Introdução ao NAS Parallel Benchmarks
Performance Relativa de Kernels Sequenciais, em ambiente de Memória Partilhada e ambiente de Memória Distribuída

Filipe Oliveira
Departamento de Informática
Universidade do Minho
Email: a57816@alunos.uminho.pt

***Abstract*—Neste estudo, analisamos a performance de kernels**

1. Introduction

This demo file is intended to serve as a "starter file" for IEEE Computer Society conference papers produced under LaTeX using IEEEtran.cls version 1.8b and later. I wish you the best of success.

Filipe Oliveira
1 Março, 2016

2. Contextualização das Benchmarks

As "NAS Parallel Benchmarks" [1] englobam 5 kernels (EP, MG, CG, FT, IS) e 3 aplicações que simulam dinâmica de fluídos (LU,SP,BT). No nosso caso de estudo, temos por interesse os 5 kernels, dado que centraremos o nosso estudo da performance relativa via alterações no paradigma de memória e forma de comunicação entre nodos, assim como a própria ferramenta de compilação e respectivas flags. Assim sendo, temos então 5 opções a analisar: EP, MG, CG, FT, IS. Resta-nos portanto analisar primeiramente quais as principais propriedades de cada um antes de qualquer avanço no trabalho.

- **EP**, tal como o próprio nome indica (Embarrassingly Parallel), que por calcular números aleatórios é um kernel implicitamente embaracosamente paralelo. Tem como propósito estabelecer o Peak Performance em "FP Operations" de um sistema de computação em teste. É então espectável obtermos os melhores resultados de performance neste kernel e, por esse mesmo motivo, este será um dos kernels com grande relevância para o nosso caso de estudo.
- **MG**, cujo kernel implementa um método numérico multigrid simplificado, numa sequência de malhas de diferentes propriedades, implicando portanto uma elevada comunicação para a resolução do algoritmo. Tanto para as versões em memória distribuída com para a versão de memória partilhada será interessante analisar o comportamento do kernel nos ambientes de teste. Será portanto também incluído no caso de estudo.
- **CG**,que recorre ao método do Conjugado do Gradiente por forma a calcular uma aproximação ao menor dos valores próprios de uma matriz esparsa de grandes dimensões. Dada o tipo de dados, este kernel testa computação e comunicação não estruturada, sendo portanto expectável uma fraca performance deste kernel quando em comparação com o **EP**. Será portanto também incluído no caso de estudo.
- **FT**, que calcula a Transformada de Fourier em 3 dimensões (3 transformadas de Fourier de uma dimensão), sendo o resultado a solução de uma equação diferencial parcial. Dado que a principal propriedade a ser estudada é comunicação, este kernel será portanto excluído do caso de estudo em detrimento do **MG**.
- **IS**, que realiza operações de sorting em inteiros. Este kernel testa tanto a capacidade de computação de um sistema em termos de operações sobre inteiros, assim como a performance de comunicação do mesmo dada a irregularidade dos acessos à memória e, quando aplicável, comunicação entre processos. Será também incluído no caso de estudo.

3. Caracterização do Hardware do ambiente de testes

Escolhidas as benchmarks, resta-me especificar os ambientes de teste nos quais pretendo realizar as benchmarks. Através da análise do hardware disponível no Search6 [1], a nossa plataforma de teste, deparamo-nos com duas realidades distintas presentes no mesmo cluster. Se por um lado temos uma grande porção dos nodos de computação com configurações de hardware relativamente homogéneas (28 dos 54 nodos disponíveis apresentam todos as mesmas características de comunicação, armazenamento local, memória RAM disponível e mesma família de processadores - Ivy Bridge), por outro lado temos os restantes 26 nodos com características relativamente distintas entre nodos (diferentes famílias de processadores, diferentes características

1. Services and Advanced Research Computing with HTC/HPC clusters

de comunicação entre nodos disponíveis, memória RAM disponível em diferente número).

Considerando a abrangência do número de nodos o ponto essencial para a escolha do tipo de nodos a teste, decidi incluir no mesmo os nodos do tipo 662, 652, 641, e 431(apenas os nodos com 48GB de memória RAM disponíveis), que passarei de seguida a caracterizar.

Denote que ao realizarmos os testes de performance nos nodos acima enumerados estamos a abrangir 33 dos 54 nodos disponíveis, preservando características entre eles fundamentais para a possibilidade de comparação como por exemplo o suporte da rede Myrinet 10Gbps, e englobando como requerido mais do que uma classe de arquitectura existe no Search6.

TABLE 1. CARACTERÍSTICAS DE HARDWARE DO NODO 662

Sistema	compute-662
# CPUs	2
CPU	Intel® Xeon® E5-2695 v2
Arquitectura de Processador	Ivy Bridge
# Cores por CPU	12
# Threads por CPU	24
Freq. Clock	2.4 GHz
Cache L1	384KB (32KB por Core)
Cache L2	3072KB (256KB por Core)
Cache L3	30720KB (partilhada)
Ext. Inst. Set	SSE4.2, AVX
#Memory Channels	4
Memória Ram Disponível	64GB
Peak Memory BW Fab. CPU	59.7 GB/s
Rede Suportada	Gigabit Ethernet, Myrinet 10Gbps

TABLE 2. CARACTERÍSTICAS DE HARDWARE DO NODO 652

Sistema	compute-652
# CPUs	2
CPU	Intel® Xeon® E5-2670 v2
Arquitectura de Processador	Ivy Bridge
# Cores por CPU	10
# Threads por CPU	20
Freq. Clock	2.5 GHz
Cache L1	320KB (32KB por Core)
Cache L2	2560KB (256KB por Core)
Cache L3	25600KB (partilhada)
Ext. Inst. Set	SSE4.2, AVX
#Memory Channels	4
Memória Ram Disponível	64GB
Peak Memory BW Fab. CPU	59.7 GB/s
Rede Suportada	Gigabit Ethernet, Myrinet 10Gbps

TABLE 3. CARACTERÍSTICAS DE HARDWARE DO NODO 641

Sistema	compute-641
# CPUs	2
CPU	Intel® Xeon® E5-2650 v2
Arquitectura de Processador	Ivy Bridge
# Cores por CPU	8
# Threads por CPU	16
Freq. Clock	2.6 GHz
Cache L1	256KB (32KB por Core)
Cache L2	2048KB (256KB por Core)
Cache L3	20480KB (partilhada)
Ext. Inst. Set	SSE4.2, AVX
#Memory Channels	4
Memória Ram Disponível	64GB
Peak Memory BW Fab. CPU	59.7 GB/s
Rede Suportada	Gigabit Ethernet, Myrinet 10Gbps

TABLE 4. CARACTERÍSTICAS DE HARDWARE DO NODO 431

Sistema	compute-431
# CPUs	2
CPU	Intel® Xeon® X5650
Arquitectura de Processador	Nehalem
# Cores por CPU	6
# Threads por CPU	12
Freq. Clock	2.66 GHz
Cache L1	192KB (32KB por Core)
Cache L2	1536KB (256KB por Core)
Cache L3	12288KB (partilhada)
Ext. Inst. Set	SSE4.2
#Memory Channels	3
Memória Ram Disponível	48GB
Peak Memory BW Fab. CPU	32 GB/s
Rede Suportada	Gigabit Ethernet, Myrinet 10Gbps

Da caracterização de hardware acima realizada podemos retirar já as diferentes possibilidades relativamente ao número de threads a testar em ambiente de memória partilhada, assim como propriedades adequadas aos testes em ambiente de memória distribuída. Para além do anteriormente enumerado obtive dados de extrema importância relativa à próxima decisão do nosso caso de estudo – a escolha das classes de dados a incluir no nosso caso de estudo. Será com base nas propriedades relativas à Memória RAM disponível, tamanho e forma de distribuição dos vários níveis de cache, assim como a Peak Memory Bandwith teórica de Memória do CPU, que centrarei a minha decisão sobre quais as classes de dados relevantes.

3.1. Inclusão de diferentes dimensões (classes) de dados

Será importante para a relevância dos testes de perfomance incluir datasets de diferentes dimensões. Ora, analisando as propriedades dos processadores presentes nos nossos nodos de computação devo portante seleccionar um dataset capaz de ser compreendido na Cache L1 (menor ou igual a 32KB), um dataset capaz de ser compreendido na Cache L2 (menor ou igual a 256KB), um dataset capaz de ser compreendido na Cache L3 (menor que 12MB) e

um dataset capaz de ser compreendido na Memória Ram disponível nos nodos de computação (menor que 48GB).

Analisadas as dimensões dos problemas para as diferentes Classes de dados e Benchmarks a incluir no nosso caso de estudo, verificamos que em nenhum dos casos os datasets são possíveis de ser compreendidos totalmente na cache L1. Assim, e focando o nosso processo de seleção com base nas benchmarks IS e MG podemos concluir que as classes de dados que melhor se associam ao nosso problema são portanto a classe S (totalmente contida na cache L2 ou L3), e as classes A, B, e C por serem consideradas classes standard das benchmarks e de fácil comparação de resultados com a comunidade do HPC.

Denote ainda que para o caso da benchmark CG a classe de dados C apresenta um tamanho de dados superior ao possível de estar contido na memória principal obrigando teoricamente a um decréscimo de perfomance derivada do aumento de IO.

No anexo X presente na página X poderá encontrar informação mais detalhada relativa à forma de cálculo do tamanho dos datasets das diferentes classes de dados.

TABLE 5. DIMENSÃO DO DATASET PARA AS DIFERENTES CLASSES E BENCHMARKS

Bench.	data type	S	A	B	C
EP	double	128 MB	2 GB	8 GB	32 GB
MG	double	256 KB	128 MB	128 MB	1024 MB
CG	double	15MB	1,46 GB	41,91 GB	167,64 GB
IS	integer	256 KB	32 MB	128 MB	512 MB

4. Caracterização do Software do ambiente de testes

O cluster Search6 é baseado no Rocks 6.1 cluster management distribution. Dado que um dos principais propósitos deste caso de estudo acenta na investigação da influência de diferentes ferramentas de compilação e diferentes configurações de ferramentas de comunicação, assim como dos diferentes paradigmas de memória, na perfomance global dos kernels, o próximo passo natural passa por identificar os módulos disponíveis no nosso ambiente de clustering que cumprem o objectivo anteriormente enumerado.

Relativamente aos compiladores de interesse para o caso de estudo temos portanto o **GCC compiler suite** e o **Intel compilers suite**. O último apresenta apenas a possibilidade de seleção da versão "icc version 13.0.1 (gcc version 4.4.6 compatibility)". Ora, pela própria informação presente na versão do compilador da Intel deveremos incluir no nosso leque de testes a versão do compilador da gnu 4.4.6. Irei também incluir a versão 4.9.0 do compilador da Gnu por ser considera a versão default deste mesmo compilador no nosso ambiente de clustering.

Relativamente às MPI stacks, nomeadamente à versão do OpenMPI a ser incluída, será em ambos os casos (Intel e GNU) a versão mais recente disponível para ambos, nomeadamente a versão 1.8.2 dos módulos com via de comunicação Gigabit Ethernet e Myrinet 10Gbps.

Relativamente às flags de compilação serão testas as respectivas optimizações disponibilizadas pelas flags de compilação -O2 e -O3, e sem qualquer tipo de otimização por flags de compilação.

Teremos portanto as seguintes combinações de software distintas no nosso caso de estudo:

- Compiladores distintos:
  - Kernels compilados com compilar da GNU gcc versão 4.4.6
  - Kernels compilados com compilar da GNU gcc versão 4.9.0
  - Kernels compilados com compilar da Intel icc versão 13.0.1
- Versão do OpenMPI:
  - versão 1.8.2 via comunicação Gigabit Ethernet para compilador da GNU
  - versão 1.8.2 via comunicação Gigabit Ethernet para compilador da Intel
  - versão 1.8.2 via comunicação Myrinet 10Gbps para compilador da GNU
  - versão 1.8.2 via comunicação Myrinet 10Gbps para compilador da Intel

5. Conclusion

The conclusion goes here.

Acknowledgments

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.