

CCF 252 - Organização de Computadores I

Trabalho Prático 01 - Benchmarks

Mateus Pinto da Silva¹, Leandro Lázaro Araújo Vieira¹, José Augusto M. Nacif¹

¹Instituto de Ciências Exatas e Tecnológicas,
Universidade Federal de Viçosa, Florestal, MG, Brasil

1. Descrição

Neste trabalho, buscamos exercitar a prática da medição e da análise de desempenho, utilizando benchmarks. O trabalho é composto de quatro partes gerais: (1) especificação dos computadores; (2) medição do desempenho de cada computador utilizando benchmarks; (3) análise de desempenho dos computadores com base nos dados obtidos; e (4) solução de problemas hipotéticos com uso das informações obtidas anteriormente. Além disso, é necessário a escrita de um tópico sobre as conclusões gerais do trabalho. Para isto, são fornecidos: (a) um computador referência, usado como base para a comparação com os demais computadores; (b) um modelo simplificado de computador e (c) um modelo simplificado de carga de trabalho para serem usados na solução dos problemas hipotéticos; (d) a indicação de benchmarks para a medição de desempenho de diferentes características computador.

A versão final deste trabalho deverá obedecer ao formato especificado na Seção 6. Os critérios para avaliação serão: a qualidade do texto (capacidade de síntese, com ênfase nos pontos mais importantes, figuras, numeração de figuras e bibliografia), a estrutura do texto (o padrão foi seguido? O texto está bem organizado?) e principalmente o mérito técnico (qualidade das soluções, da análise de desempenho, das justificativas e das conclusões).

2. Escolha e Especificação de Computadores

Na primeira etapa do trabalho, os alunos devem escolher 2 (dois) computadores diferentes a que tenham acesso (dê preferência para o computador que cada componente do grupo possua em casa). Utilizando o benchmark indicado na Seção 7, o aluno deverá especificar cada um desses computadores preenchendo os requisitos abaixo. Os computadores escolhidos serão chamados de CompA e CompB, além do CompRef (computador referência). Deverão ser especificadas as seguintes características de cada um dos computadores:

1. Processador: marca, modelo e frequência (GHz)
2. Sistema Operacional: nome
3. Chipset: marca e modelo
4. Placa-mãe: marca e modelo
5. Memória Primária (RAM): tamanho (MB), frequência e tipo (ex: SDRAM etc.)
6. Memória Cache L1 (dados e instruções): tamanho (KB), associatividade (n-way) e tamanho do bloco em bytes (line size)
7. Memória Cache L2: tamanho (KB), associatividade (n-way) e tamanho do bloco em bytes (line size)

3. Medição de Desempenho

Para cada um dos computadores escolhidos e especificados, o desempenho de cada parte do computador deverá ser determinado, por meio do benchmark e métricas especificadas abaixo:

1. Unidade Lógica e Aritmética (ALU): MIPS
2. Unidade de ponto flutuante (FPU): MFLOPS
3. Cache L1: latência (latency)
4. Cache L2: latência
5. Memória Primária: latência e largura de banda (bandwidth)

4. Análise de Desempenho Experimental

Nesta terceira parte, com base nos dados medidos, deverão ser feitos gráficos de desempenho de cada uma das métricas medidas, ou seja, a apresentação dos resultados. Em seguida, deverá ser feita uma análise do desempenho de cada uma das métricas medidas, com base nas questões indicadas. Note que os gráficos e questões dizem respeito às informações geradas pelo benchmark em relação aos computadores escolhidos (CompA, CompB e CompRef).

- Fazer 1 gráfico de barras: Integer Score e Floating Point Score para os 3 computadores. Questão: Justifique por que o processador mais rápido foi o mais rápido para cada caso (ALU (Integer Score) e FPU (Floating Point Score)).
- Fazer 1 gráfico de barras: latência x tipo de memória (para cache L1, L2, memória principal) para os 3 computadores. Obs: este gráfico terá 9 barras ao todo. Questão: Em relação a cada tipo de memória, qual computador obteve o melhor desempenho?

4.1. Computador Referência (CompRef)

O computador referência possui 1000 pontos no modo de único processador (single core) no Geekbench.

4.1.1. Especificação do computador de referência

Processador: Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz

Sistema Operacional: Microsoft Windows 10 (64-bit)

Chipset: Acer Core7H

Placa-mãe: Acer Nitro AN515-51 Acer Nitro 5

Memória Primária (RAM): 2 x 8GB 2.4GHz DDR4 SDRAM

Memória Cache de Instrução L1 : 4x 32kB, 8-Way, 64bytes Line Size, 2 Thread(s)

Memória Cache de Dados L1 : 4x 32kB, 8-Way, 64bytes Line Size, 2 Thread(s)

Memória Cache L2: 4x 256kB, ECC, 4-Way, 64bytes Line Size, 2 Thread(s)

Memória Cache L3: 6MB, ECC, 12-Way, Fully Inclusive, 64bytes Line Size, 16 Thread(s)

4.1.2. Métricas de Desempenho do computador de referência

CPU Aritmético:

MIPS: 102000

MFLOPS: 63230

Cache L1: Latência – 5 ciclos = $(5 \times 1/\text{Frequência}) = 5/2,8 \text{ GHz} = 1,78 \text{ ns}$

Cache L2: Latência – 11,5 ciclos = $(11,5 \times 1/\text{Frequência}) = 11,5/2,8 \text{ GHz} = 4,1 \text{ ns}$

Memória Primária:

Latência: 40.8 ns

Largura de Banda:

Write float: 21,92 GB/s

Write int: 21,65 GB/s

5. Análise de Desempenho Analítica

Nesta parte, os alunos deverão utilizar as informações coletadas nas partes anteriores, juntamente com os modelos de arquitetura e de carga de trabalho para solucionar, analiticamente, as situações-problema hipotéticas. Abaixo, apresentamos os modelos de carga de trabalho e arquitetura.

5.1. Modelo de Carga de Trabalho

Um programa é composto por instruções de inteiros, instruções de ponto flutuante e número de acessos à memória. O número de acessos à memória é dividido em porcentagem de acessos à memória cache L1, cache L2, e memória primária (RAM).

5.2. Modelo de Carga de Trabalho

No nosso modelo de arquitetura, um computador é composto de unidade lógica aritmética (ULA), unidade de ponto flutuante (UPF) e três memórias independentes: cache L1, cache L2 e memória principal. A ULA é responsável pelo processamento das instruções de inteiros, por meio da equação 1, na qual o Tempo é dado em segundos

$$\text{Tempo} = \text{Número de Instruções} / \text{MIPS}$$

A UPF é responsável pelo processamento das instruções de ponto flutuante, por meio da equação 2, na qual o Tempo é dado em segundos.

$$\text{Tempo} = \text{Número de Instruções de Ponto Flutuante} / \text{MFLOPS}$$

Por último, o tempo gasto para realizar um determinado número de acessos a qualquer uma das memórias é dado pela equação 3. A latência ou tempo de acesso (em segundos) é o tempo gasto para cada acesso a memória.

$$\text{Tempo} = \text{Número de Acessos} \times \text{Latência}$$

5.3. Modelo de Carga de Trabalho

Suponha um programa X com as seguintes características:

- Número de Instruções de Inteiros: 10000 MI (milhões de instruções)
- Número de Instruções de Ponto Flutuante: 5000 MI (milhões de instruções)
- Número de Acessos à Cache L1: 1000000 acessos
- Número de Acessos à memória primária: 10000 acessos

Pergunta-se: quando executado no CompRef, quantos segundos o programa X gastará?

Solução:

$$TULA = 10000 \text{ MI} / 102000 \text{ MIPS} = 0,09 \text{ s (instruções de inteiros)}$$

$$TUPF = 5000 \text{ MI} / 63230 \text{ MFLOPS} = 0,07 \text{ s (instruções de ponto flutuante)}$$

$$TCL1 = 1000000 \times 1,78 \text{ ns} = 178000 \text{ ns} = 0,00178 \text{ s (acessos a cache L1)}$$

$$TRAM = 10000 \times 40,8 \text{ ns} = 408000 \text{ ns} = 0,00408 \text{ s (acessos à memória principal)}$$

$$\text{TempoTotal} = TULA + TUPF + TCL1 + TDSC$$

$$\text{TempoTotal} = 0,09\text{s} + 0,07\text{s} + 0,00178\text{s} + 0,00408\text{s}$$

$$\text{TempoTotal} = 0,16\text{s}$$

Portanto, o programa gastaria 0,16 segundos para ser executada no Computador Referência.

5.4. Problemas Hipotéticos

Nesta seção, apresentamos os problemas hipotéticos cuja solução deverá ser apresentada. Para isto, suponha o programa BABE com as seguintes características:

- Número de Instruções: 2×10^{12}
 - 20% de inteiro
 - 80% de ponto flutuante
- Número de Acessos a Memória: 1×10^7
 - 72% à cache L1
 - 15% à cache L2
 - 13% à memória primária

Considerando o programa BABE e os computadores analisados, o grupo deve resolver analiticamente os problemas abaixo:

1. Calcular o tempo gasto para executar o programa BABE em cada um dos três computadores analisados.

2. Calcular o ganho de desempenho (aceleração ou speedup) de cada computador em relação ao computador referência e gerar o gráfico: Computador x Tempo de Resposta.
3. Escolher o computador (BEST) que executou o programa em menor tempo e utilizá-lo nas próximas questões.
4. Calcular o tempo de resposta do programa BABE executado no computador BEST sem a memória cache L2 (os acessos a cache L2 passaram para a RAM). Quanto pior ficou o desempenho do computador BEST?
5. Calcular o tempo de resposta do programa BABE executado no computador BEST somente com a memória principal (todos os acessos irão para a RAM). Quanto pior ficou o desempenho do computador BEST?
6. Calcular o tempo de resposta do programa BABE executado no computador BEST somente com a memória cache L2 (todos os acessos irão para a cache L2). Quanto pior/melhor ficou o desempenho do computador BEST?
7. Considerando a questão anterior, calcule o preço da cache L2 se ela tivesse o tamanho da memória principal (cada 1M de memória cache L2 custa em torno de R\$ 50,00). Compare se o ganho no desempenho (tempo) vale o custo adicional (memória)
8. Calcular o tempo de resposta do programa BABE, caso o processador possuísse 2, 4, 8 e 16 ULAs e UPFs. Gerar o gráfico de ganho de desempenho (aceleração ou speedup) em relação ao computador BEST
9. Se melhorarmos o tempo de execução das instruções de ponto-flutuante em 50% e reduzirmos a latência de acesso à cache L1 em 30% em cada computador, qual seria o ganho geral no tempo de execução do programa BABE em todos os computadores considerados (CompA, B e Ref)?
10. Gerar o novo gráfico: Computador x Tempo de Resposta, considerando os novos valores obtidos no problema (i).
11. Após as otimizações, em qual dos computadores escolhidos (CompA, B ou Ref), este programa obteria o melhor desempenho?

6. Conclusão

Começar por exemplo, com: “Considerando as medidas e análises realizadas neste trabalho, concluímos que ...” Colocar resumo das conclusões tiradas. Neste parágrafo inicial.

Depois, cada conclusão deve ser expandida e explicada com mais detalhes. A dica para uma boa conclusão é: reler o trabalho feito e concluir ressaltando os pontos que foram observados durante a execução do trabalho (esses pontos não foram necessariamente explicitados). Para criar a sua conclusão, escreva um texto tentando responder as questões abaixo:

1. No geral, qual computador apresentou o melhor desempenho geral? Sugerindo alguns tipos de aplicações que obteriam um maior desempenho neste computador do que nos outros (ex: jogos, processamento gráfico, compilador, compactador etc).
2. Quais as vantagens e desvantagens encontradas em cada benchmark na medida de desempenho?
3. Quais são as melhores métricas para a análise de desempenho de um computador?
4. Qual o resultado que mais surpreendeu o grupo? Por quê?

7. Formato de Entrega

Respeitar o padrão “SBC”. Escrever um texto com o mínimo de 4 e o máximo de 10 páginas de acordo com o formato. Como o espaço é limitado, os critérios para avaliação serão: a qualidade do texto (a capacidade de síntese, com ênfase nos pontos mais importantes, figuras e bibliografia), a estrutura do texto (Você seguiu o padrão? Respeitou o número de páginas? Seu texto está bem organizado?) e principalmente o mérito técnico (qualidade de análise de desempenho, das justificativas e da conclusão)

8. Disponibilidade dos Benchmarks

Os benchmarks utilizados nos testes do computador referência foram Geekbench, CPU-WORLD e SiSoft Sandra.

9. Links

- CPUWorld - Informações sobre o CPU
- GeekBench - Benchmark a ser utilizado para descobrir o score
- SiSoft Sandra - Benchmark a ser utilizado para descobrir o MIPS, MFLOPS e latência da cache
- SiSoft Sandra - Download direto para Windows