

Grupo 01

André Luís Mendes Fakhoury

Debora Buzon da Silva

Gustavo Vinicius Vieira Silva Soares

Thiago Preischadt Pinheiro

Projeto do algoritmo paralelo utilizando a metodologia PCAM

Particionamento

A matriz de dimensão $N \times M$ foi transposta de forma que os dados da amostra (representados por uma coluna) ficam contíguos na memória pois ficaram distribuídos em linhas. Após essa transposição, a matriz ficará com dimensões $M \times N$.

O cálculo das métricas é dividido em $6 \times M$ tarefas e para cada linha, são agrupadas da seguinte forma: primeiro, é feita a ordenação; após isso é feito o cálculo da média aritmética, média harmônica, mediana e moda. Por fim, é feito, em uma única tarefa, o cálculo da variância, desvio padrão e coeficiente de variação.

Comunicação

A tarefa de ordenação recebe o vetor respectivo à uma linha e o seu resultado é passado às outras tarefas que atuam no mesmo vetor. Além disso, o resultado da tarefa da média aritmética é passado para a tarefa que calcula a variância em uma linha. Todas as métricas são armazenadas em uma matriz final.

Aglomerção

Dado que a plataforma alvo é um *cluster de computadores*, as $6 \times M$ tarefas serão agrupadas em P processos. Cada processo é responsável por $6 \times \text{floor}(M / P)$ tarefas e as $6 \times (M \bmod P)$ tarefas restantes são mapeadas para os $(M \bmod P)$ primeiros processos, de forma que as tarefas relacionadas a mesma linha são colocadas no mesmo processo.

Mapeamento

Caso o desempenho dos nós do cluster seja homogêneo, o mapeamento de cada um dos P processos em *PROC* elementos de processamentos pode ocorrer por meio de uma fila circular. Caso o desempenho dos nós seja diferente, o mapeamento pode deixar de ser estático e ser dinâmico, atribuindo cada processo ao nó com menor carga de trabalho (definida por alguma métrica de desempenho).