

MapReduce

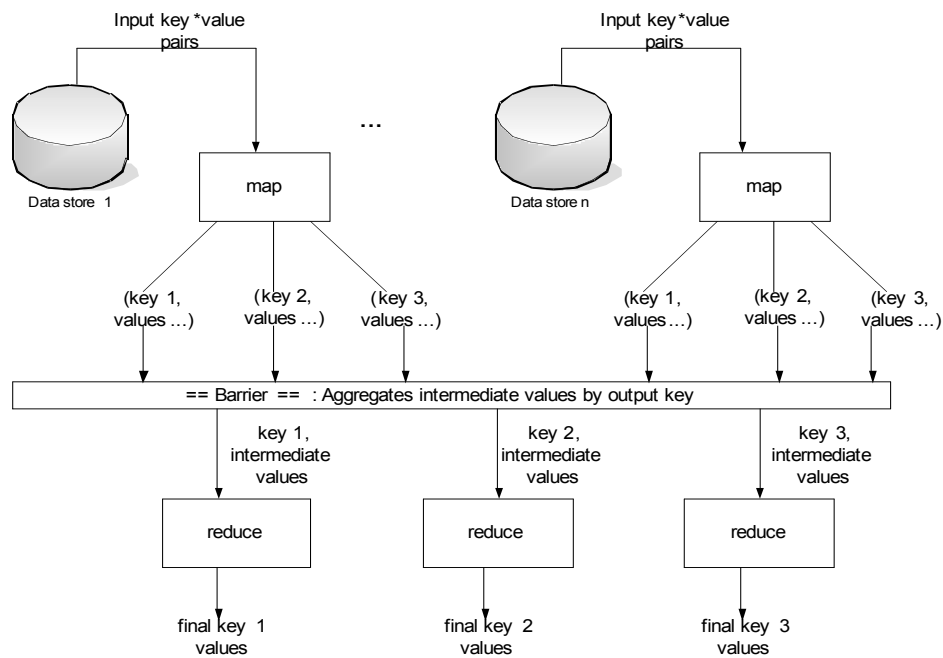


Programming Model

```
map (in_key, in_value) ->  
    (out_key, intermediate_value) list
```

```
reduce (out_key, intermediate_value list) ->  
    out_value list
```

MapReduce Control Flow

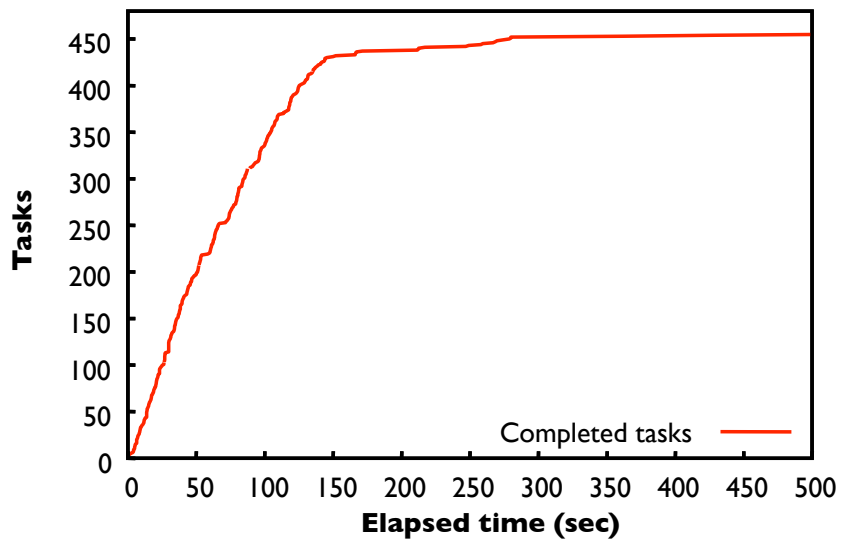


Word Count Example

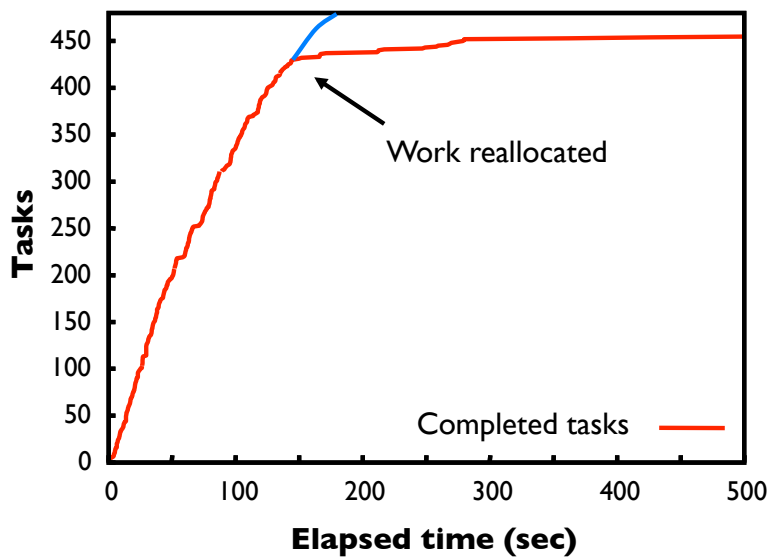
```
map(String input_key, String input_value):
    // input_key: document name
    // input_value: document contents
    for each word w in input_value:
        EmitIntermediate(w, "1");

reduce(String output_key, Iterator
        intermediate_values):
    // output_key: a word
    // output_values: a list of counts
    int result = 0;
    for each v in intermediate_values:
        result += ParseInt(v);
    Emit(AsString(result));
```

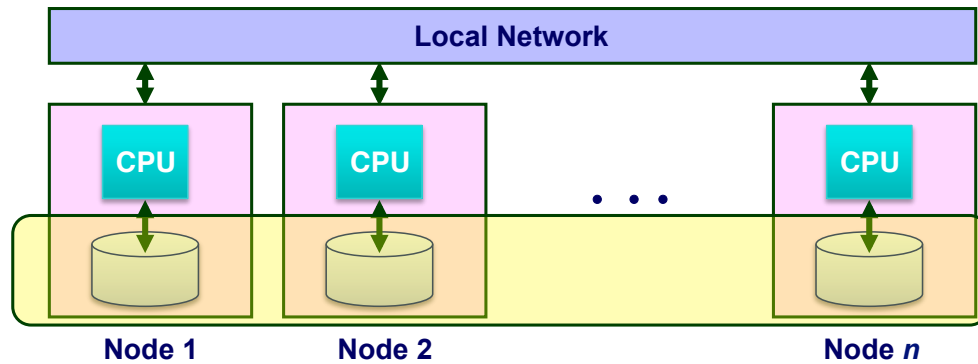
Stragglers



Stragglers



Apache Hadoop

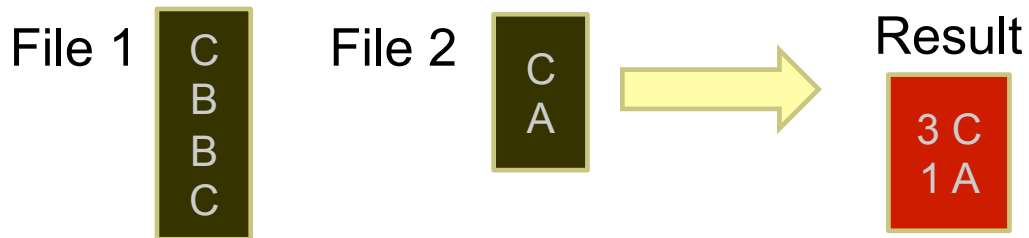


Programming Model

```
map (in_key, in_value) ->  
    (out_key, intermediate_value) list
```

```
reduce (out_key, intermediate_value list) ->  
    out_value list
```

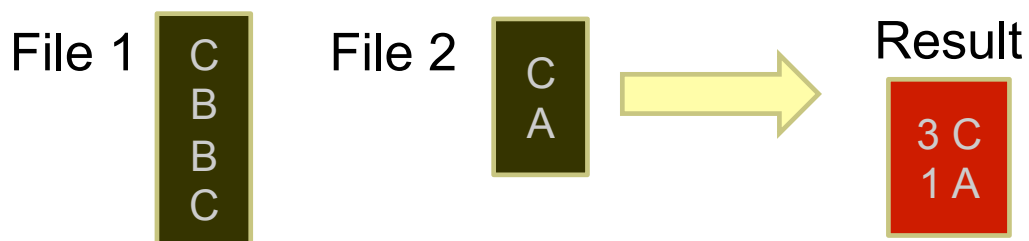
Web Log Parsing



“Count the number of times pages matching **A|C** were fetched”

(e.g., or matching “bowdoin.edu”)

Web Log Parsing



Map tasks:

(f1, C) -> [(C, 1)]

(f1, B) -> []

(f1, B) -> []

(f1, C) -> [(C, 1)]

(f2, C) -> [(C, 1)]

(f2, A) -> [(A, 1)]

Reduce tasks:

(A, [1]) -> (A, 1)

(C, [1, 1, 1]) -> (C, 3)