



INSTITUTO NACIONAL DE PESQUISA ESPACIAL – INPE

PROVA DISCURSIVA

TG13

AMBIENTE DE SUPERCOMPUTAÇÃO E HPC (HIGH-PERFORMANCE COMPUTING)



SUA PROVA

- Além deste caderno contendo **5 (cinco)** questões discursivas **com as respectivas folhas de rascunho**, você receberá do fiscal de prova as folhas de textos definitivos;



TEMPO

- Você dispõe de **4 (quatro) horas** para a realização da prova;
- **2 (duas) horas** após o início da prova, é possível retirar-se da sala, sem levar o caderno de questões;
- A partir dos **30 (trinta) minutos** anteriores ao término da prova é possível retirar-se da sala **levando o caderno de questões**.



NÃO SERÁ PERMITIDO

- Qualquer tipo de comunicação entre os candidatos durante a aplicação da prova;
- Anotar informações relativas às respostas em qualquer outro meio que não seja no caderno de questões e nas folhas de textos definitivos;
- Levantar da cadeira sem autorização do fiscal de sala;
- Usar o sanitário ao término da prova, após deixar a sala.



INFORMAÇÕES GERAIS

- Verifique se seu caderno de questões está completo, sem repetição de questões ou falhas. Caso contrário, **notifique imediatamente o fiscal da sala**, para que sejam tomadas as devidas providências;
- Confira seus dados pessoais, especialmente nome, número de inscrição e documento de identidade e leia atentamente as instruções para preencher as folhas de textos definitivos;
- Para o preenchimento das folhas de textos definitivos, use somente caneta esferográfica, fabricada em material transparente, com tinta preta ou azul;
- Assine seu nome apenas no(s) espaço(s) reservado(s) no cartão de respostas;
- Caso você tenha recebido caderno de cargo **diferente** do impresso em suas folhas de textos definitivos, o fiscal deve ser **obrigatoriamente** informado para o devido registro na ata da sala;
- O preenchimento das folhas de textos definitivos é de sua responsabilidade e **não será permitida a troca de folha de texto definitivo em caso de erro cometido pelo candidato**;
- Para fins de avaliação, serão levadas em consideração apenas os textos das folhas de textos definitivos;
- A FGV coletará as impressões digitais dos candidatos na lista de presença;
- Os candidatos serão submetidos ao sistema de detecção de metais quando do ingresso e da saída de sanitários durante a realização das provas.
- **Boa prova!**

Questão 1

Em arquitetura de computadores, um assunto importante é a arquitetura das memórias. Existe uma forte relação entre as tecnologias e níveis das memórias utilizadas em um computador e a velocidade do processamento. Uma forma de melhorar essa velocidade, uma vez que as memórias podem ter tecnologias diferentes e velocidades menores que o processador, é criar vários níveis de memória com tecnologias e velocidades diferentes. Esse é a base do conceito chamado Hierarquia de Memória.

Diante do contexto apresentado, responda ao que se pede a seguir.

- A) Explique o que é Hierarquia de Memória.
- B) Faça um desenho cuja forma represente a Hierarquia de Memória de um computador, com os seus respectivos níveis e nomes. Além disso, através de duas setas verticais ao lado deste desenho principal, uma apontando para cima e outra apontando para baixo, também utilizando as palavras “maior”, “menor”, “mais” e “menos”, indique a relação de capacidade, velocidade e custo dos níveis da Hierarquia de Memória.
- C) Além da Hierarquia de Memória, outro conceito importante relacionado às memórias com a finalidade de melhorar seu desempenho é o Princípio da Localidade. Explique o que é o Princípio da Localidade.
- D) Explique o que é Localidade Temporal no contexto do Princípio da Localidade.
- E) Explique o que é Localidade Espacial no contexto do Princípio da Localidade.

Questão 2

Em ambientes de supercomputação (HPC), é cada vez mais comum o uso de aceleradores, como as GPUs, para a realização de operações sobre grandes quantidades de dados, em paralelo, para a resolução de problemas computacionais complexos.

Considerando arquiteturas paralelas e o uso de aceleradores, responda ao que se pede a seguir.

- A) Segundo a taxonomia de Flynn, indique como as GPUs são tipicamente classificadas. Justifique.
- B) Em quais situações o paralelismo em GPU é recomendado? Justifique.
- C) Considere um vetor de inteiros de 32 posições e um laço de repetição que irá varrer este vetor, executando uma operação de potência elevada ao quadrado (x^2) e armazenando o resultado na mesma posição do vetor. Com base na arquitetura CUDA, identifique as etapas necessárias para a execução de um kernel que realize essas mesmas operações em uma GPU.

Questão 3

Algoritmos de ordenação baseados em comparação interna possuem limite inferior da complexidade computacional $O(n \log n)$. O *Heap Sort* e o *Merge Sort* são exemplos de algoritmos de ordenação com esta complexidade computacional.

Vamos imaginar um algoritmo de ordenação hipotético, chamado Algoritmo *A*, com complexidade computacional $O(n \log n)$. Submetendo a este algoritmo uma instância de tamanho n , a ordenação será obtida em exatamente $n \log_2 n$ operações elementares. Considere que cada operação elementar é executada em uma unidade de tempo.

Podemos paralelizar a execução do Algoritmo *A* dividindo o conjunto de dados em tamanhos iguais por k processadores, aplicando o algoritmo *A* nas partições do conjunto e intercalando os resultados obtidos por cada processador.

Em função da forma de paralelização do algoritmo, concluiu-se que o passo de ordenação é paralelizável, mantendo seu desempenho, entretanto, um novo passo de intercalação linear não paralelizável é necessário. Considere que este passo de intercalação é executado em n operações elementares.

Levando em consideração este método de paralelização, responda ao que se pede a seguir.

- A) Haverá ganho de tempo de execução do algoritmo *A* paralelizado em relação ao executado por um único processador? Justifique sua resposta.
- B) Considerando a estratégia adotada, haverá alteração da complexidade computacional do Algoritmo *A* devido a paralelização? Justifique sua resposta.
- C) Supondo que submetemos uma instância de tamanho 10^6 e paralelizamos o Algoritmo *A* em 10 processadores e que o passo de intercalação executa em exatamente n operações não paralelizáveis, determine o *Speedup* obtido, se houver. Use $\log 2 = 0,3$.

Questão 4

Uma série de números A_n possui a seguinte lei de formação recursiva:

$$A_0 = 0,35$$

$$A_1 = 2,30$$

$$A_n = 5^{\cos(\pi A_{n-1})} \cdot \log_2\{\sin(\pi A_{n-2})\}, \text{ para } n > 1.$$

Com o objetivo de acelerar o processamento, têm-se à disposição dois *threads* extras para paralelizar a execução dos cálculos. O esqueleto do código está mostrado a seguir:

```
double exp2(double x); /* função que retorna 2^x */
double cos(double x); /* função que retorna cos(x) */
double sin(double x); /* função que retorna sen(x) */
double log2(double x); /* função que retorna log2(x) */
#define M_PI 3.14159
... /* declarações de variáveis compartilhadas */
void thread1(void)
{
    ...
}

void thread2(void)
{
    ...
}

void calculaAn(int numtermos)
{
    ... /* aqui usam-se algumas variáveis compartilhadas entre os threads */
    ... /* aqui são feitos cálculos com os resultados do processamento dos threads */
} /* existe também um controle de loop */
```

Usando as funções declaradas, proponha um código eficiente e sincronizado para as rotinas “calculaAn”, “thread1” e “thread2”, dividindo as operações matemáticas, de forma a acelerar o cálculo dos coeficientes da série.

Questão 5

Considere um sistema computacional de alto desempenho (*HPCC - High Performance Computing Cluster*) que consiste de um conjunto de servidores de processamento e de armazenamento de dados interligados através de uma rede de alta velocidade operando como um sistema unificado, capaz de receber e executar tarefas computacionais de forma paralela. Nesse tipo de ambiente, os usuários definem as tarefas ('jobs') e as submetem ao sistema para serem executadas. As tarefas, que podem ser independentes ou acopladas, são submetidas a um sistema de escalonamento ('job scheduler' ou 'job manager') e então executadas de acordo com a necessidade de recursos especificada por cada usuário e com a disponibilidade desses recursos.

Com base nas definições acima, responda ao que se pede a seguir.

- A) Considerando um HPCC formado por servidores multiprocessados com arquitetura do tipo MIMD (*multiple instruction, multiple data*), defina os conceitos de arquitetura de memória compartilhada e de memória distribuída. Nesses termos, esse cluster pode ser considerado um sistema de arquitetura híbrida? Justifique.
- B) Descreva o que você entende por escalonamento de tarefas em um sistema do tipo HPCC e cite dois escalonadores comumente usados em sistemas desse tipo.
- C) Cite duas tecnologias de interconexão mais comumente usadas nesses sistemas; descreva sucintamente cada uma delas e comente as vantagens e desvantagens da adoção de cada tecnologia em relação à outra.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30

Realização

