MapReduce



Programming Model

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

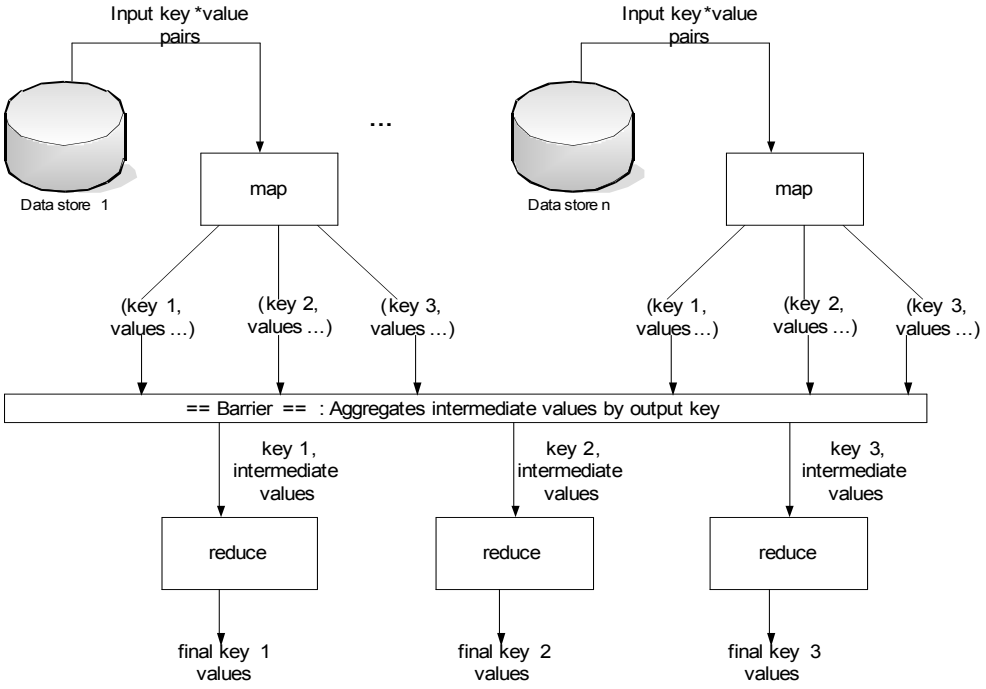

reduce (out_key, intermediate_value list) ->
     out_value list
```

Word Count Example

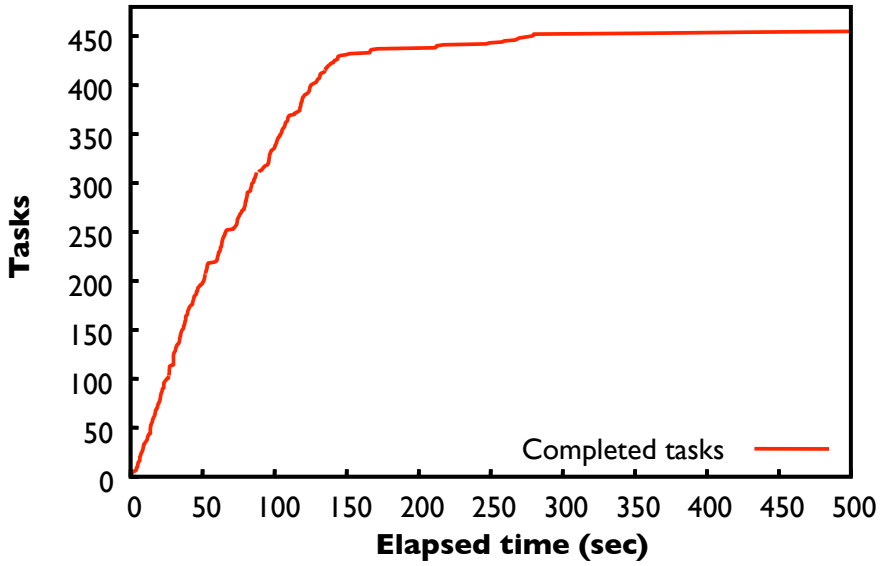
```
map(String input_key, String input_value):
  // input_key: document name
  // input_value: document contents
  for each word w in input_value:
    EmitIntermediate(w, "1");

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));
```

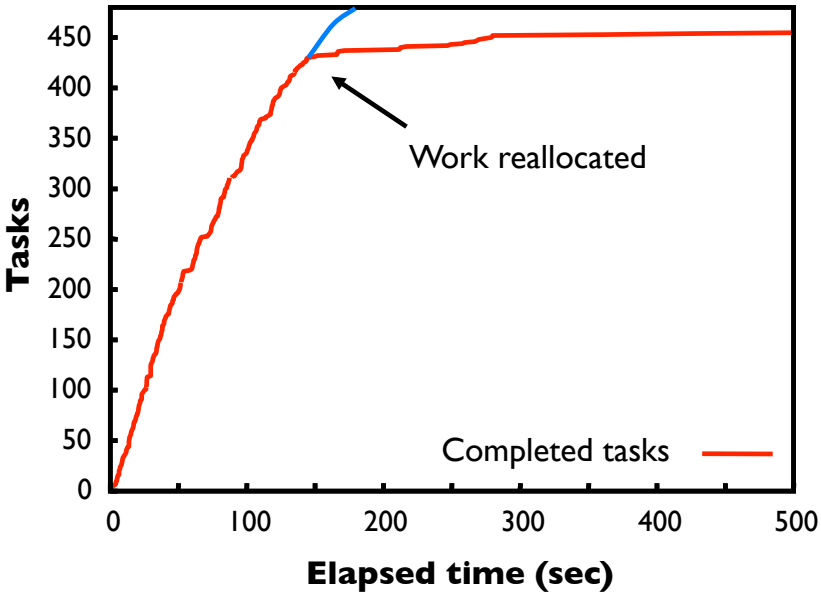

MapReduce Data Flow



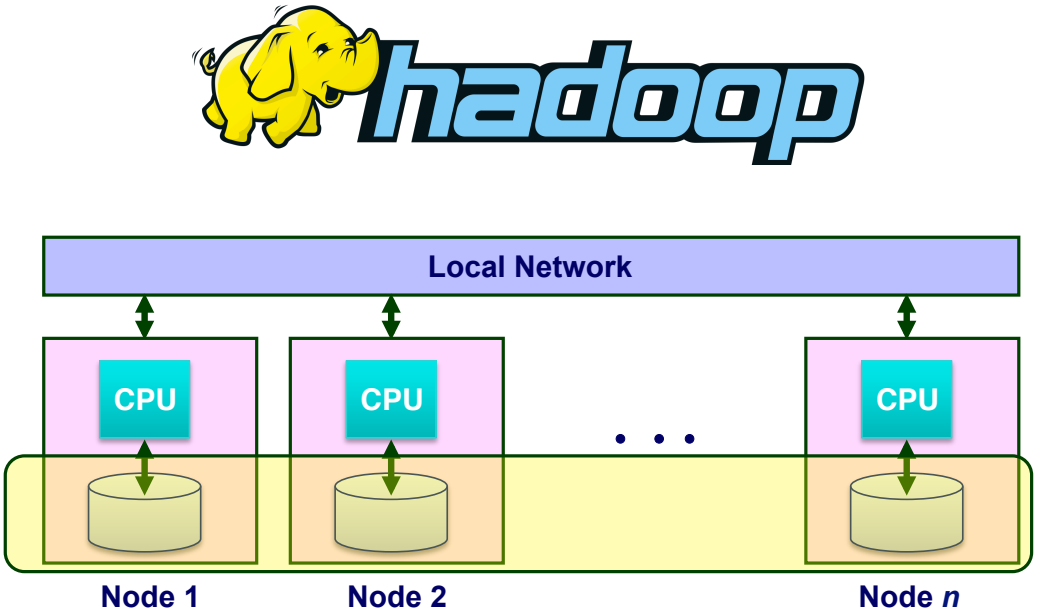
Stragglers



Reallocating Tasks



Apache Hadoop

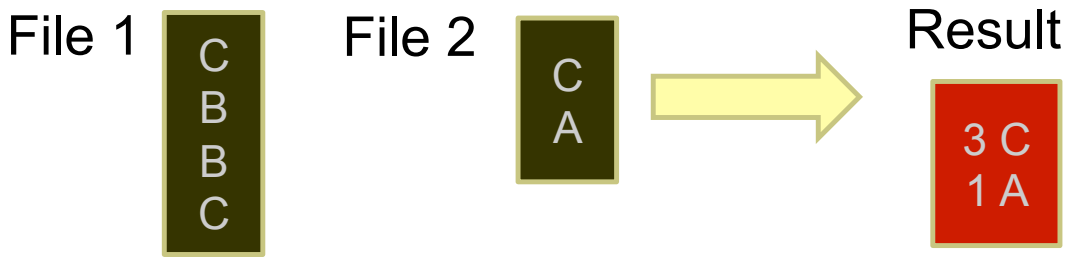


Programming Model

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

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

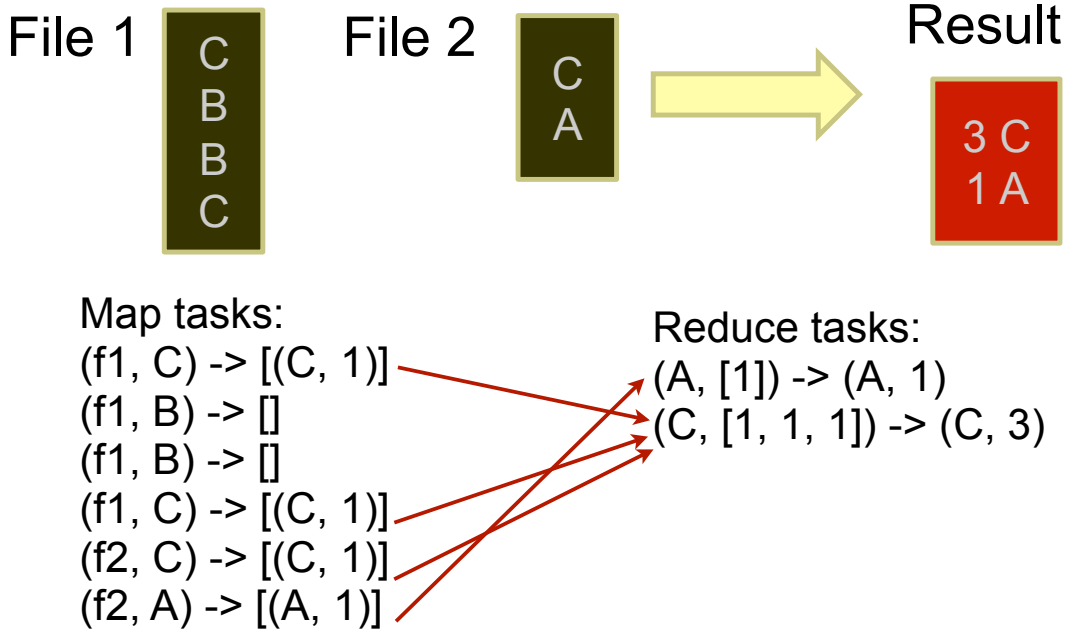
Web Log Parsing



Example: "Count the # of times pages A or C were fetched"

(or any kind of search; e.g., pages matching "bowdoin.edu")

Web Log Parsing



Web Link Analysis

