

Centro Federal de Educação Tecnológica de Minas Gerais

Departamento de Computação

Algoritmos paralelos de ordenação em Hadoop

Mariane Raquel Silva Gonçalves

Orientadora: Cristina Duarte Murta

Belo Horizonte
13 de março de 2012

Sumário

1 Proposta do Projeto.....	2
1.1 Tema do trabalho	2
1.2 Relevância	2
1.3 Objetivos	2
1.4 Resultados esperados	3
1.5 Metodologia	3
1.6 Infraestrutura necessária	3
1.7 Cronograma de trabalho	4
Referências.....	5

1 Proposta do Projeto

1.1 Tema do trabalho

A ordenação

A computação paralela ...

1.2 Relevância

Uso crescente de computação paralela para sistemas computacionais.

1.3 Objetivos

O projeto propõe a

(i) Estudar a programação paralela, seus algoritmos e suas possibilidades de implementação em ambiente paralelo multicore; (ii) Implementar e avaliar o desempenho de um algoritmo de ordenação paralela; (iii) Estudar e implementar a solução no modelo MapReduce, com o software Hadoop; (iv) Comparar os resultados obtidos com os algoritmos de ordenação e analisar seu desempenho com relação à quantidade de dados a serem ordenados, variabilidade dos dados de entrada e número máquinas utilizadas.

1.4 Resultados esperados

Com a realização do trabalho, busca-se ampliar e consolidar conhecimentos adquiridos na área de computação paralela, assim como a capacidade de análise e desenvolvimento de algoritmos paralelos. Os resultados esperados incluem entendimento do modelo MapReduce, implementação de algoritmos de ordenação paralela em ambiente Hadoop, assim como testes para comparação de desempenho entre dois ou algoritmos.

1.5 Metodologia

vamos estudar a importancia da computacao paralela. da ordenacao. Vamos estudar o map reduce. do hadoop. implementar. testar. melhorar se possível. refazer os testes comparando implementacoes e analisar os resultados.

1.6 Infraestrutura necessária

A infra estrutura necessária ao desenvolvimento do projeto será fornecida pelo Laboratório de Redes e Sistemas (LABORES) do Departamento de Computação (DECOM). Esse laboratório possui um cluster formado por cinco máquinas Dell Optiplex 380, que serão utilizadas na realização dos testes dos algoritmos. Os algoritmos serão desenvolvidos em linguagem Java, de acordo com o modelo MapReduce, no ambiente Hadoop.

O cluster a ser utilizado apresenta as seguintes características:

- 5 nodos
- Processador Intel Core 2 Duo de 3.0 GHz
- Disco rígido SATA de 500 GB 7200 RPM
- Memória RAM de 4 GB
- Placa de rede Gigabit Ethernet

Fevereiro	Definição do tema de trabalho
Março	Pesquisa bibliográfica e escrita da proposta de projeto
Abril	Pesquisa bibliográfica e início da escrita do projeto
Maio	Aplicação da metodologia proposta no desenvolvimento de algoritmo e realização de testes
Junho	Finalização da escrita do projeto e entrega
Julho	Desenvolvimento e testes de algoritmos paralelos
Agosto	Desenvolvimento e testes dos algoritmos desenvolvidos
Setembro	Análise dos resultados obtidos até o momento, em busca de pontos de melhorias no projeto
Outubro	Teste final dos algoritmos, análise e escrita dos resultados
Novembro	Finalização da escrita do projeto final, entrega e apresentação

Tabela 1: Cronograma proposto para o projeto

- Sistema operacional Linux Ubunbu 10.04 32 bits (kernel 2.6.**XX**)
- Sun Java JDK 1.6.0 19.0-b09
- Hadoop 0.20.2

1.7 Cronograma de trabalho

[Kale e Solomonik 2010]

[Manferdelli, Govindaraju e Crall 2008]

[Dean e Ghemawat 2008]

[Asanovic et al. 2009]‘

Referências

- [Asanovic et al. 2009]ASANOVIC, K. et al. A view of the parallel computing landscape. *Commun. ACM*, ACM, New York, NY, USA, v. 52, n. 10, p. 56–67, out. 2009. ISSN 0001-0782.
- [Dean e Ghemawat 2008]DEAN, J.; GHEMAWAT, S. Mapreduce: simplified data processing on large clusters. *Commun. ACM*, ACM, New York, NY, USA, v. 51, n. 1, p. 107–113, jan. 2008. ISSN 0001-0782.
- [Kale e Solomonik 2010]KALE, V.; SOLOMONIK, E. Parallel sorting pattern. In: *Proceedings of the 2010 Workshop on Parallel Programming Patterns*. New York, NY, USA: ACM, 2010. (ParaPLoP '10), p. 10:1–10:12. ISBN 978-1-4503-0127-5.
- [Manferdelli, Govindaraju e Crall 2008]MANFERDELLI, J. L.; GOVINDARAJU, N. K.; CRALL, C. Challenges and opportunities in Many-Core computing. *Proceedings of the IEEE*, v. 96, n. 5, p. 808–815, maio 2008. ISSN 0018-9219.