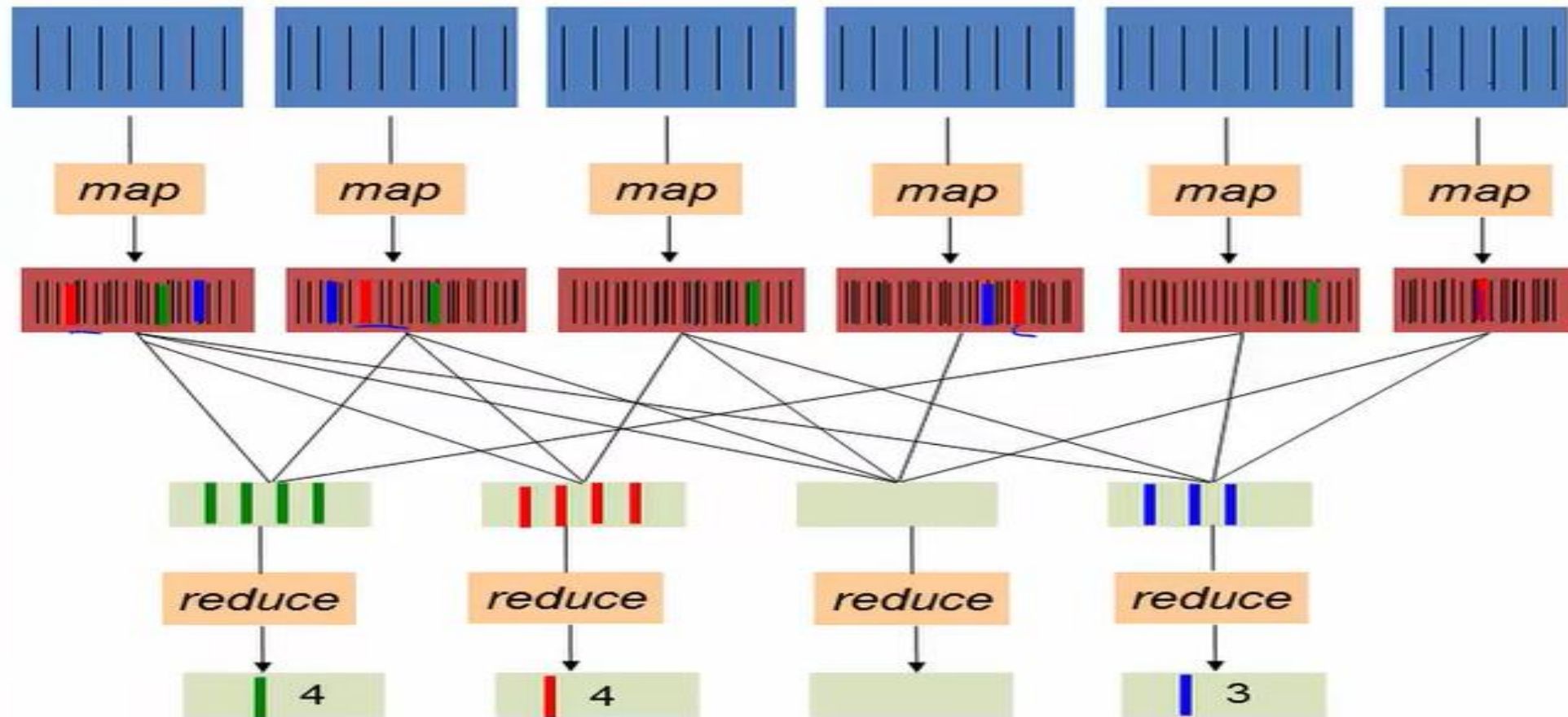


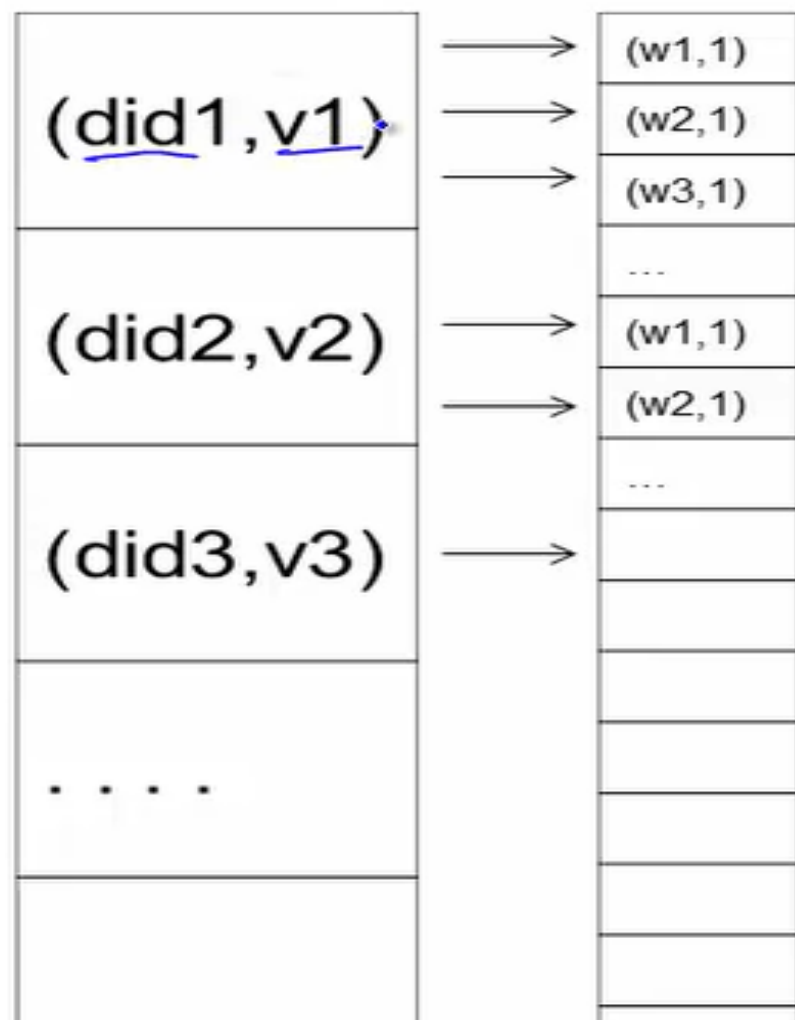
Map reduce

# MapReduce Abstraction

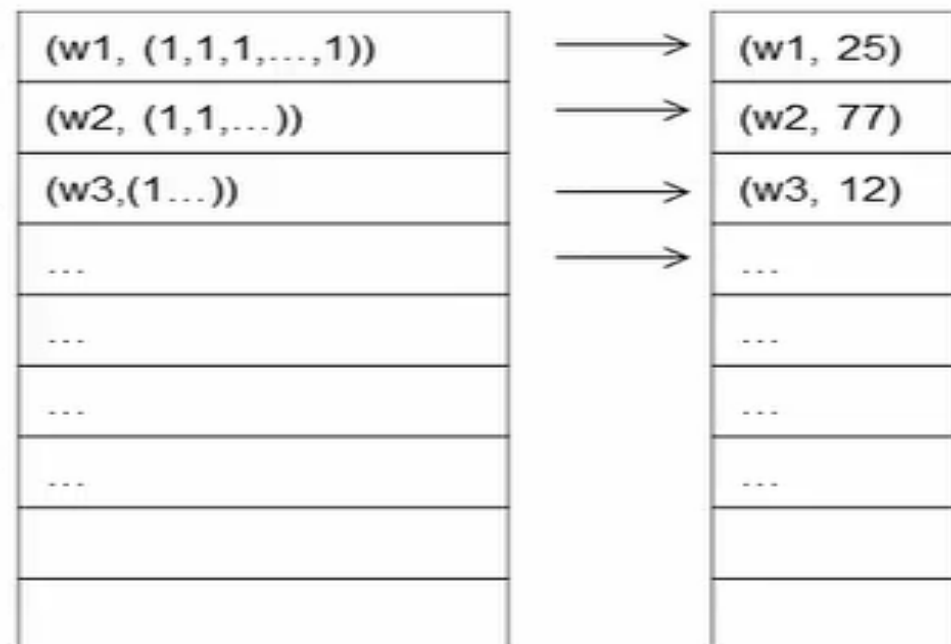
Count word occurrences across all documents



## MAP



Shuffle



## REDUCE

# Map Reduce

- Google: paper published 2004
- Free variant: Hadoop
- Map-reduce = high-level programming model and implementation for large-scale parallel data processing

# Data Model

## Files !

A file = a bag of **(key, value)** pairs

A map-reduce program:

- Input: a bag of **(inputkey, value)** pairs
- Output: a bag of **(outputkey, value)** pairs

## Step 1: the **MAP** Phase

User provides the **MAP**-function:

- Input: (input key, value)
- Output:  
bag of (intermediate key, value)

System applies the map function in parallel to all (input key, value) pairs in the input file



## Step 2: the REDUCE Phase

User provides the REDUCE function:

- Input:  
(intermediate key, bag of values)
- Output: bag of output (values)

The system will group all pairs with the same intermediate key, and passes the bag of values to the REDUCE function

# MapReduce Programming Model

- Input & Output: each a set of key/value pairs
- Programmer specifies two functions:

**map (in\_key, in\_value) -> list(out\_key, intermediate\_value)**

- Processes input key/value pair
- Produces set of intermediate pairs

**reduce (out\_key, list(intermediate\_value)) -> list(out\_value)**

- Combines all intermediate values for a particular key
- Produces a set of merged output values (usually just one)

*Inspired by primitives from functional programming languages such as Lisp, Scheme, and Haskell*



## Example: What does this do?

**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 intermediate\_key, Iterator intermediate\_values):**

// intermediate\_key: word

// intermediate\_values: ????

**int result = 0;**

**for each v in intermediate\_values:**

**result += v;**

**EmitFinal(intermediate\_key, result);**

## Example: What does this do?

```
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);
```

("history", 5)

1 or 5

---

```
reduce(String intermediate_key, Iterator intermediate_values):
```

```
// intermediate_key: word
```

```
// intermediate_values: ????
```

```
int result = 0;
```

```
for each v in intermediate_values:
```

```
    result += v;
```

```
EmitFinal(intermediate_key, result);
```

(history, 25)



# MapReduce Example: Word Length Histogram

## Example: Word length histogram

### Abridged Declaration of Independence

A Declaration By the Representatives of the United States of America, in General Congress Assembled. When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident; that all men are created equal and independent; that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying it's foundation on such principles and organizing it's power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will dictate that governments long established should not be changed for light and transient causes; and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies, and such is now the necessity which constrains them to expunge their former systems of government, the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unswerving by falsehood.

How many "big", "medium", and "small" words are used?

## MapReduce Example: Word Length Histogram

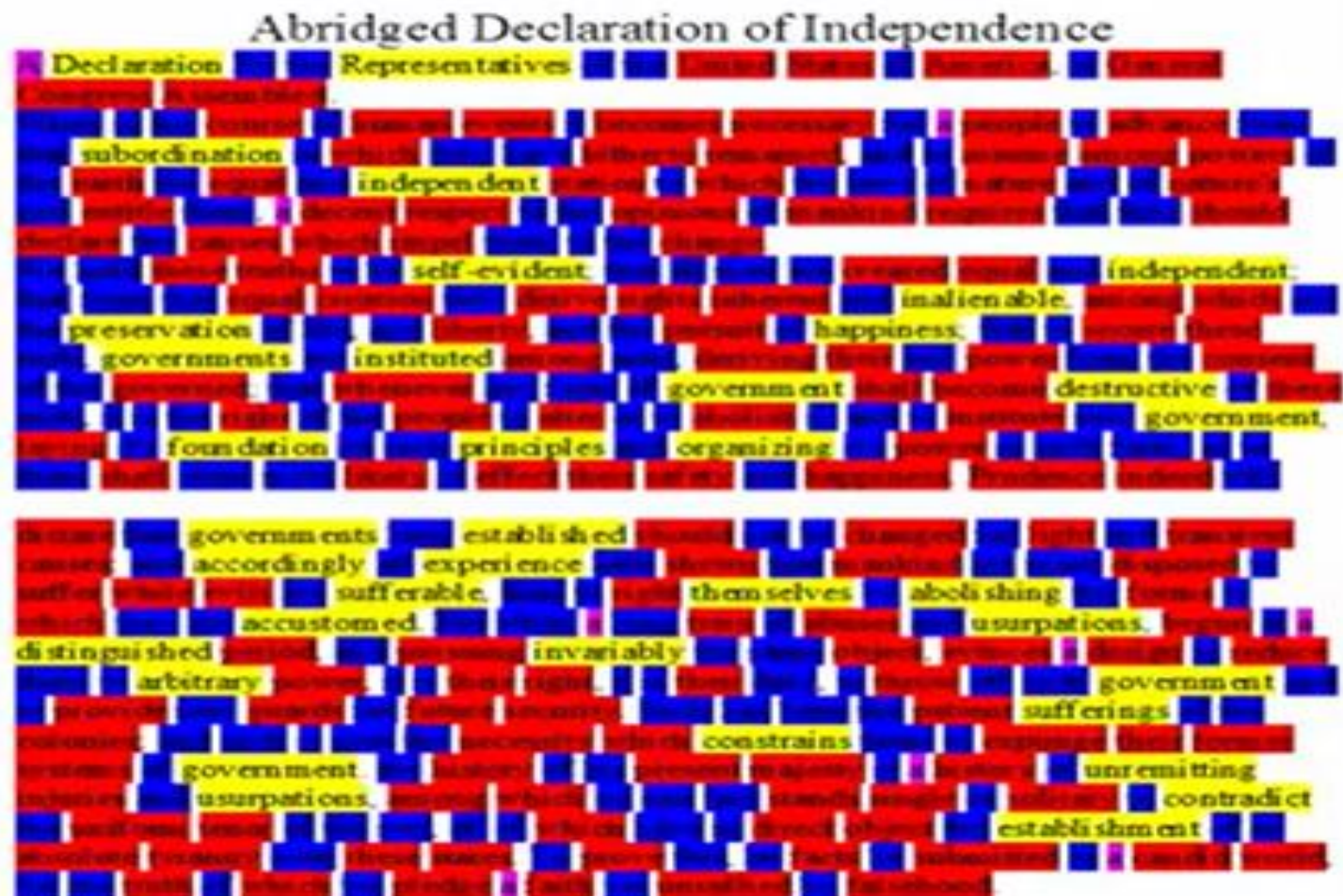
## Example: Word length histogram

Big = Yellow = 10+ letters

Medium = Red = 5..9 letters

Small = Blue = 2..4 letters

Tiny = Pink = 1 letter





# MapReduce Example: Word Length Histogram

## Example: Word length histogram

Split the document into chunks and process each chunk on a different computer

Chunk 1

Chunk 2

### Abridged Declaration of Independence

A Declaration By the Representatives of the United States of America, in General Congress Assembled.

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident; that all men are created equal and independent; that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying its foundation on such principles and organizing its power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will

dictate that governments long established should not be changed for light and transient causes: and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies; and such is now the necessity which constrains them to expunge their former systems of government. the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unsullied by falsehood.

# Example: Word length histogram

Map Task 1  
(204 words)

Abridged Declaration of Independence

Declaration Representatives United States American Congress  
Assembled constitute a united front becomes necessary a people advance  
subordination which is to be maintained and to extend its power  
independent states which are to be maintained and to extend its power  
declare the causes which impel them to the change  
self-evident that all men are created equal independent;  
equal rights derive rights inherent inalienable among which  
preservation of life liberty and pursuit of happiness; secure these  
governments instituted among men deriving their just powers from the consent  
of the governed; whenever a government becomes destructive these  
rights of the people shall be restored to secure them government  
foundation principles organizing a new government  
that they shall effect their safety and happiness. Prudence indeed

(key, value)

(yellow, 17)

(red, 77)

(blue, 107)

(pink, 3)

dictate governments established should changed in light of  
accordingly experience and should be disposed  
sufferable; and to abolish themselves abolishing  
which are accustomed to abuse of power usurpations; because  
distinguished period; for some invariably to be abolished  
arbitrary power; that right; that they shall throw government  
provide for the future security of the people sufferings  
colours; and to be necessary to be constrained to expose their names  
government; history of the present is a history of unrelenting  
usurpations; among which is a history of unrelenting  
unrelenting tyrannies; these tyrannies are a history of unrelenting  
unrelenting tyrannies; these tyrannies are a history of unrelenting



# Example: Word length histogram

“Shuffle step”

## Map task 1

A Declaration By the Representatives of the United States of America, in General Congress Assembled  
When in the course of human events it becomes necessary for a people to assume, from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.  
We hold these truths to be self-evident, that all men are created equal, and independent, that from that equal station they derive rights, which no man has a right to tread upon. Among these are the preservation of life, and liberty; and the pursuit of happiness. That to secure these ends, governments are instituted among men, deriving their just powers from the consent of the governed, that whenever any form of government becomes destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying its foundation on such principles and organizing its powers in such form, so that their safety and their happiness be best secured. Prudence would not

(yellow, 17)  
(red, 77)  
(blue, 107)  
(pink, 3)

## Map task 2

discontented governments long established should not be changed for light and transient causes; and accordingly all experience hath shown that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to absolute power, it is their right, it is their duty, to throw off such government, and to provide new guards for their future security. Truth has been the power of the world, and such is now the common sense of mankind that to enquire their former systems of government, the history of this present age is a history of searching inquiry and usurpations, among which no one form is single or solitary to condemn the wisdom of the rest, all of which have in direct object the establishment of an absolute tyrannical power. It is now time, therefore, to submit to a candid world, for the rule of which we pledge a faith yet untried by blood.

(yellow, 20)  
(red, 71)  
(blue, 93)  
(pink, 6)

## Reduce tasks

(yellow, 17)  
(yellow, 20) (yellow, 37)  
(red, 77)  
(red, 71) (red, 148)  
(blue, 93)  
(blue, 107) (blue, 200)  
(pink, 6)  
(pink, 3) (pink, 9)

# More Examples: Build an Inverted Index

## Input:

tweet1, ("I love pancakes for breakfast")  
tweet2, ("I dislike pancakes")  
tweet3, ("What should I eat for breakfast?")  
tweet4, ("I love to eat")

## Desired output:

"pancakes", (tweet1, tweet2)  
"breakfast", (tweet1, tweet3)  
"eat", (tweet3, tweet4)  
"love", (tweet1, tweet4)  
...



