MAPREDUCE

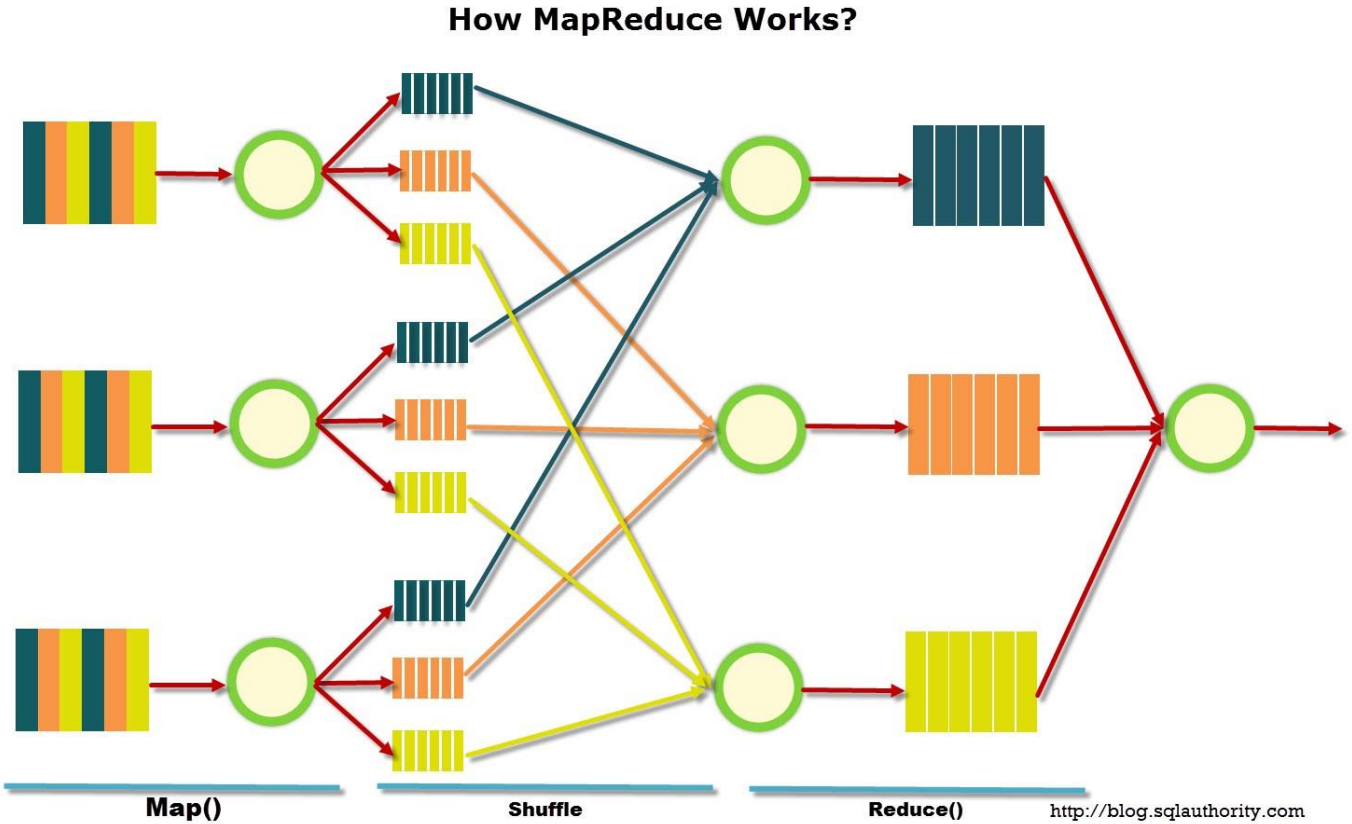
Prof. Faber Henrique

MapReduce

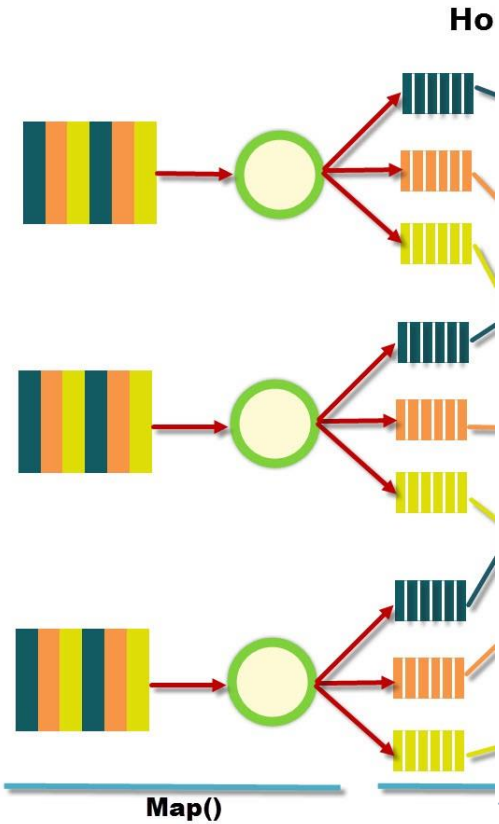
MapReduce


VOCATION BREWERY
ESTD. MMXV
DIVIDE
& CONQUER
HEBDEN
BRIDGE
ROASTED
VELVET
BLACK PALE ALE: 6.5%

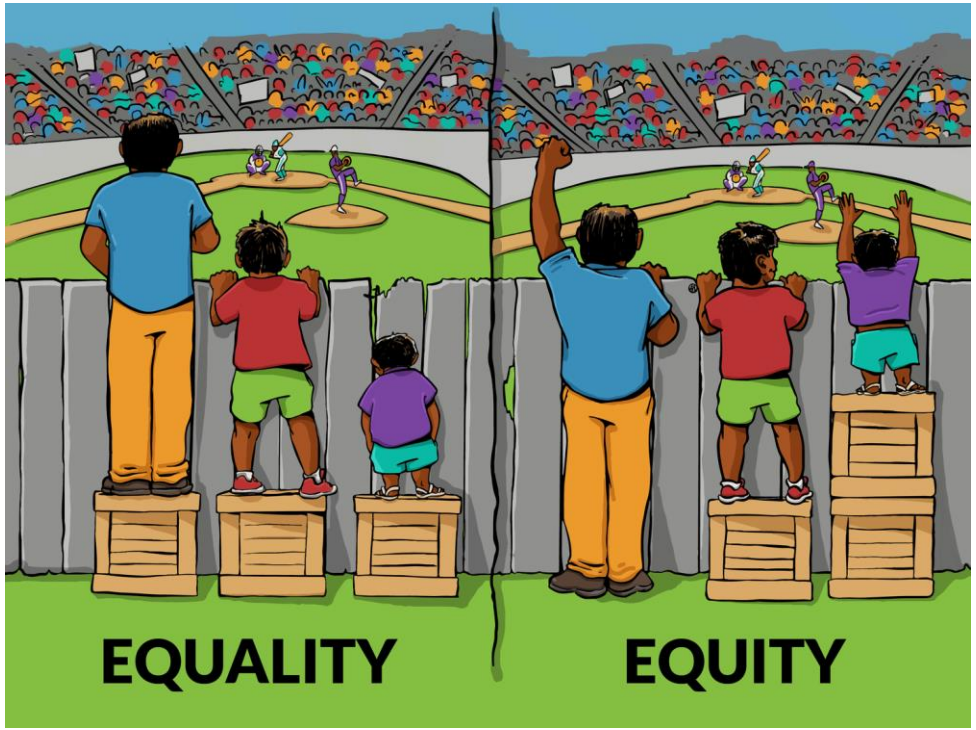
MapReduce


How MapReduce Works?
Map()
Shuffle
Reduce()
http://blog.sqlauthority.com

MapReduce



MapReduce



MapReduce

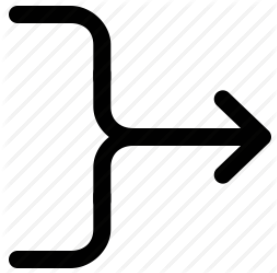
$$\forall < key, valor > \rightarrow \begin{cases} Sub(< key, valor >) \\ Sub(< key, valor >) \end{cases}$$

MAP

MapReduce

$$Sub(< key, valor >)$$

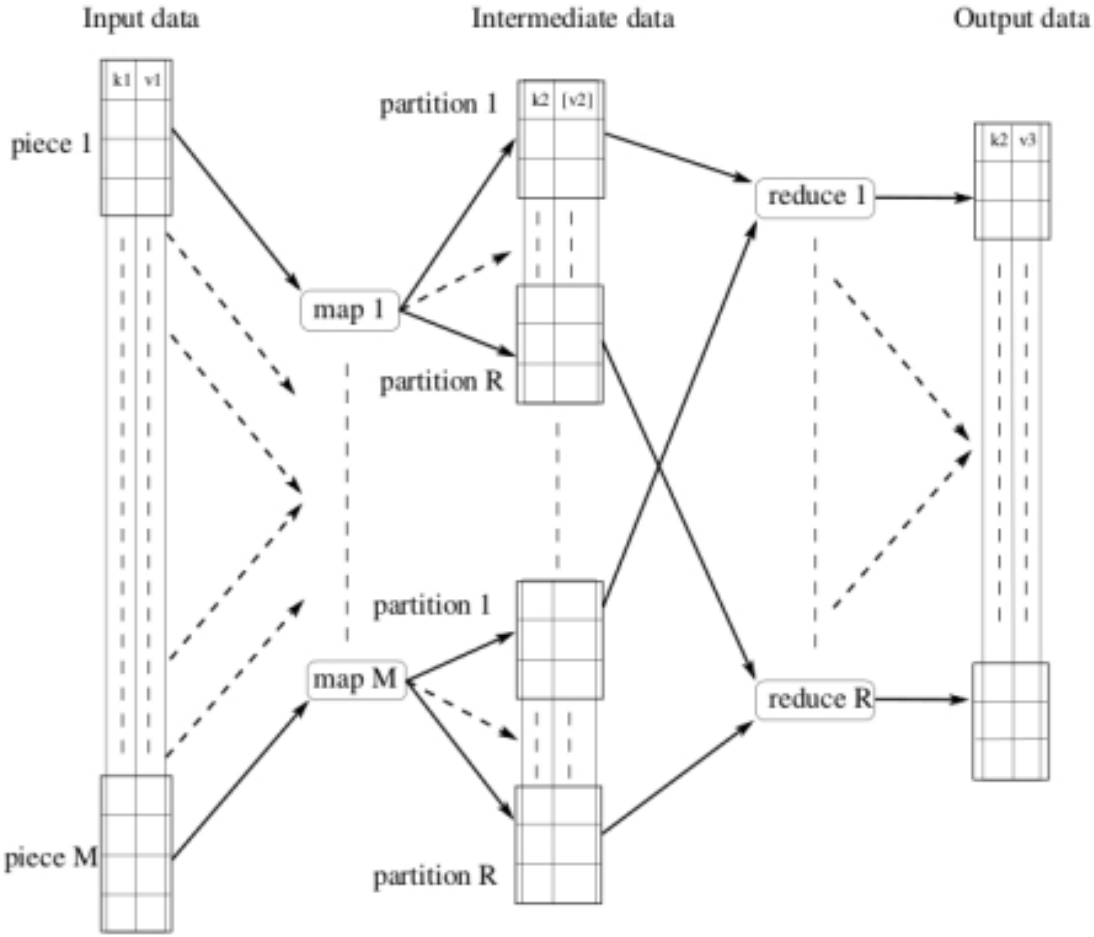
$$Sub(< key, valor >)$$



$$< key, valor >$$

REDUCE

MapReduce


Input data
Intermediate data
Output data
k1
v1
piece 1
piece M
map 1
map M
partition 1
partition R
partition 1
partition R
k2
[v2]
reduce 1
reduce R
k2
v3

MapReduce

Algoritmo 1 Map

```
Require: String key: nome do documento
         String value: conteúdo do documento
  for each word w in value do
    emitIntermediate(w, "1");
  end for
```

MapReduce

Algoritmo 2 Reduce

```
Require: String key: uma palavra
         Iterator value: uma lista de contadores
  int result = 0;
  for each v in value do
    result += parseInt(v);
  end for
  emit(key, asString(result));
```

MapReduce

1. Particionar os dados

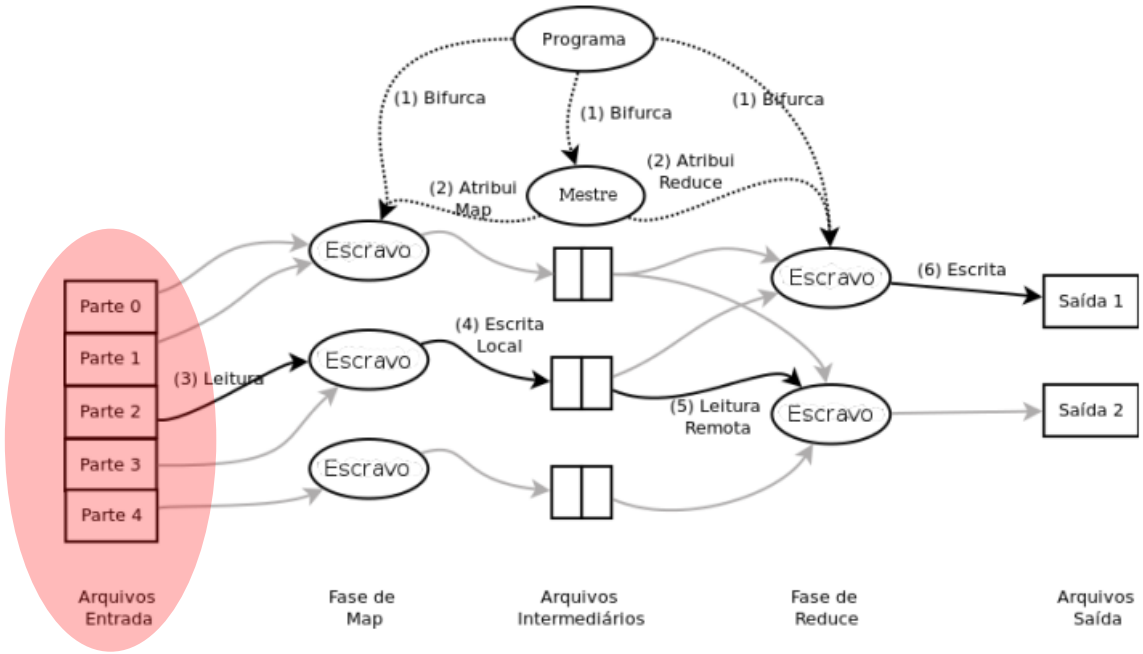


Figura 2. Fluxo geral do MapReduce

MapReduce

2. Definir as tarefas

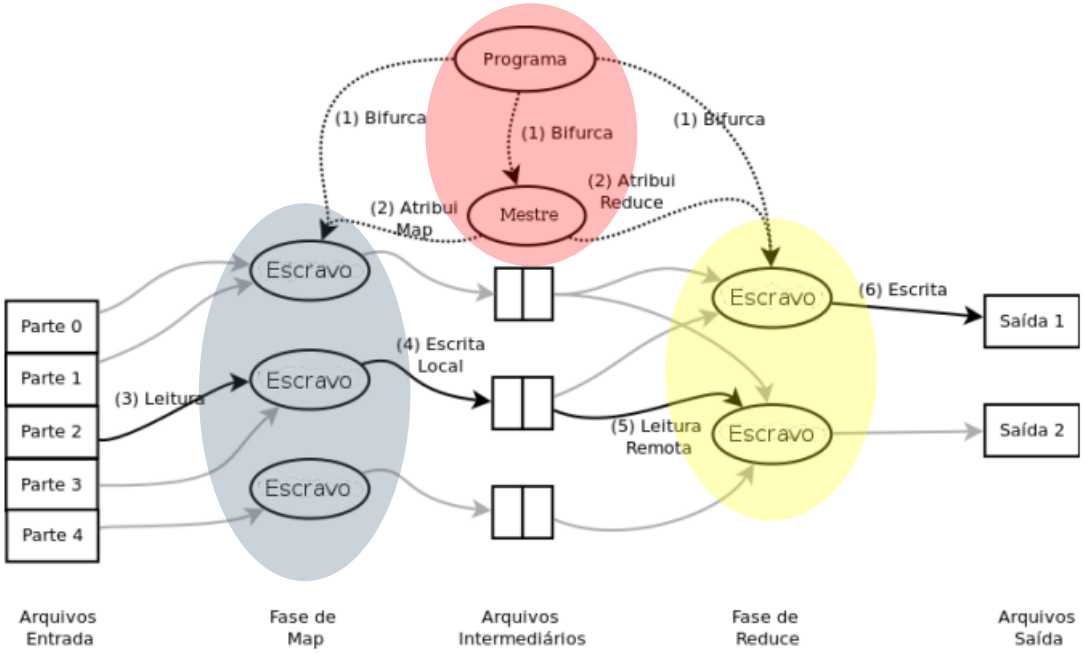


Figura 2. Fluxo geral do MapReduce

MapReduce

3. Leitura dos arquivos

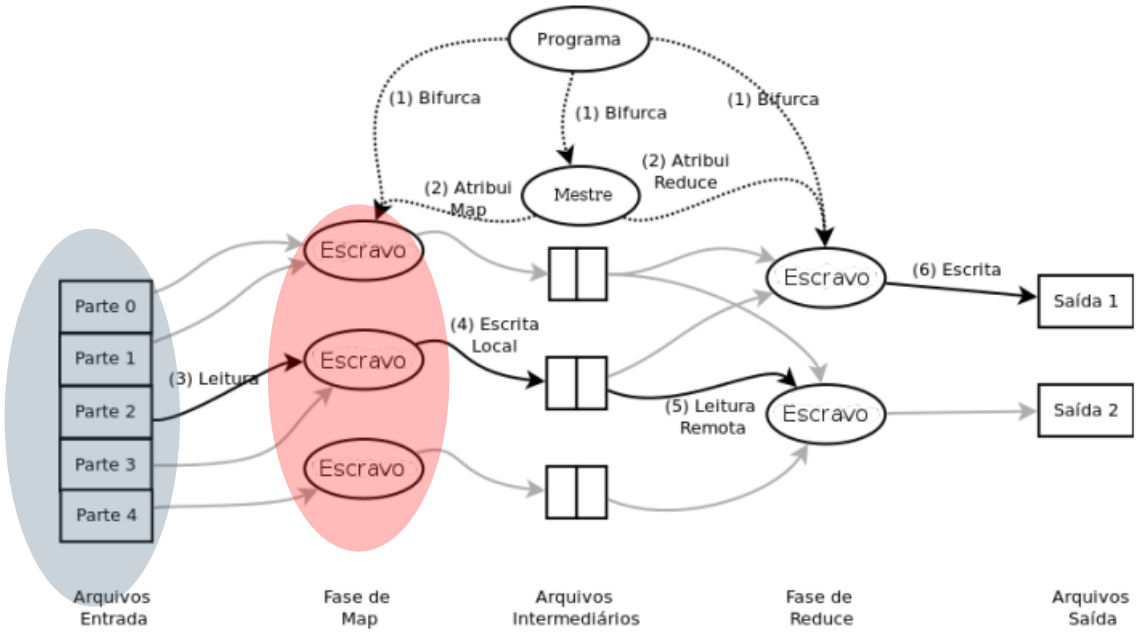


Figura 2. Fluxo geral do MapReduce

MapReduce

4. Criar região temporária

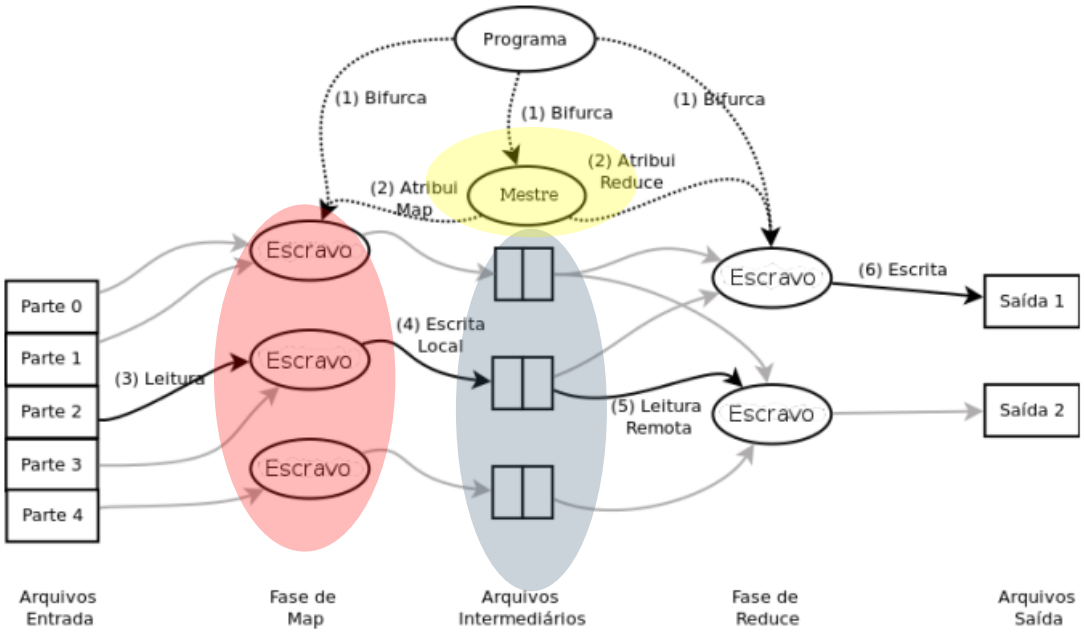


Figura 2. Fluxo geral do MapReduce

MapReduce

4/2. Notificar reduces

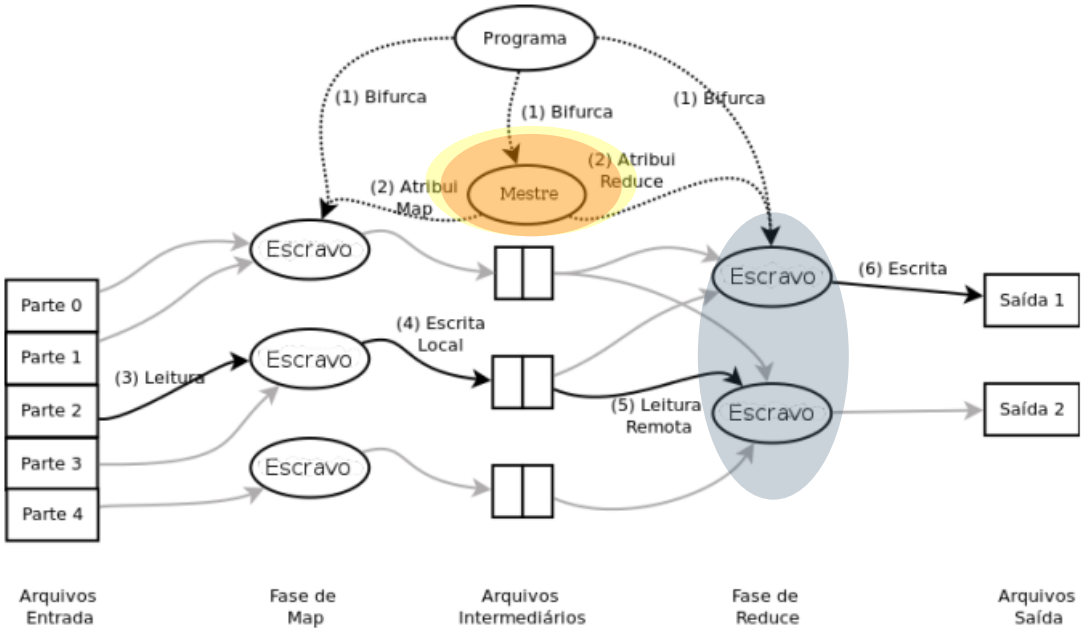


Figura 2. Fluxo geral do MapReduce

MapReduce

5. Ler e ordenar pelas keys

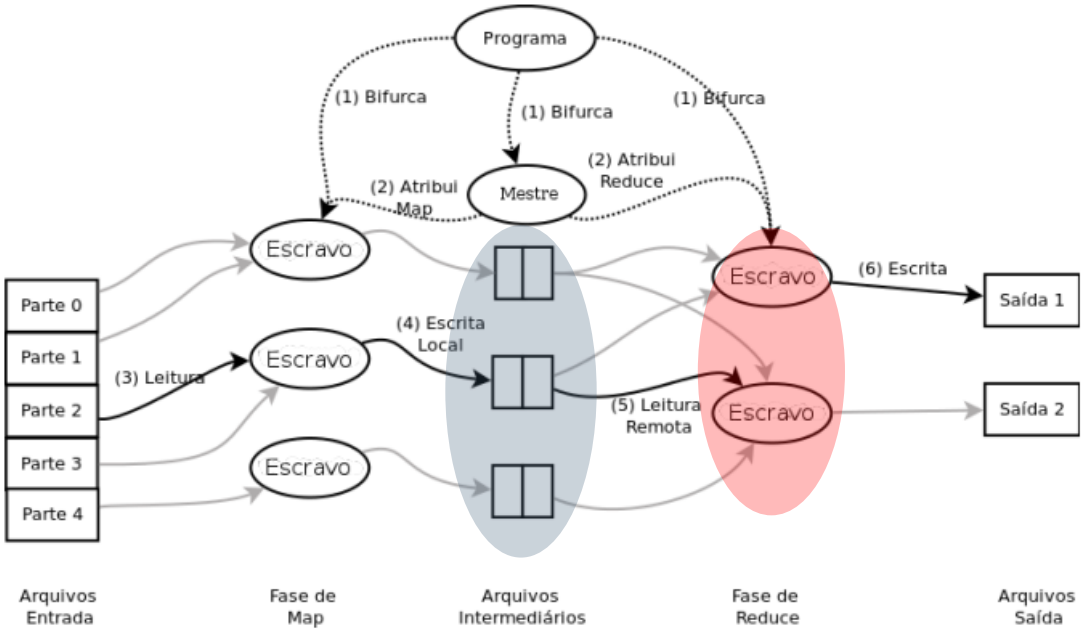


Figura 2. Fluxo geral do MapReduce

MapReduce

6. Agregar e preparar resultado (**Função de redução**)

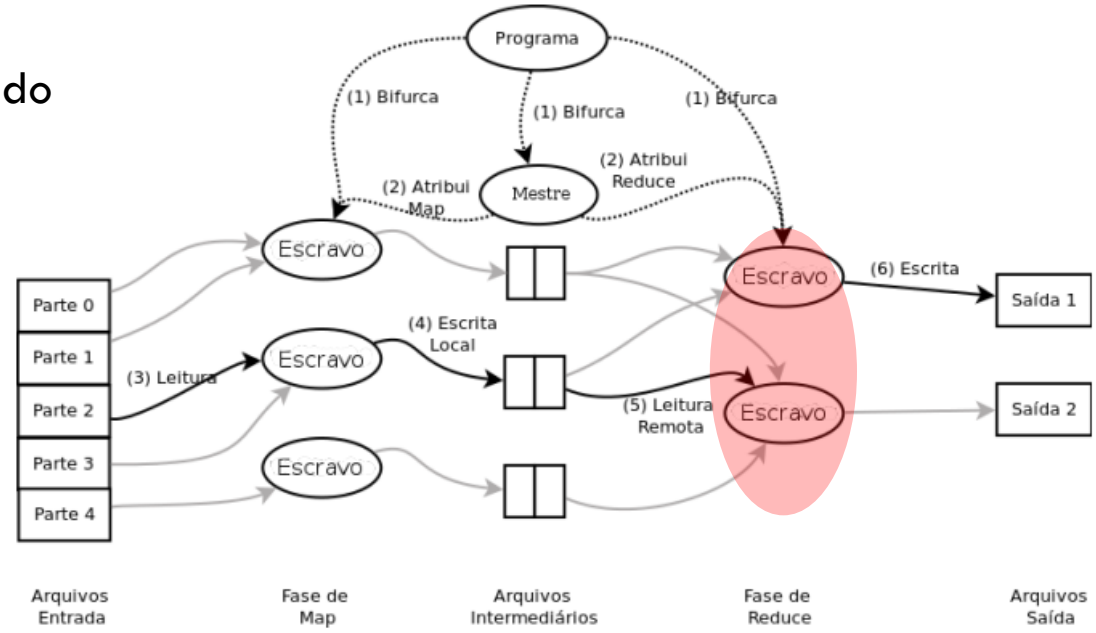


Figura 2. Fluxo geral do MapReduce

MapReduce

6. Retornar ao usuário

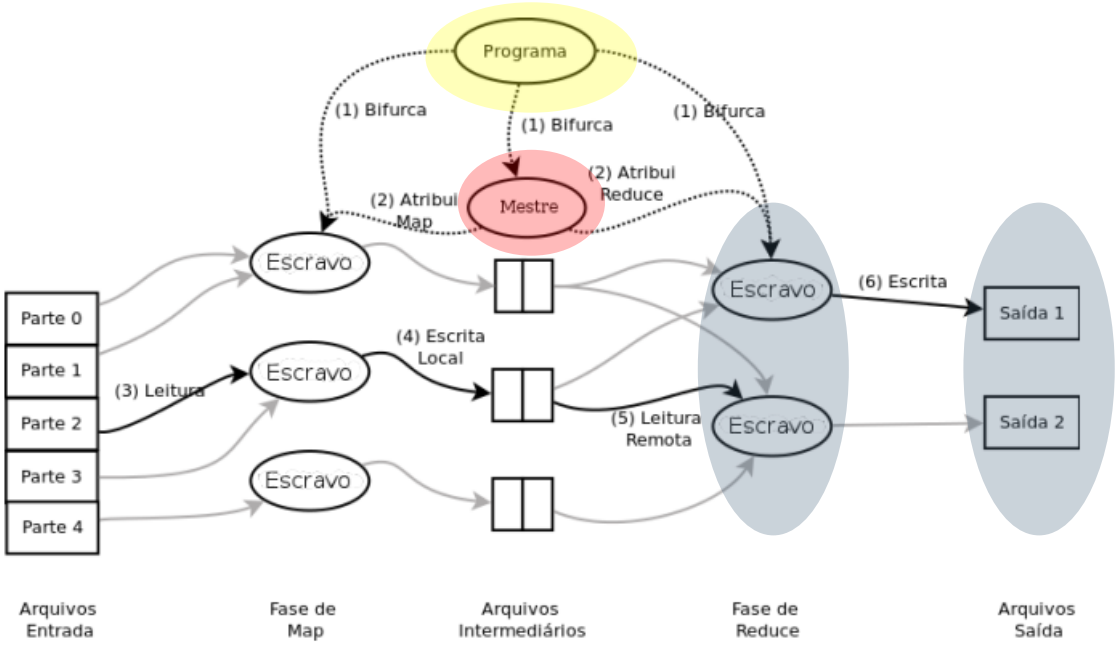


Figura 2. Fluxo geral do MapReduce

MapReduce

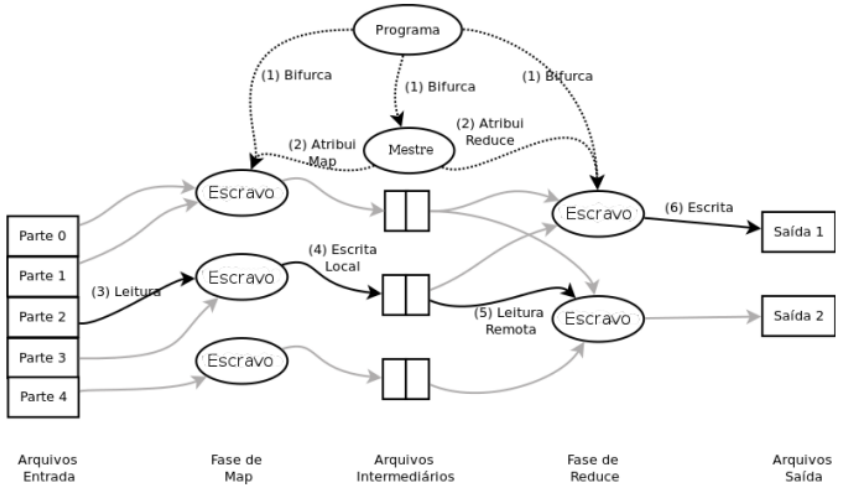
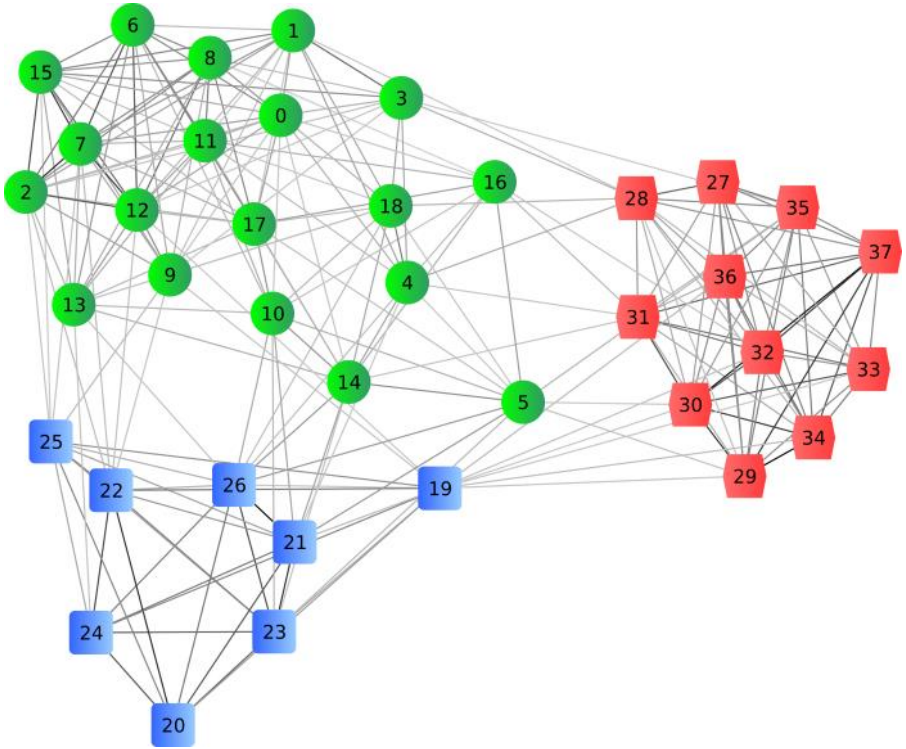


Figura 2. Fluxo geral do MapReduce

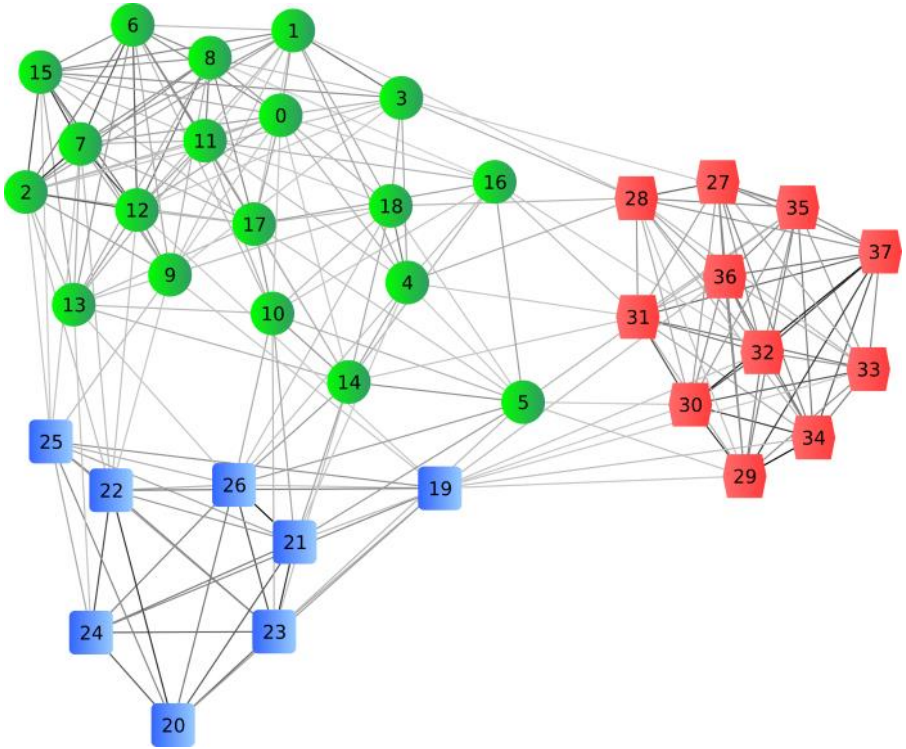
MapReduce

Node fail



MapReduce

1. Pings
2. Caso nó seja perdido
    1. Tarefas são resetadas
    2. Rescalonamento



Node Master fail

- Salve instancias periodicamente

OBRIGADO.