

Algoritmo k-Means

Implementação em
memória partilhada e
distribuída



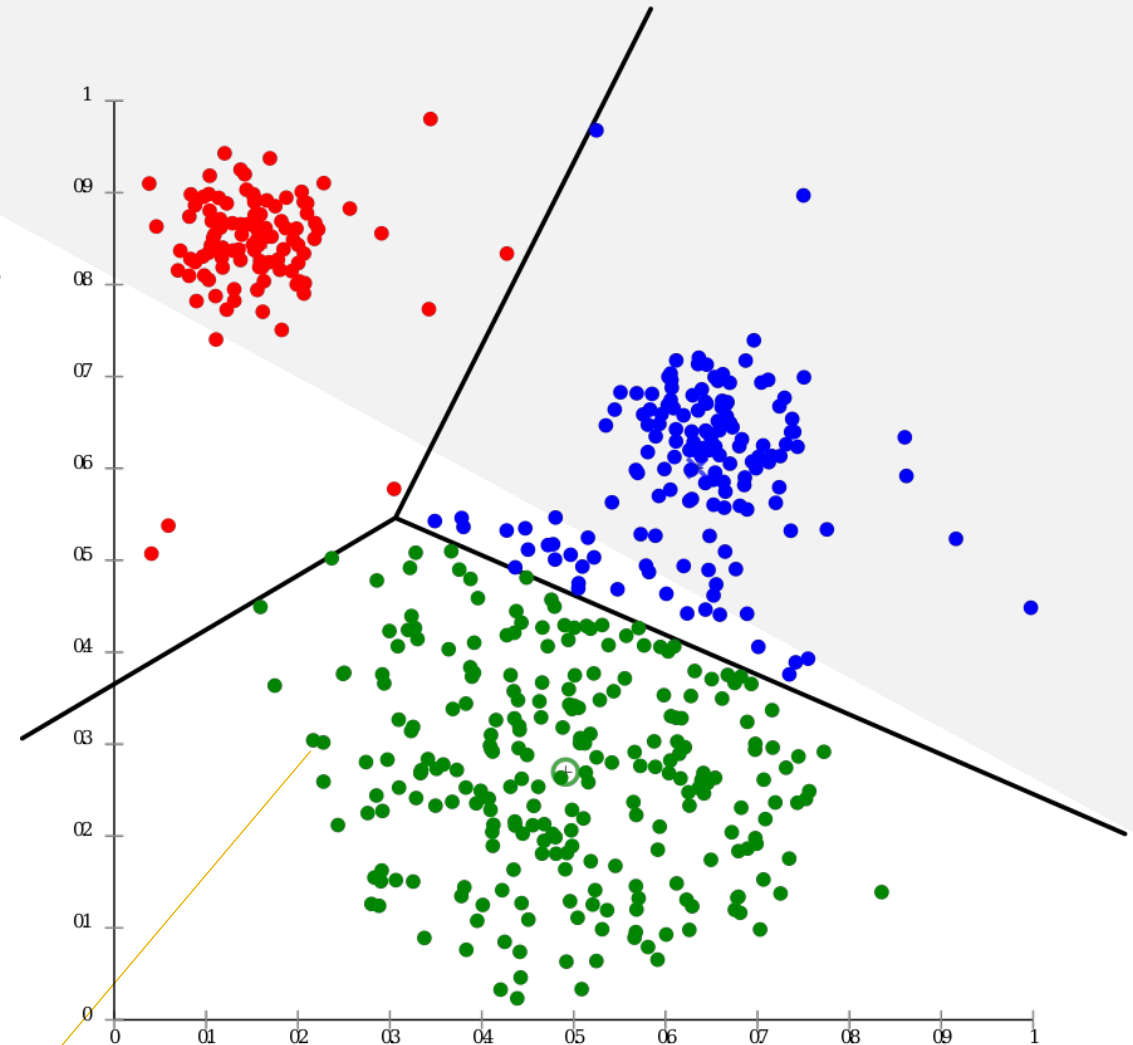
K-Means

- Particiona informação em K grupos seguindo critérios de semelhança.

Caso de estudo

Pontos em coordenadas cartesianas

- Seleciona k centroides;
- Adiciona cada ponto ao cluster cujo centroide está mais perto.
- Calcula o novo centroide de cada cluster
- Calcula o erro

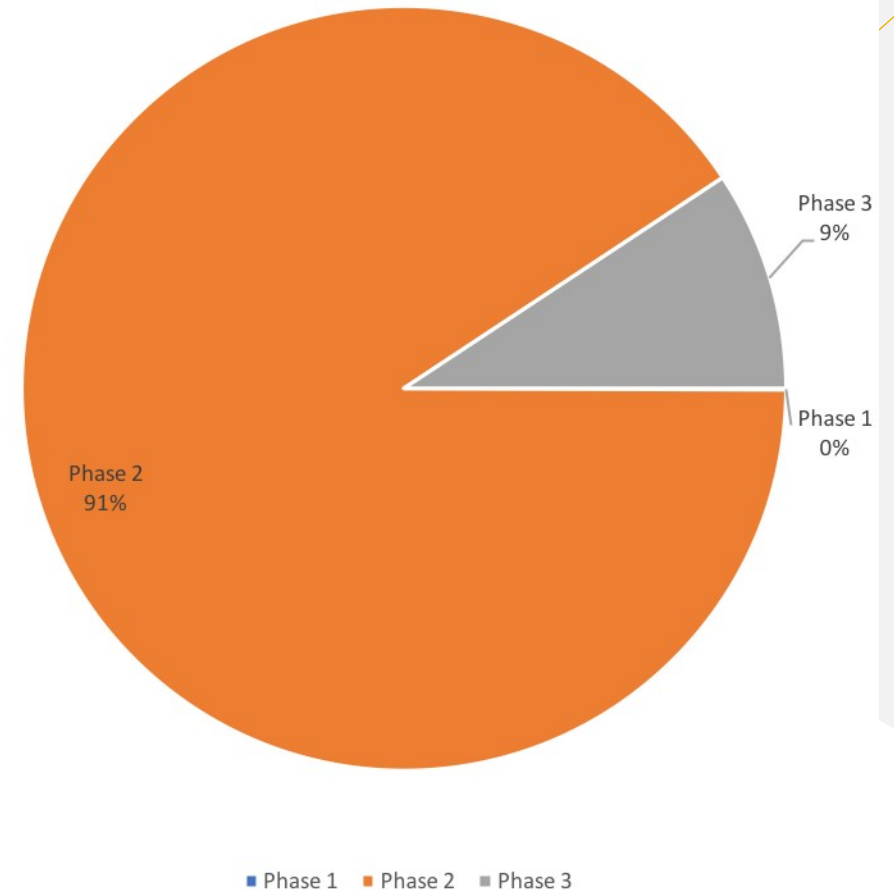


Implementação

Memória partilhada

Fase 2

- Utilização de redução de arrays
- Resolve problemas de false sharing
- Evita writing-conflicts

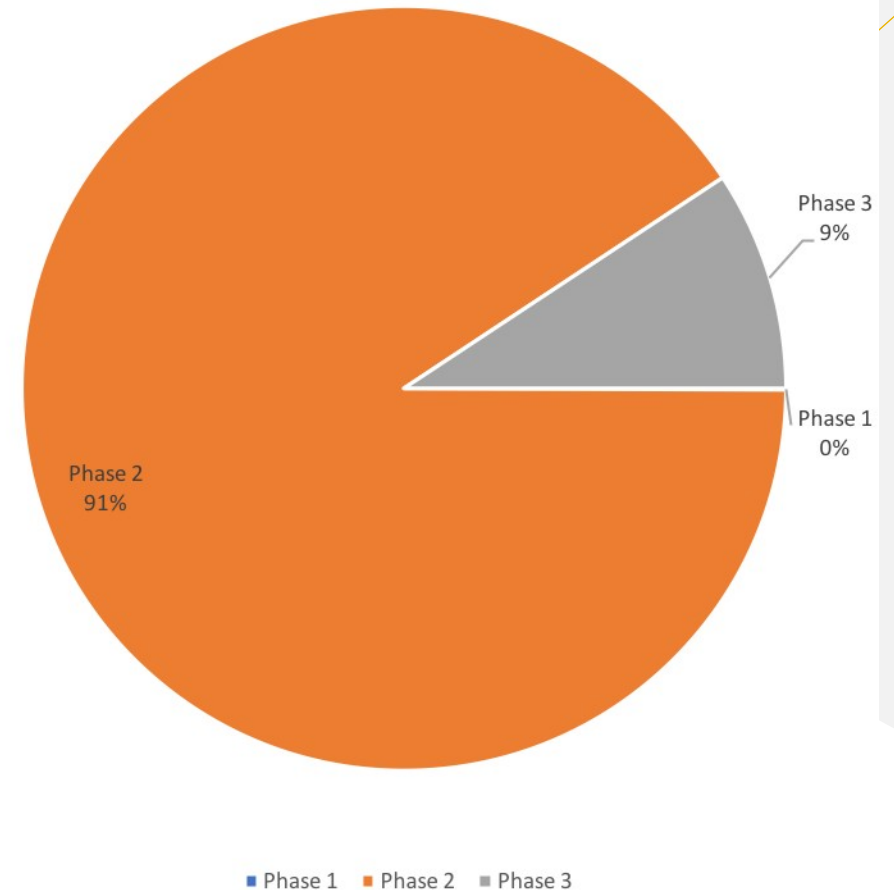


Implementação

Memória partilhada

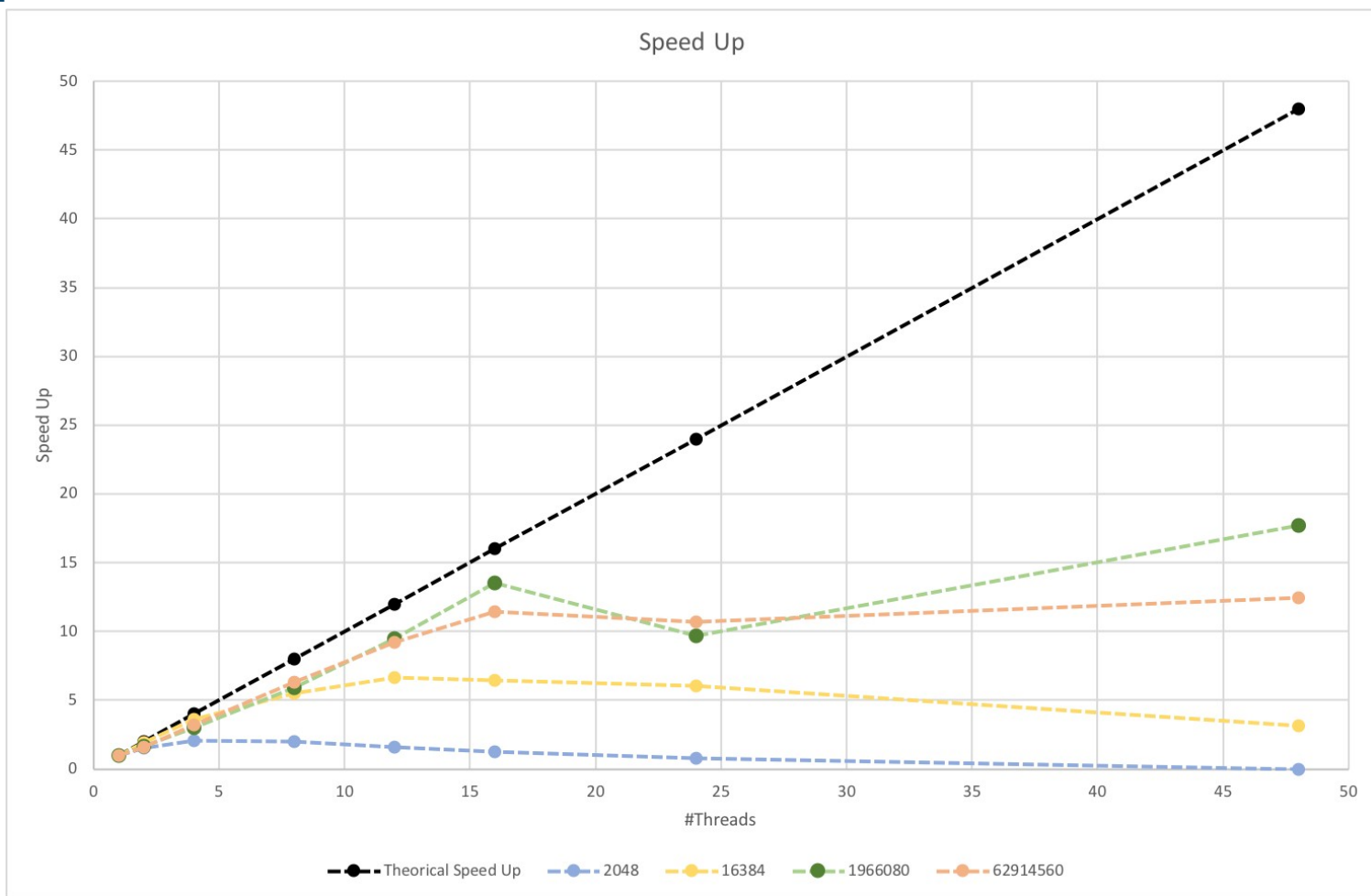
Fase 3

- Utilização de redução de arrays
- Componente x
- Componente y



Results

Memória partilhada

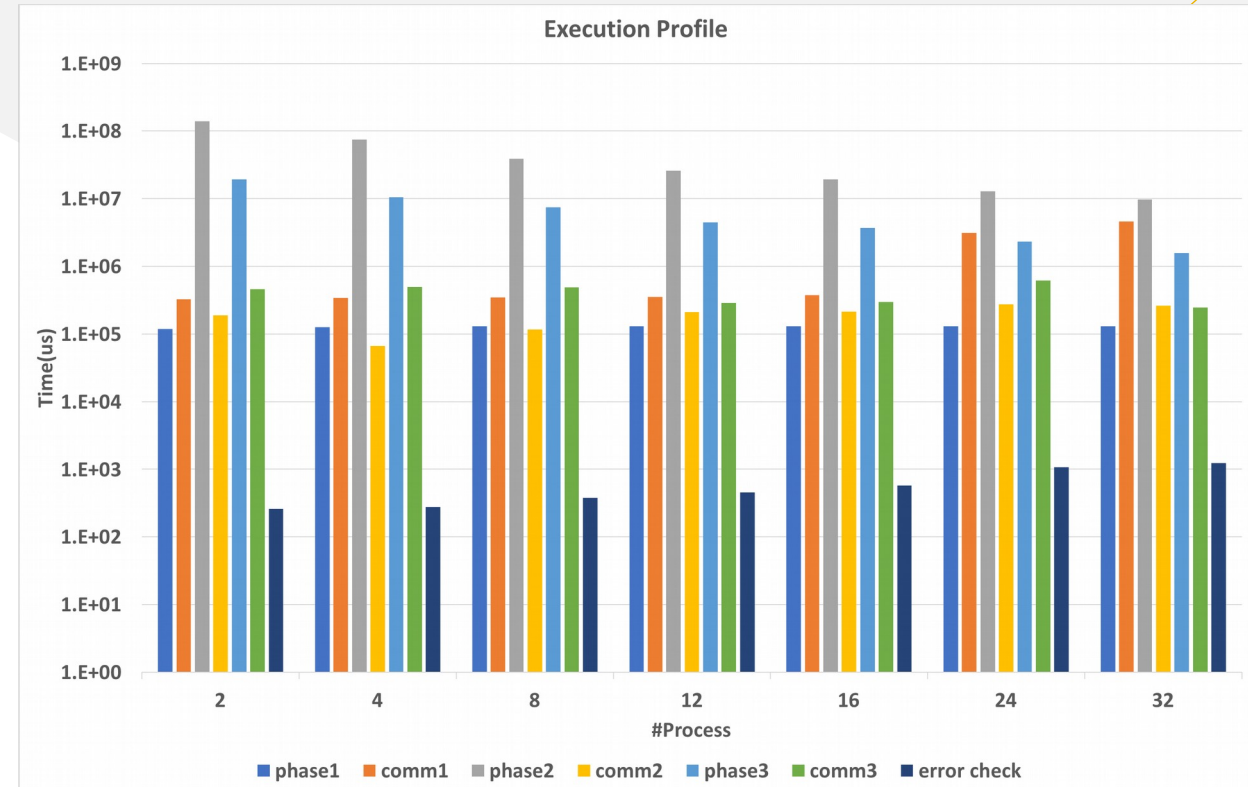


Implementação

Memória Distribuída

Fase 2 e 3 de computação

- Computação é igualmente distribuída
- Utilização de primitivas coletivas
- Duas estratégias estudadas
BroadCast/reduce e allReduce

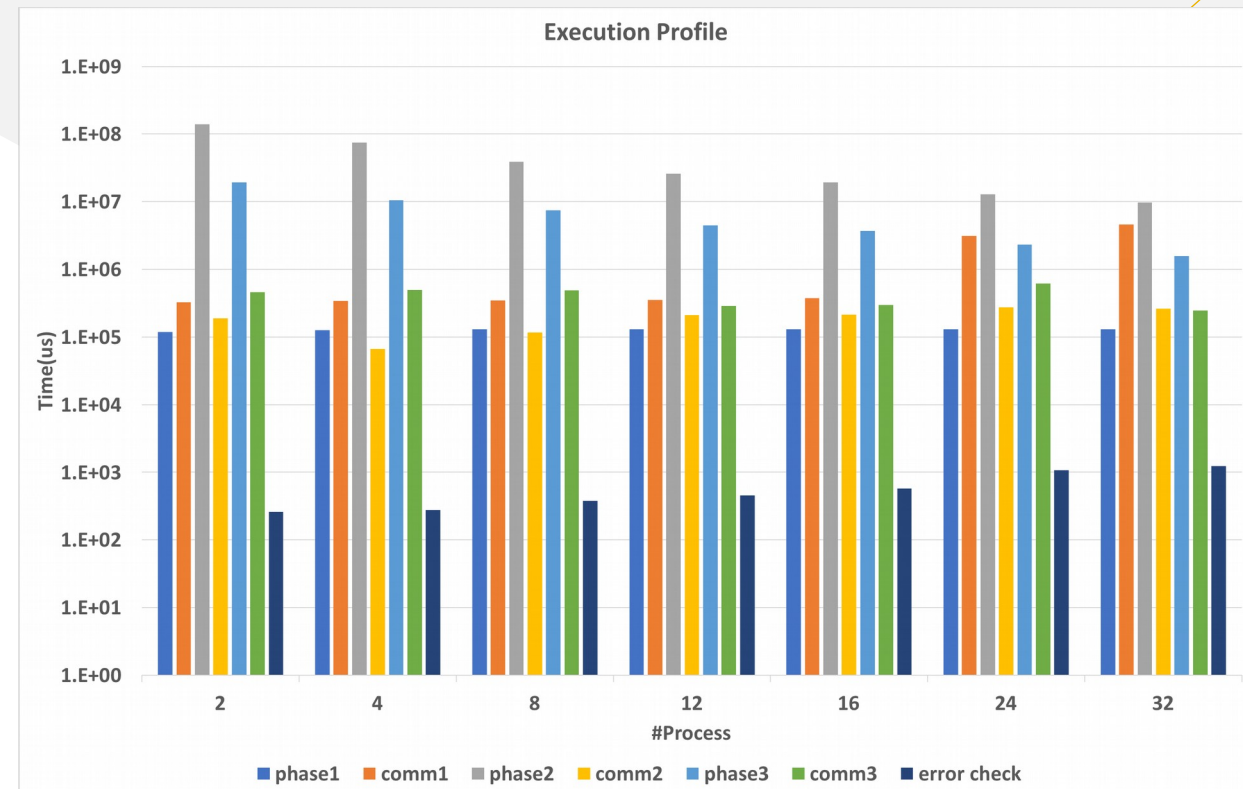


Implementação

Memória Distribuída

Comunicação 1→2 Fase

- O data set e os centroides são repartidos por N processos

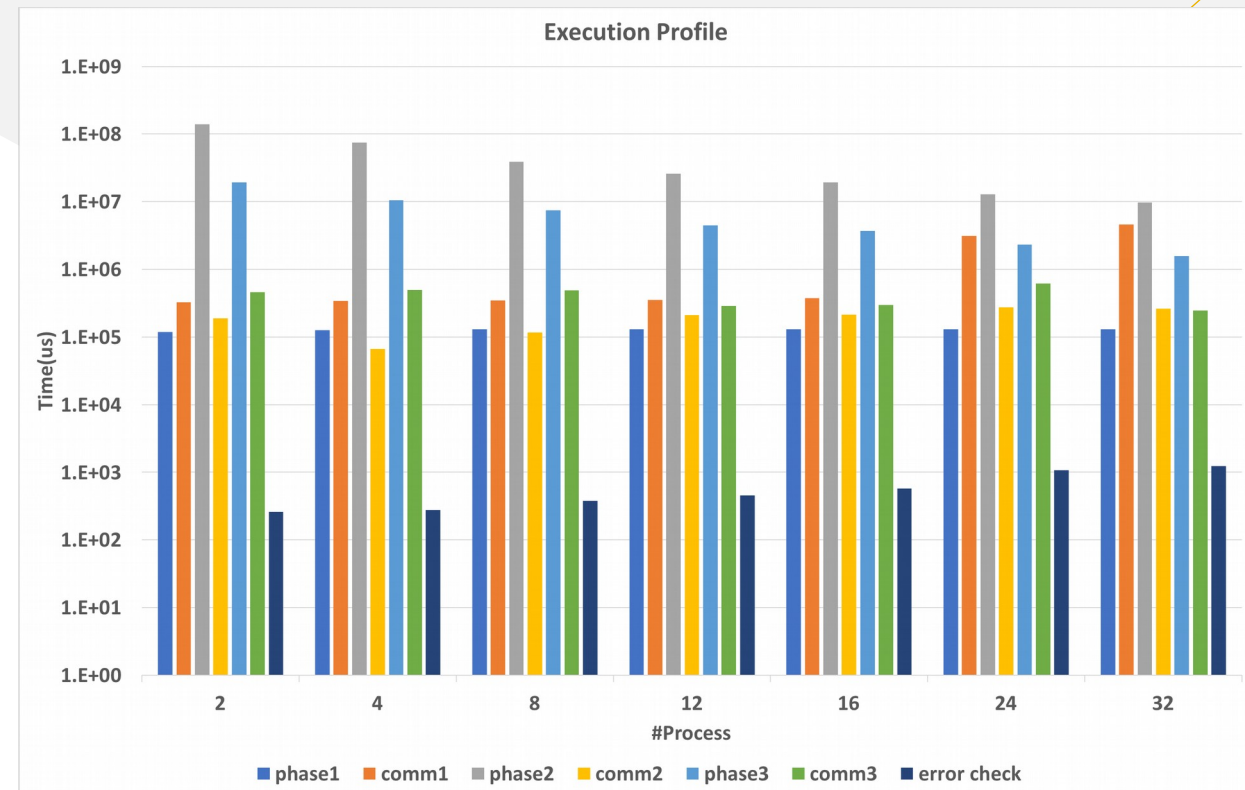


Implementação

Memória Distribuída

Comunicação 2→3 Fase

- O data set e os centroides de cada processo são partilhados

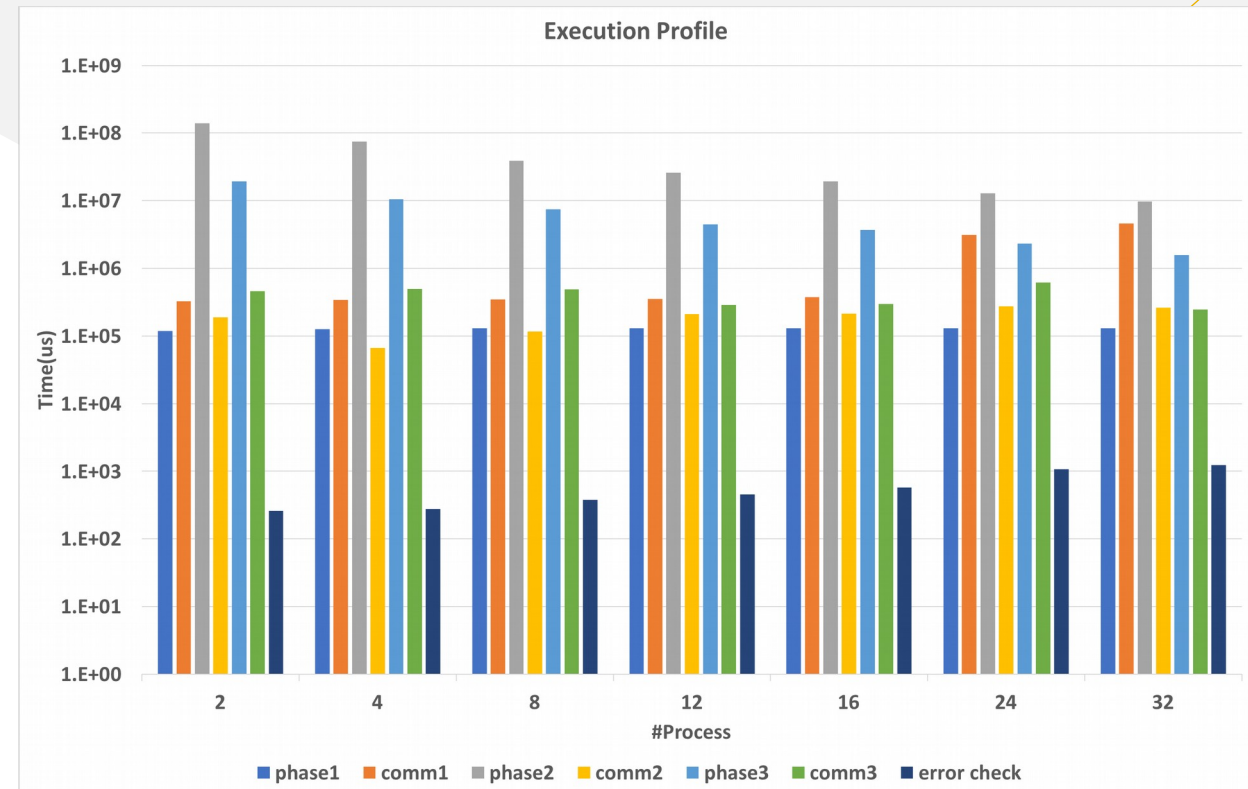


Implementação

Memória Distribuída

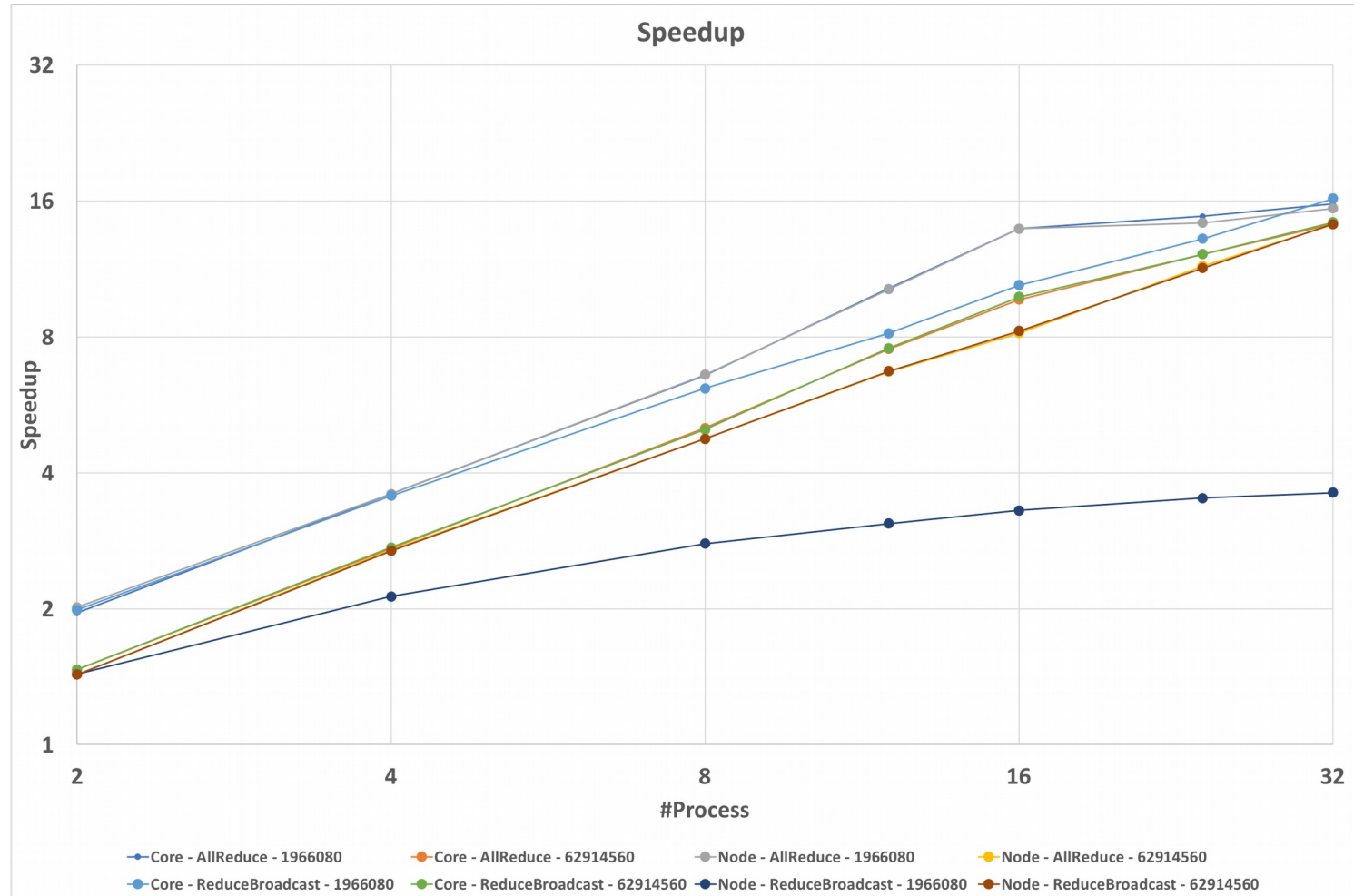
Comunicação 3→2 Fase

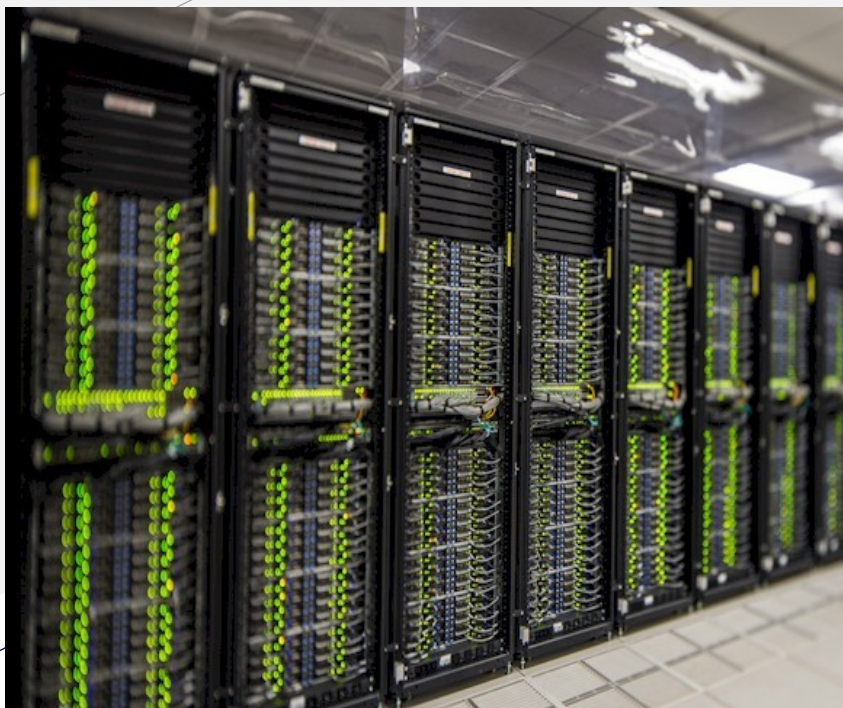
- Os centroides são partilhados
- O processo master envia um inteiro para saber se o processo é repetido.
- Em caso de convergencia, os sets são transferidos para o Master.



Results

Memória Distribuída





Algoritmo k-Means

Implementação em
memória partilhada e
distribuída