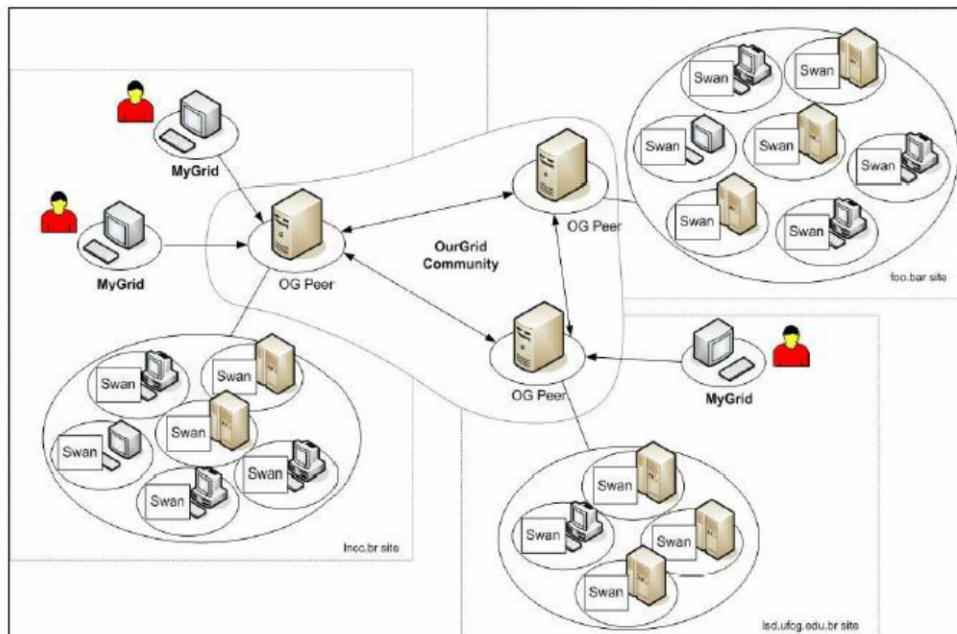


Figura 2. Arquitetura da solução OurGrid (OurGrid 3.3.1 Online Manual, 2006)

Segundo Galvão et al. (2006) a utilização de grades computacionais acrescenta preocupações para o desenvolvimento do software e uma dessas preocupações é a heterogeneidade de sistemas operacionais dos diversos recursos da grade computacional. O sistema de grades computacionais OurGrid, por exemplo, apesar de requerer que as máquinas referentes à peer machine e à home machine operem com sistema operacional Linux, enviam suas tarefas para diversas grid machines, as quais podem executá-las tanto com o sistema operacional Linux quanto com Windows.

### 3. RESULTADOS E DISCUSSÃO

Os experimentos realizados por Galvão et al. (2006) foram conduzidos utilizando a versão 3.3 do OurGrid. Quando da realização dos experimentos, a comunidade OurGrid utilizando esta versão, estava constituída por 8 peers, totalizando aproximadamente 120 máquinas na grade. Os resultados obtidos demonstraram uma redução no tempo de simulação em torno de 94%.

Neste trabalho os experimentos foram realizados com a grade OurGrid, utilizando versão 3.3.1, constituída por 12 peers, totalizando aproximadamente 225 máquinas. Os resultados obtidos são apresentados na Tabela 1, onde tem-se um resumo do desempenho obtido com as simulações na grade computacional, cada vez que a aplicação foi submetida, com: a) tempo de execução: tempo gasto, em minutos, e a redução percentual obtida ao comparar este tempo com o necessário para executar a aplicação em um único computador local; b) máquinas utilizadas: quantidade de máquinas diferentes que foram recebidas da grade ao longo da execução, e quantidade de máquinas que finalizaram com sucesso ao menos uma réplica; réplicas: quantidade de réplicas com estado *Finished*, *Failed* e *Aborted*.

Pôde-se verificar que em todas as execuções houve melhora considerável na performance de processamento ao utilizar a grade computacional para realizar as simulações, com uma de redução no tempo de processamento variando em torno de 62,6 % a 96,5 %. O menor tempo de execução registrado foi de 1,2 minuto, correspondente ao menor número de simulações (10), enquanto que o maior tempo de execução foi obtido com 10000 simulações, 167 minutos. O melhor desempenho foi observado para 5000 simulações em virtude do número de réplicas (*Failed* e *Aborted*).

Da mesma forma que foi verificado por Galvão et al. (2006) a relação do número de máquinas recebidas com o tempo de execução não segue uma relação definida. Uma importante constatação é que o desempenho da grade ocorreu satisfatoriamente tanto para um menor como para um maior número de simulações. Isso significa que a utilização da grade pode ser justificada mesmo em problemas de menor porte. Entretanto, o melhor desempenho se obtém com um maior número de simulações. O aumento do problema não implicou em um aumento proporcional no tempo de execução da grade.

Tabela 1. Resumo dos experimentos realizados.

<i>Jobs*</i>	Tempo de execução [minutos]			Máquinas utilizadas		Réplicas		
	Local	Grid	Redução	Total	Com sucesso	<i>Finished</i>	<i>Failed</i>	<i>Aborted</i>
10	3,3	1,2	62,6 %	56	18	10	7	68
100	33	5	84,8 %	104	79	100	41	136
500	165	10	93,9 %	110	85	500	78	97
2500	825	48	94,2 %	92	61	2500	90	65
5000	1652	57	96,5 %	109	85	5000	82	81
10000	3300	167	94,9 %	136	114	10000	187	63

\* arquivo de texto gerado para executar as aplicações na grade, com a descrição da aplicação, denominado de arquivo jdf – job description file.

#### 4. CONCLUSÕES

A grade computacional OurGrid potencializa um considerável ganho de desempenho em termos de economia de tempo de processamento, em aplicações integradas ao EPANET, que exigem elevado número de simulações. As aplicações das grades computacionais constituem excelente meio de minimizar deficiências ou mesmo melhorar a infraestrutura computacional das instituições de ensino e pesquisa.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

- BARBOSA, J.; SANTANA, C.; GALVÃO, C. (2003). Descrição de parte do sistema adutor da cidade de Campina Grande, Paraíba, Brasil. UFCG.
- CARRIJO, I. (2004). Extração de regras operacionais ótimas de sistemas de distribuição de água através de algoritmos genéticos multiobjetivo e aprendizado de máquina. Tese (Doutorado em Hidráulica e Saneamento) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- CHEUNG, P. (2004). Análise de reabilitação de redes de distribuição de água para abastecimento via algoritmo genético multi-objetivo. Tese (Doutorado em Hidráulica e Saneamento) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos.
- CIRNE, W. (2002). Grids Computacionais: Arquiteturas, Tecnologias e Aplicações (Computational Grids: Architectures, Technologies and Applications). Anais do Terceiro Workshop em Sistemas Computacionais de Alto Desempenho (Proceedings of the Third Workshop on High Performance Computing Systems), October 2002.
- CIRNE, W.; BRASILEIRO, F.; ANDRADE, N.; COSTA, L.; ANDRADE, A.; NOVAES, R.; MOWBRAY, M. (2006). Labs or the World, Unite!!!. Accepted for publication in Journal of Grid Computing.
- GALVÃO, C. O.; BARBACENA, M. M.; MACHADO, E. C. M.; LACERDA, I. S. (2006). Simulações intensivas de redes de abastecimento de água utilizando o EPANET em grade computacional. In: VIII Seminário de Recursos Hídricos do Nordeste, Gravatá.
- GALVÃO, C.; CIRNE FILHO, W.; BRASILEIRO, F.; BRASILEIRO, E., ARAÚJO, E. (2004). Computação em grade aplicada à análise de sistemas de abastecimento de água. In: Seminário Hispano-Brasileiro sobre Sistemas de Abastecimento Urbano de Água - IV SEREA, João Pessoa.
- LÓPEZ-IBÁÑEZ, M.; PRASAD, D.; PAECHTER, B. (2005). Multi-objective pump scheduling optimisation using evolutionary strategies. In IEEE Congress on Evolutionary Computation, volume 1, p. 435 – 442, Edinburgh, UK.
- LÓPEZ-IBÁÑEZ, M. (2006) EPANET Toolkit. Disponível em <http://sbe.napier.ac.uk/~manuel/epanetlinux>.
- LUCENA, K. (2005). Uso do EPANET no dimensionamento de um sistema de irrigação. In Seminario Iberoamericano sobre Planificación, Proyecto y Operación de Sistemas de Abastecimiento de Agua. Valência (España).
- OURGRID 3.3 ONLINE MANUAL (2006). Disponível em [http://www.ourgrid.org/twikipublic/bin/view/OG/ManualOnline3\\_3](http://www.ourgrid.org/twikipublic/bin/view/OG/ManualOnline3_3).
- PEDROSA FILHO, L.; GOMES, H.; REY, P. (2005). Otimização da operação do sistema adutor da cidade de João Pessoa – Brasil. In Seminario Iberoamericano sobre Planificación, Proyecto y Operación de Sistemas de Abastecimiento de Agua. Valência (España).
- PY, M. X. (2005). Grid computacional. [www.inf.ufg.br/~jkv/ap-grid.pdf](http://www.inf.ufg.br/~jkv/ap-grid.pdf). Acesso em 14/10/06.
- ROSSMAN, L. A. (2000). EPANET 2 – Users Manual. U.S. Environmental Protection Agency –EPA. Cincinnati, Ohio. (Disponível em <http://www.epa.gov/nrmrl/wswrd/epanet.html>).

