

Verificando a influência do processador em aplicações paralelas de álgebra linear densa

CMP 270 - Introduction to High Performance Computing

Lucas Barros de Assis

Motivação

- As máquinas tupi1~tupi4 do PCAD contam ambas com placas NVIDIA RTX4090, mas com processadores diferentes

Motivação

- As máquinas tupi1~tupi4 do PCAD contam ambas com placas NVIDIA RTX4090, mas com processadores diferentes
- Considerando a enorme influência da GPU no poder de processamento, a diferença de custo entre os processadores é proporcional ao ganho obtido?

Motivação

- As máquinas tupi1~tupi4 do PCAD contam ambas com placas NVIDIA RTX4090, mas com processadores diferentes
- Considerando a enorme influência da GPU no poder de processamento, a diferença de custo entre os processadores é proporcional ao ganho obtido?

Intel Xeon E5-2620 V4 Octa-core [8 Core] 2.10 Ghz Processor - Socket R3 [lga2011-3]oem Pack - 2 Mb - 20 Mb Cache - 8 Gt/s Qpi - 64-bit Processing - 3 Ghz Overclocking Speed - 14 Nm - 85 W -

[Visit the Intel Store](#)

5.0 ★★★★★ 1 rating | [Search this page](#)

\$189⁹⁵

Intel® Core™ i9-14900KF New Gaming Desktop Processor 24 cores (8 P-cores + 16 E-cores) - Unlocked

[Visit the Intel Store](#)

4.6 ★★★★★ 1,116 ratings | [Search this page](#)

Amazon's Choice

in Computer CPU Processors by Intel

100+ bought in past month

-11% \$548⁹⁹

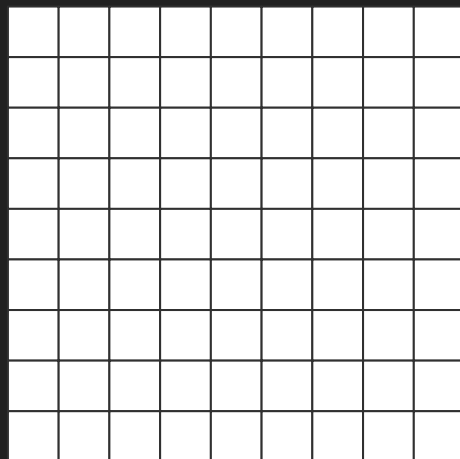
Experimentos

- O desempenho dos nós foi analisado, utilizando o binário de teste da biblioteca paralela de álgebra linear densa *Chameleon*, aplicando a operação de fatoração *LU*

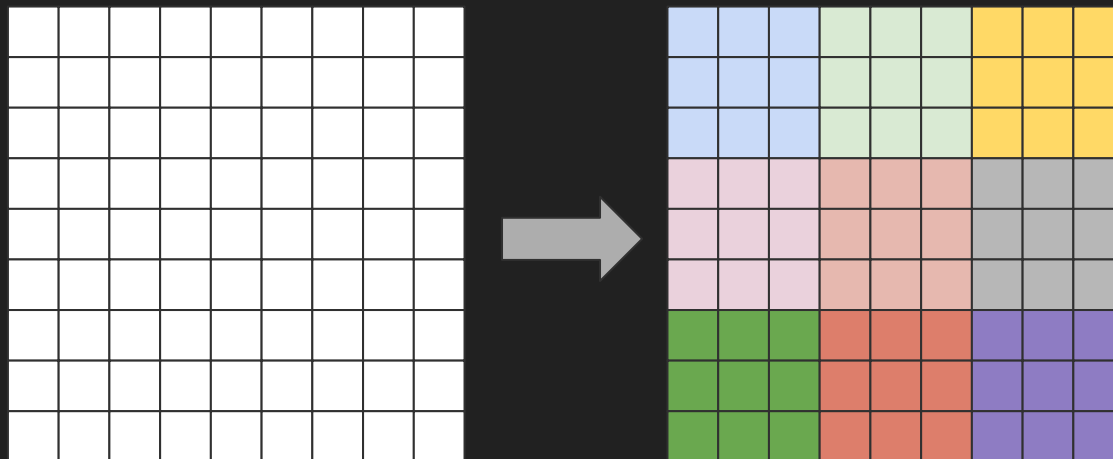
Experimentos

- O desempenho dos nós foi analisado, utilizando o binário de teste da biblioteca paralela de álgebra linear densa *Chameleon*, aplicando a operação de fatoração *LU*
- O paralelismo foi obtido utilizando o suporte de execução *StarPU*, já integrado à biblioteca *Chameleon*

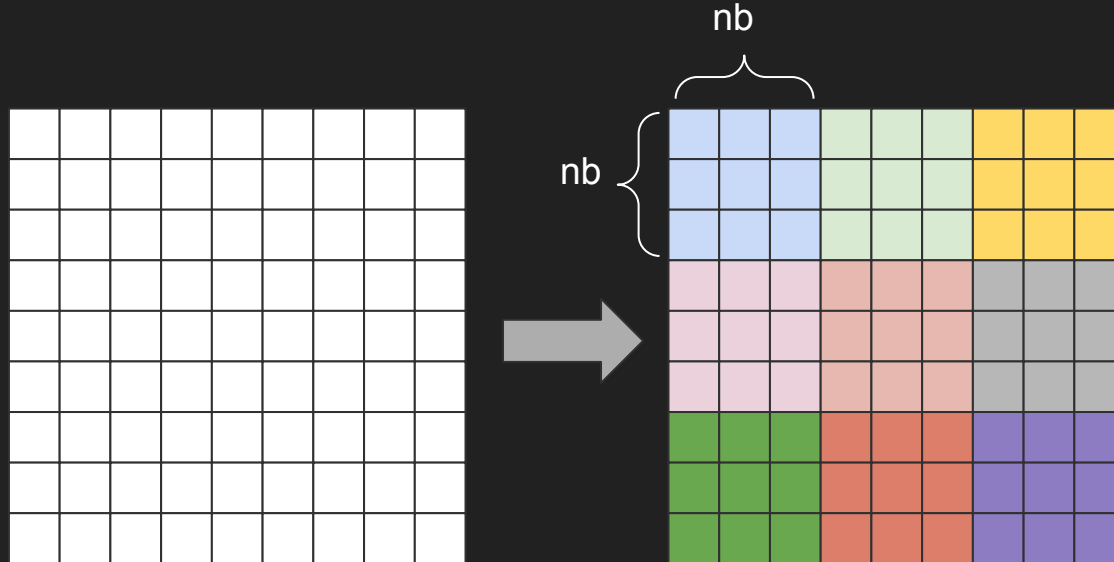
Matrizes ladrilhadas



Matrizes ladrilhadas



Matrizes ladrilhadas



Testando os processadores

- Ambos processadores foram testados realizando a fatoraçoão LU em uma matriz quadrada de ordem 65535 e em precisão simples

Testando os processadores

- Ambos processadores foram testados realizando a fatoração LU em uma matriz quadrada de ordem 65535 e em precisão simples
- Analisando o Xeon, foram considerados os casos sem *Hyper-threading* e com *Hyper-threading*

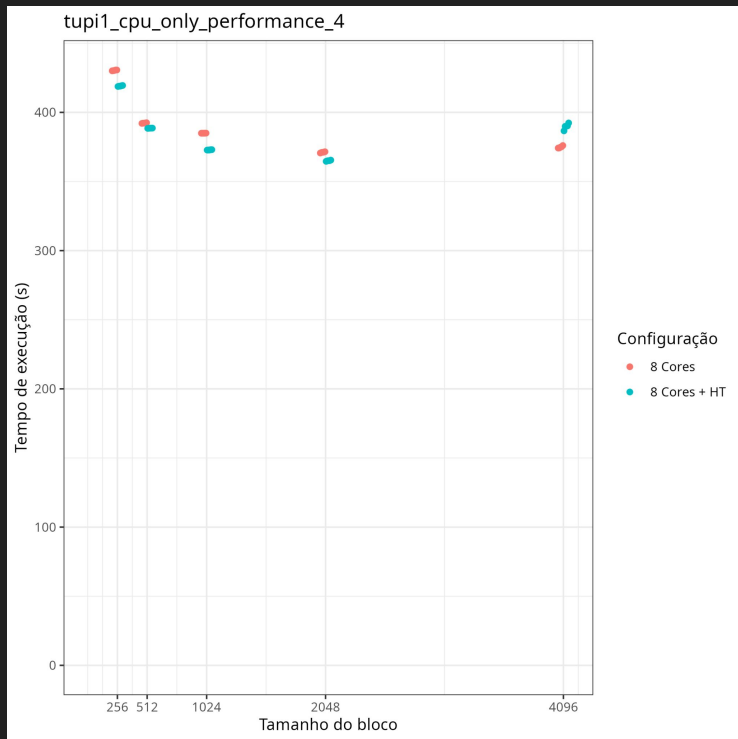
Testando os processadores

- Ambos processadores foram testados realizando a fatoração LU em uma matriz quadrada de ordem 65535 e em precisão simples
- Analisando o Xeon, foram considerados os casos sem *Hyper-threading* e com *Hyper-threading*
- Analisando o i9, além de 8 núcleos convencionais, chamados *P-cores*, e capazes de utilizar *Hyper-threading*, também existem 16 *E-cores*, concebidos para maior eficiência energética e sem a possibilidade de utilizar *Hyper-threading*

Testando os processadores

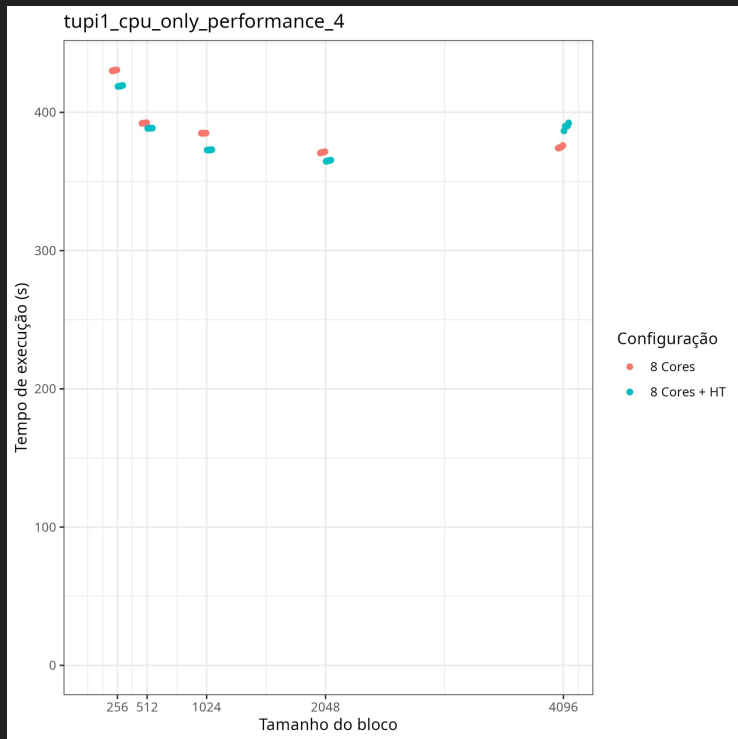
- Ambos processadores foram testados realizando a fatoração LU em uma matriz quadrada de ordem 65535 e em precisão simples
- Analisando o Xeon, foram considerados os casos sem *Hyper-threading* e com *Hyper-threading*
- Analisando o i9, além de 8 núcleos convencionais, chamados *P-cores*, e capazes de utilizar *Hyper-threading*, também existem 16 *E-cores*, concebidos para maior eficiência energética e sem a possibilidade de utilizar *Hyper-threading*
- Os experimentos foram realizados com diferentes tamanhos de bloco com o intuito de encontrar o melhor desempenho de cada processador

Testando os processadores - Xeon



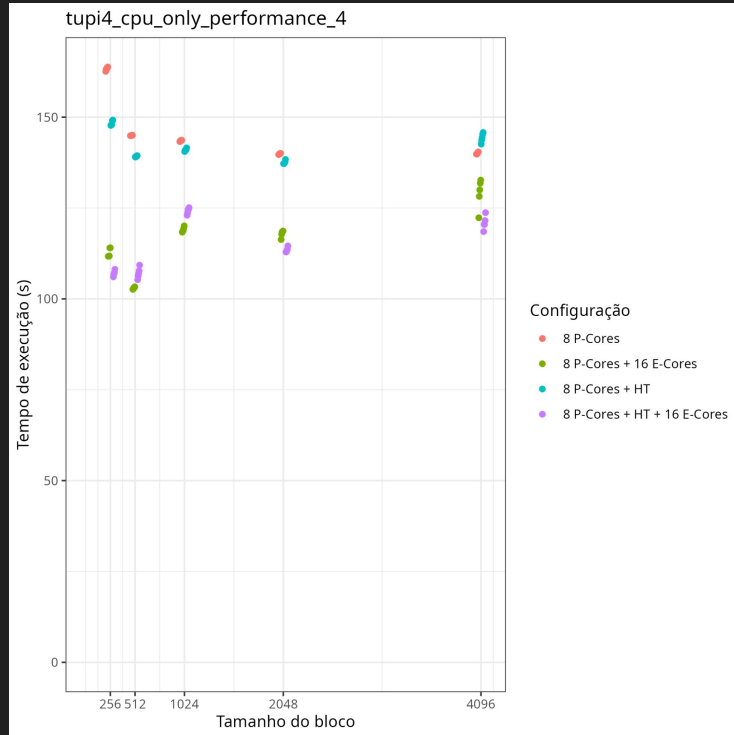
- Blocos muito pequenos sub-utilizam o poder de processamento, mas permitem um ganho ao utilizar o *Hyper-threading*

Testando os processadores - Xeon



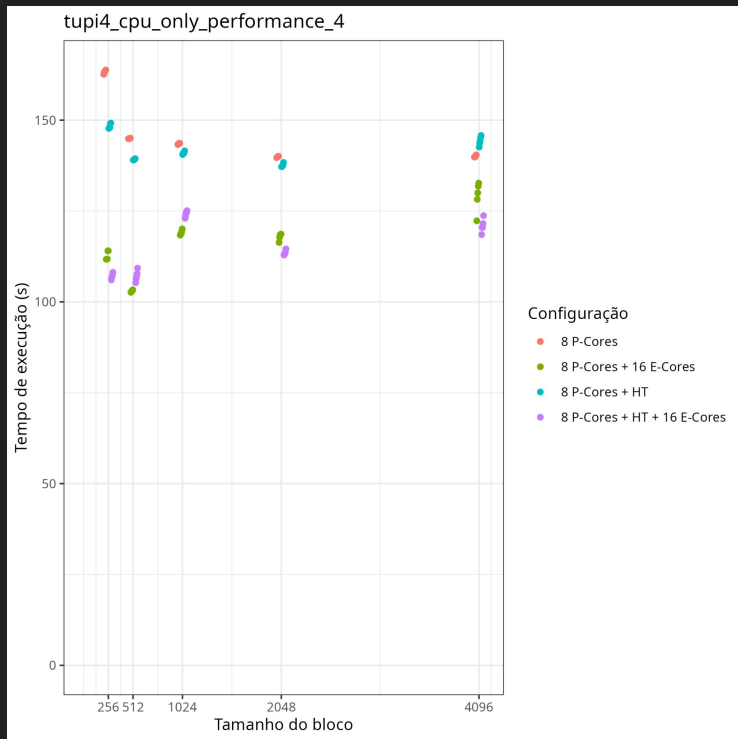
- Blocos muito pequenos sub-utilizam o poder de processamento, mas permitem um ganho ao utilizar o *Hyper-threading*
- Blocos maiores deixam de apresentar ganhos com o *Hyper-threading*, provavelmente por exceder o espaço disponível na memória cache

Testando os processadores - i9



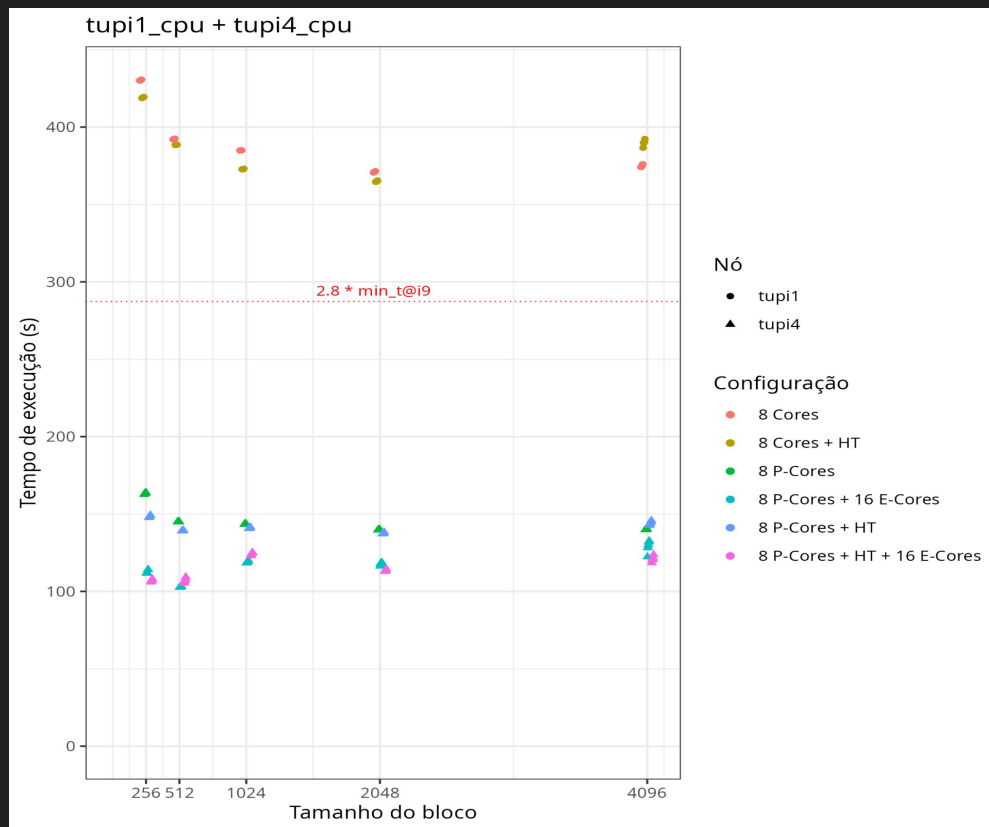
- Os casos utilizando apenas P-cores apresentam um comportamento semelhante ao visto no Xeon

Testando os processadores - i9



- Os casos utilizando apenas P-cores apresentam um comportamento semelhante ao visto no Xeon
- Os casos utilizando os E-cores apresentam um aumento no desempenho, mas também aparentam ter uma variabilidade maior

Comparação entre os processadores



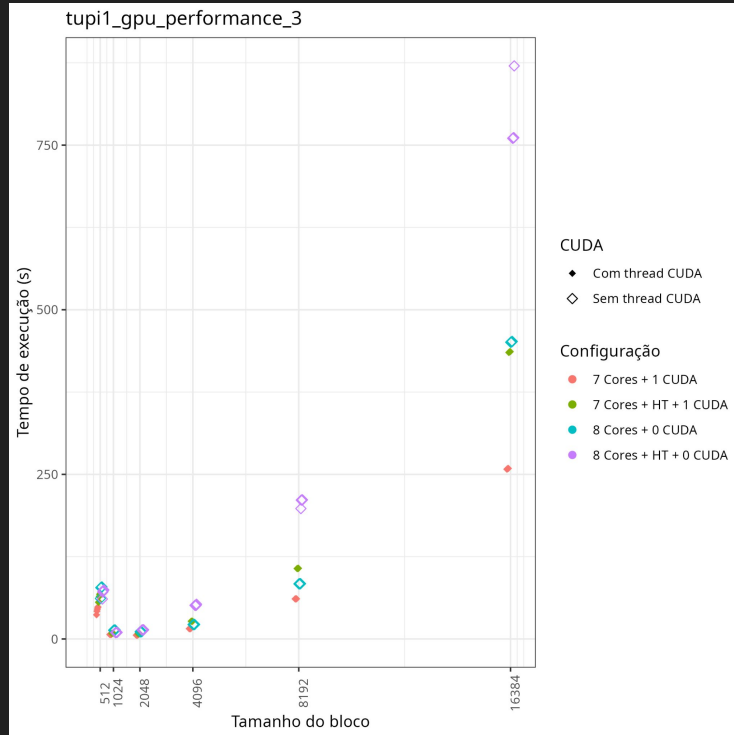
Testando os nós

- Para compararmos de fato o desempenho dos nós, adicionamos a GPU aos experimentos anteriores

Testando os nós

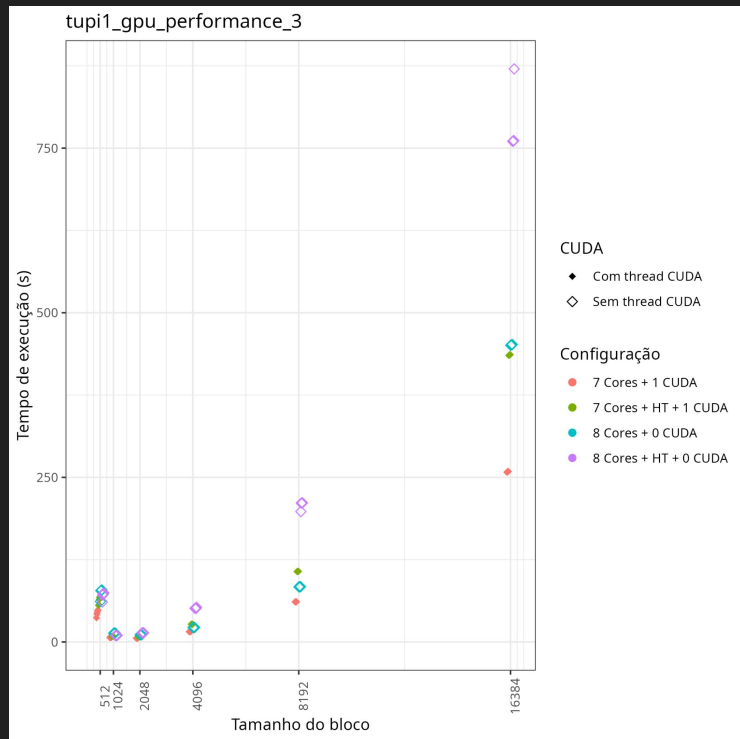
- Para compararmos de fato o desempenho dos nós, adicionamos a GPU aos experimentos anteriores
- Ao utilizar uma *GPU*, uma *thread* fica responsável pelas operações *CUDA* necessárias. O suporte de execução *StarPU* permite configurar se essa *thread* trabalhará exclusivamente com as operações *CUDA* ou se trabalhará também nos cálculos realizados. Os experimentos realizados contemplaram ambos os casos

Testando os nós - Xeon



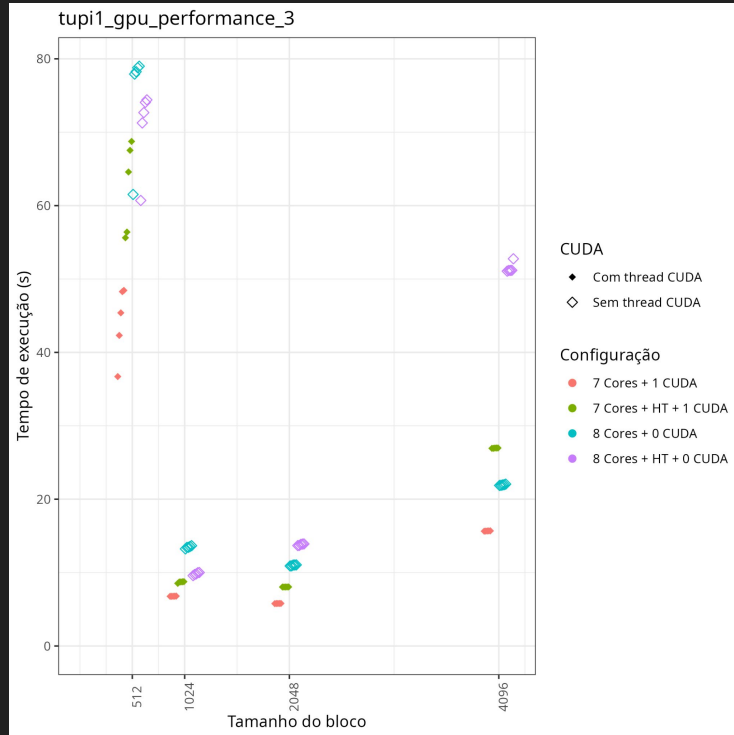
- Para uma mesma configuração de *HT*, a exclusividade da *thread CUDA* se mostrou sempre superior ao outro caso

Testando os nós - Xeon



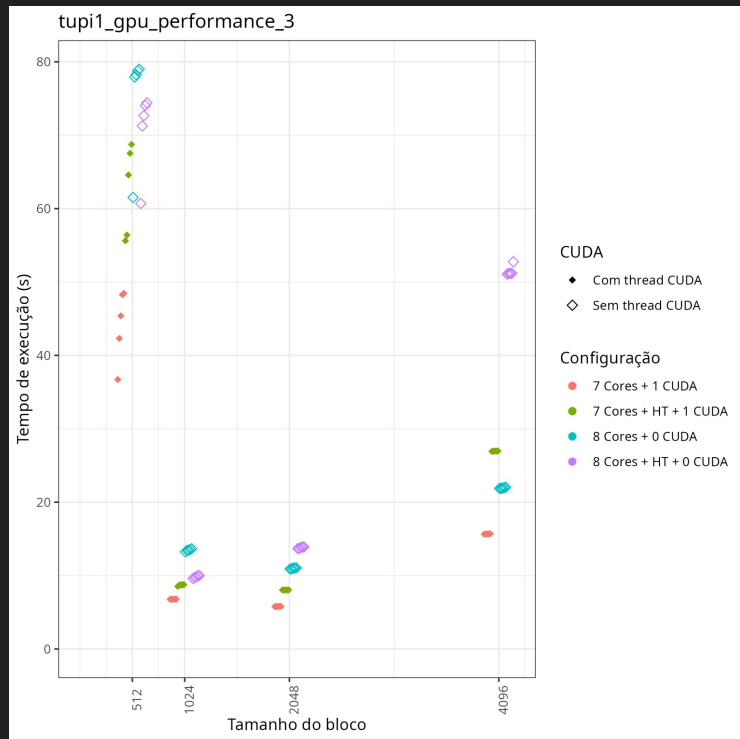
- Para uma mesma configuração de *HT*, a exclusividade da *thread CUDA* se mostrou sempre superior ao outro caso
- O desempenho ao usar *HT* diminui a medida em que o tamanho do bloco aumenta

Testando os nós - Xeon



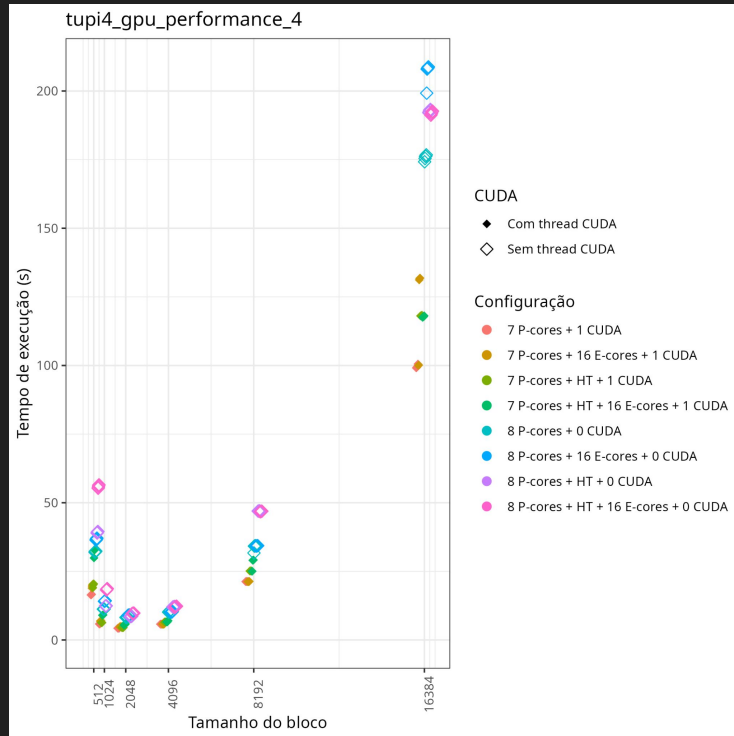
- Focando nos melhores resultados, fica nítida a desvantagem em não dedicar uma *thread* para operações *CUDA*

Testando os nós - Xeon



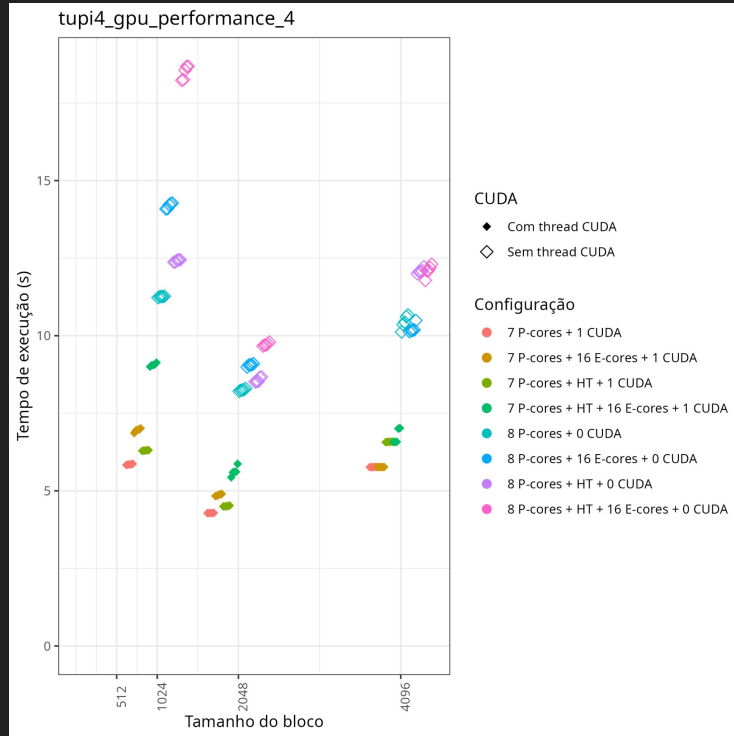
- Focando nos melhores resultados, fica nítida a desvantagem em não dedicar uma *thread* para operações *CUDA*
- Excluindo esse critério, o comportamento é muito similar ao visto utilizando apenas o processador

Testando os nós - i9



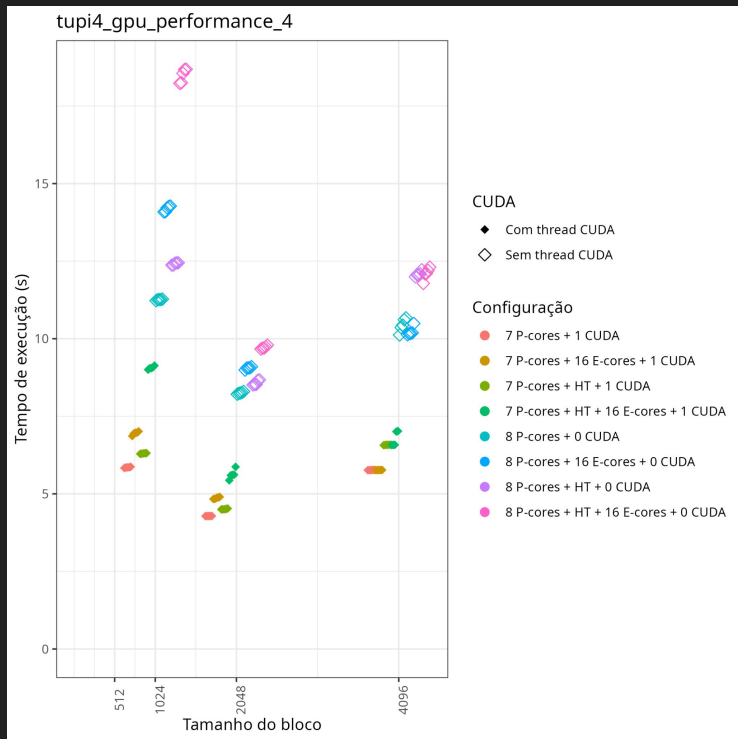
- O i9 reforça o interesse em dedicar uma *thread* às operações *CUDA*

Testando os nós - i9



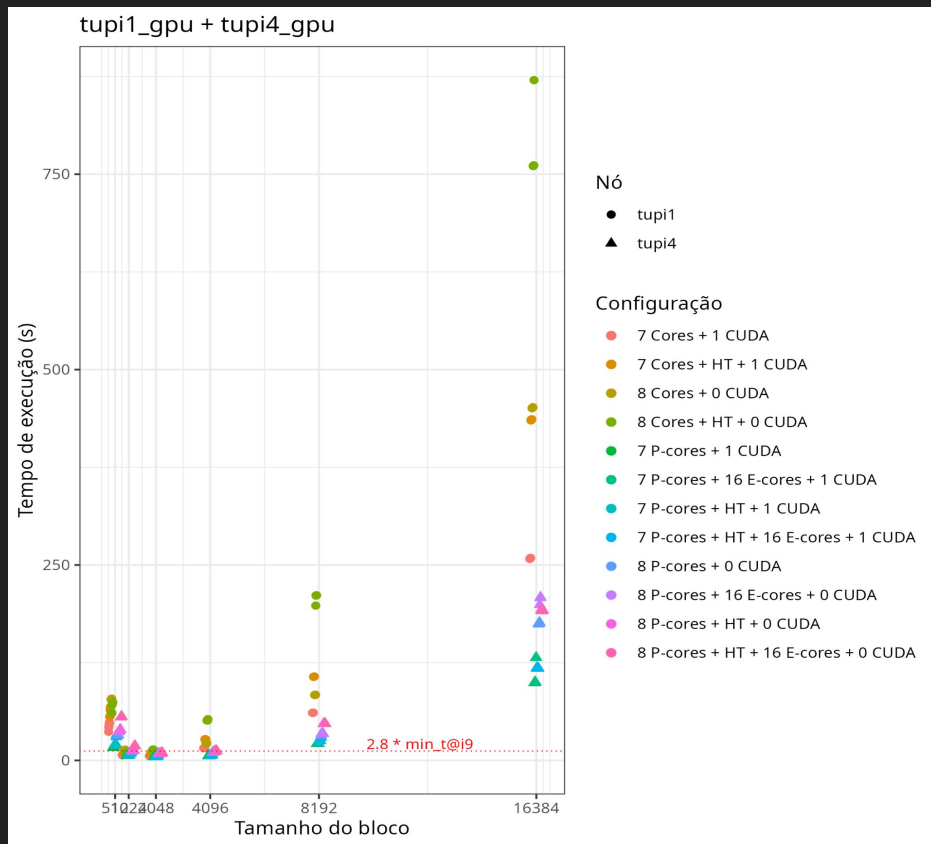
- Em geral, os *E-cores* prejudicaram o desempenho dos testes realizados

Testando os nós - i9

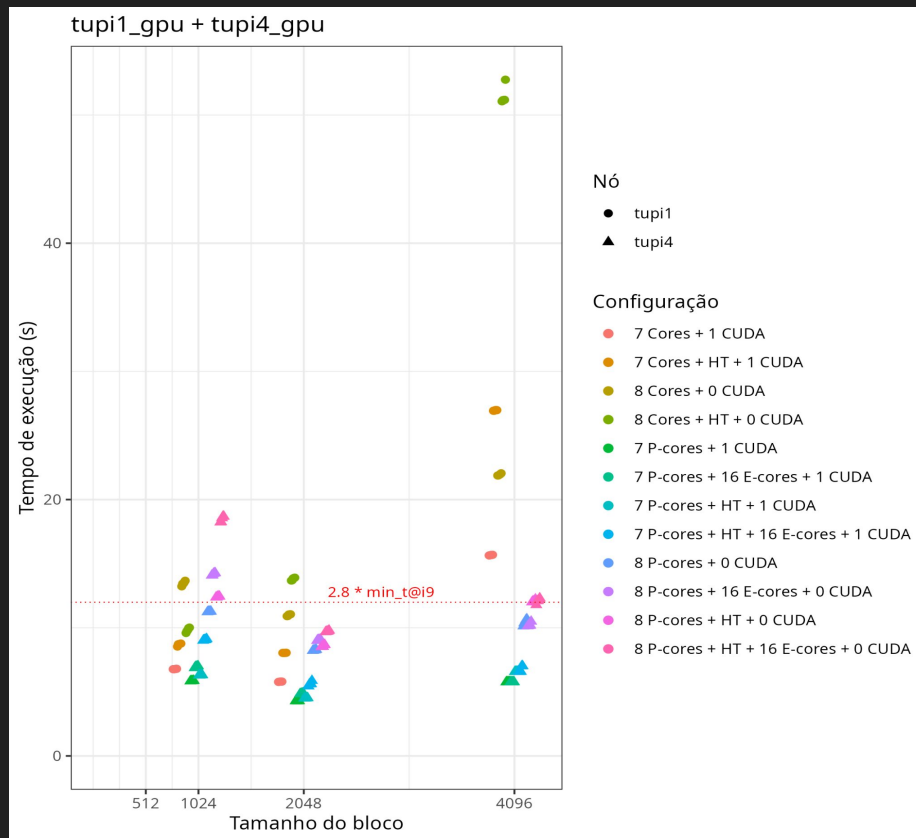


- Em geral, os *E-cores* prejudicaram o desempenho dos testes realizados
- Para grandes tamanhos de bloco, os *E-core* passam a apresentar uma contribuição, possivelmente graças à sua memória cache

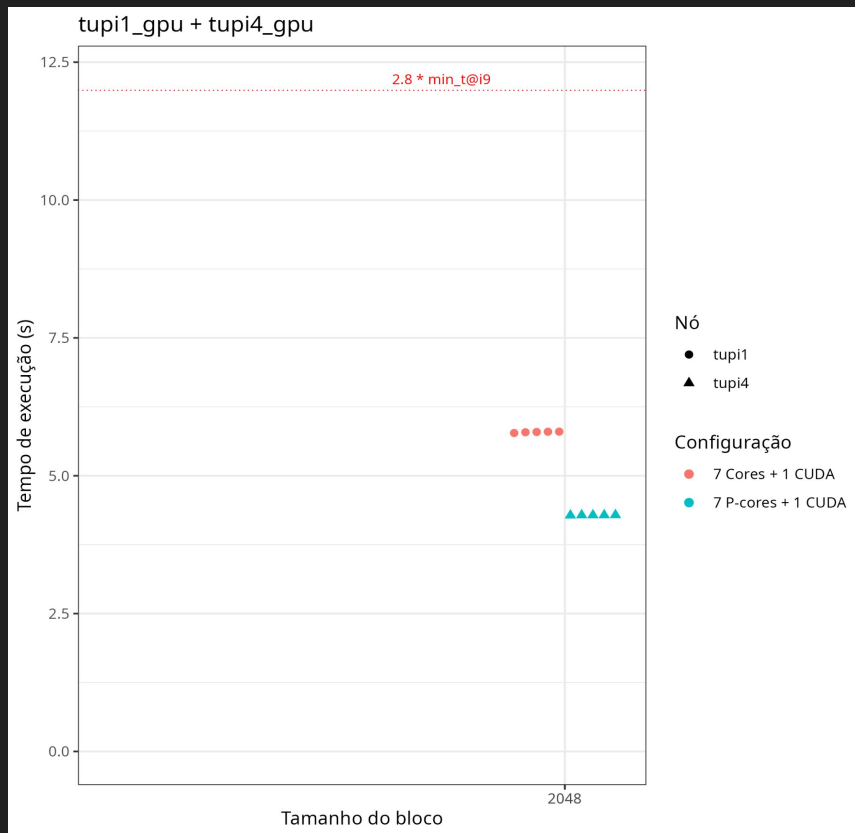
Comparação entre os nós



Comparação entre os nós

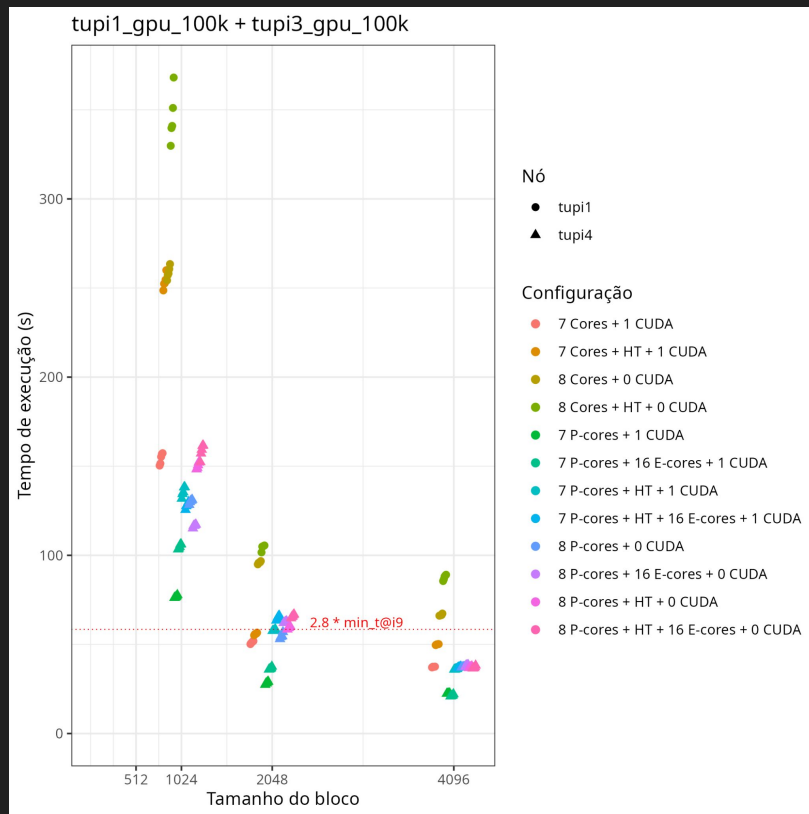


Comparação entre os nós

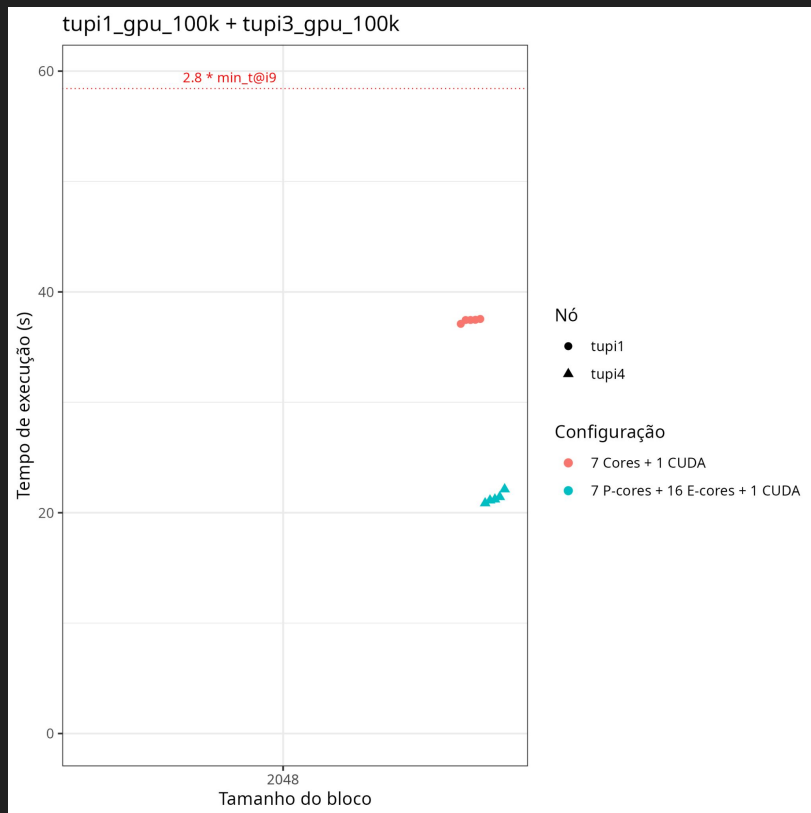


- O tempo de execução do melhor caso da *tupi1* foi aproximadamente **1,34** vezes maior que o da *tupi4*

Ocupando a GPU



Ocupando a GPU



- Com cerca de 90% da memória da GPU utilizada, o tempo de execução do melhor caso da *tupi1* foi aproximadamente **1,77** vezes maior que o da *tupi3*

Conclusão

- Apesar de oferecer um desempenho superior, a diferença de custo entre os processadores é maior que o ganho de performance obtido

Conclusão

- Apesar de oferecer um desempenho superior, a diferença de custo entre os processadores é maior que o ganho de performance obtido
- Podemos observar que a técnica de *Hyper-threading* só é vantajosa utilizando tamanhos de bloco menores do que o valor ótimo encontrado nos testes

Conclusão

- Apesar de oferecer um desempenho superior, a diferença de custo entre os processadores é maior que o ganho de performance obtido
- Podemos observar que a técnica de *Hyper-threading* só é vantajosa utilizando tamanhos de bloco menores do que o valor ótimo encontrado nos testes
- A utilização de uma *thread* exclusivamente para as operações *CUDA* se mostrou sempre a melhor alternativa

Obrigado pela atenção

Dúvidas?

