

MC970/MO644 – Programação Paralela

Laboratório 13 – Multiplicação de Matrizes Distribuída

Professor: Guido Araújo

Monitor: Hervé Yviquel

1 Enunciado

Neste laboratório, o objetivo é paralelizar multiplicação de matrizes ($C=A.B$) na nuvem usando Apache Spark.

- A primeira tarefa é implementar a multiplicação de matrizes da maneira que quiserem mas usando as funções de paralelização de Spark. A implementação está totalmente livre: pode ser em Scala ou Python usando um Notebook Jupyter ou não.
- A segunda tarefa é criar um cluster Spark no Microsoft Azure e testar a execução na nuvem. Depois da execução, devem analisar a paralelização usando a interface gráfica de profiling do Spark e tentar otimizar ao máximo o tempo de execução. Também, devem modificar o tamanho das matrizes para observar o efeito na execução. A implementação deve ser testada com matrizes de float geradas aleatoriamente e de vários tamanhos para mostrar a escalabilidade da implementação (por exemplo 2000, 8000, e 16000).

Caso tenha alguma dúvida, use o Google Groups - para este trabalho está liberado discutir a solução direta do problema.

2 Submissões

A submissão deve ser **um arquivo** (em zip) contendo o **código** (com os arquivos necessários para compilar e testar) e um **relatório** (em pdf). O relatório deve descrever a implementação, apresentar e analisar os resultados das experimentações na nuvem, e descrever as otimizações que foram aplicadas. Além disso, o relatório deve conter uma captura de tela da interface gráfica de perfilamento do Spark mostrando a execução nos núcleos do cluster em paralelo.