USC Viterbi
School of Engineering

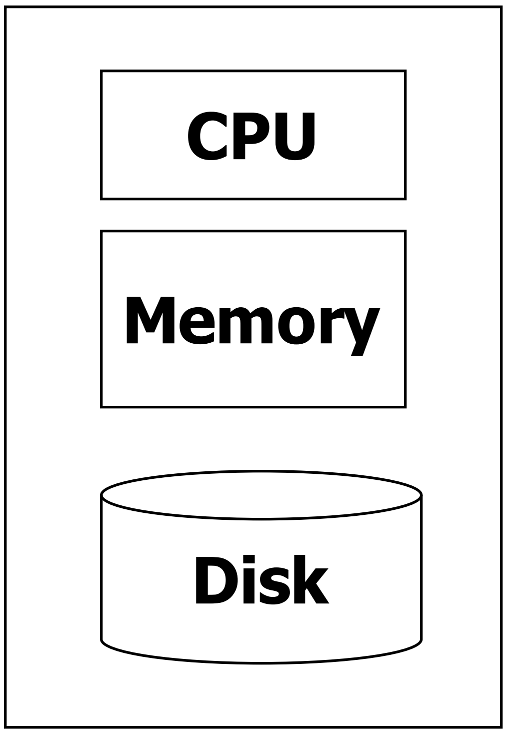
INF 553: Foundations and Applications of Data Mining

Large-Scale File Systems and Map-Reduce

Dr. Anna Farzindar
farzinda@usc.edu

Thanks for source material to: J. Leskovec, A. Rajaraman, J. Ullman: Mining of Massive Datasets

Single-node architecture



Machine Learning, Statistics

"Classical" Data Mining

MapReduce

- Much of the course will be devoted to **large scale computing** for **data mining**
- **Challenges:**
  - How to distribute computation?
  - Distributed/parallel programming is hard
- **MapReduce** addresses all of the above
  - Google's computational/data manipulation model
  - Elegant way to work with big data.

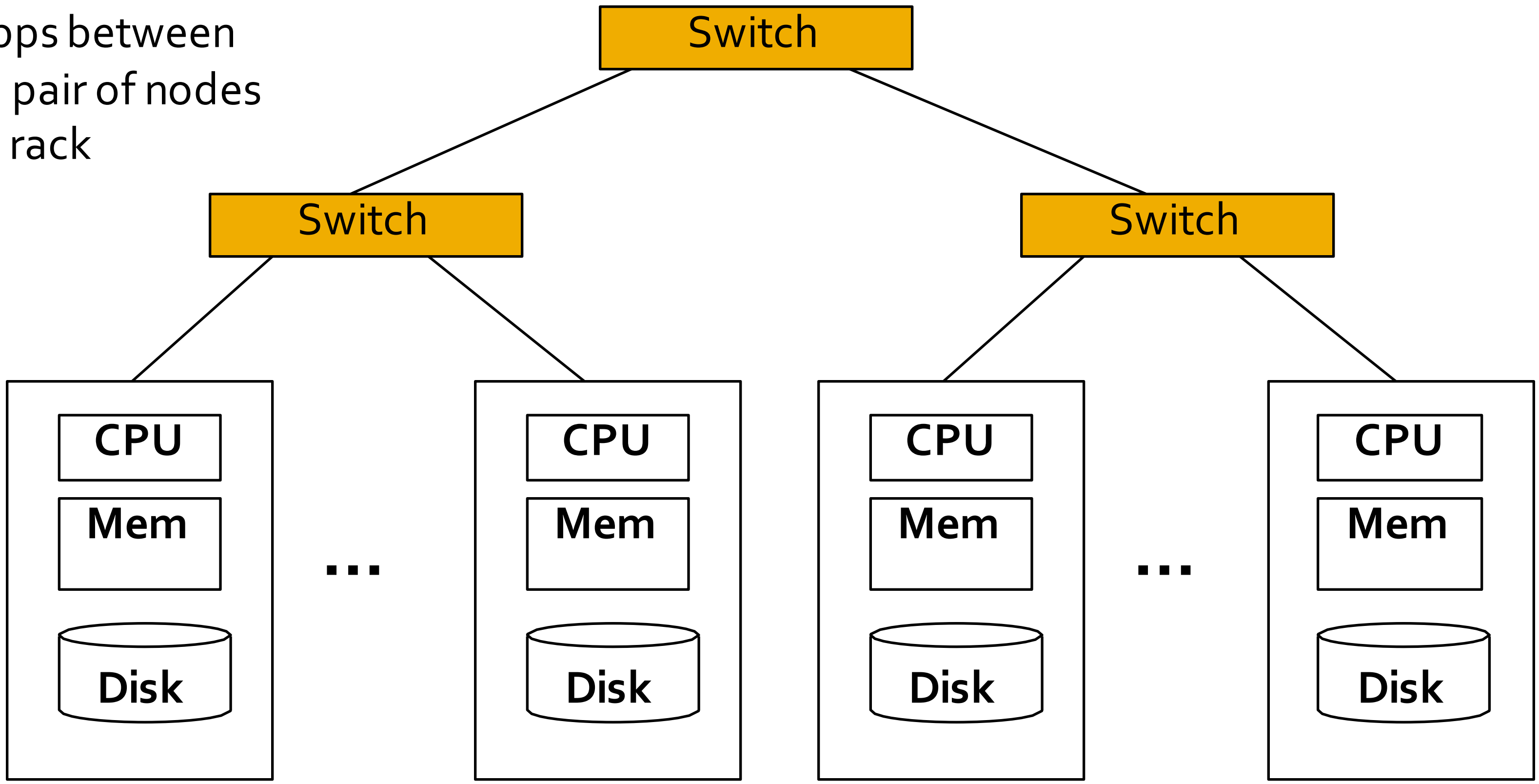
Motivation: Google Example

- 20+ billion web pages x 20KB average size of a website = 400TB
- 1 computer reads 30-35 MB/sec from disk
  - ~4 months to read the web
- ~1,000 hard drives to store the web
- Takes even more to **do** something useful with the data!
- **Recently standard architecture for such problems emerged:**
  - Cluster of commodity Linux nodes
  - Commodity network (ethernet) to connect them.

Cluster Architecture

2-10 Gbps backbone between racks

1 Gbps between any pair of nodes in a rack



Each rack contains 16-64 nodes

In 2011 it was guestimated that Google had 1M machines, http://bit.ly/Shh0RO

Large-scale Computing

- **Large-scale computing** for **data mining** problems on **commodity hardware**

- **Challenges:**
  - How do you **distribute computation?**
  - How can we make it easy to **write distributed programs?**

Cluster Computing Challenges

- **Machines fail:**
  - One server may stay up 3 years (1,000 days)
  - If you have 1,000 servers, expect to loose 1/day
  - With 1M machines 1,000 machines fail every day!
- How to **store data** persistently and keep it available if nodes can fail?
- How to deal with nodes failures during a **long running computation?**

Cluster Computing Challenges (cont.)

- Network bottleneck
  - Network bandwidth= 1 Gbps
  - Moving 10TB takes approximately 1 day

- Distributed programming is hard!
  - Need a simple model that hides most of the complexity.

Map-Reduce

- Map-Reduce addresses the challenges of cluster computing
  - Store data redundantly on multiple nodes for persistence and availability
  - Move computation close to data to minimize data movement
  - Simple programming model to hide the complexity of all this magic.

Idea and Solution

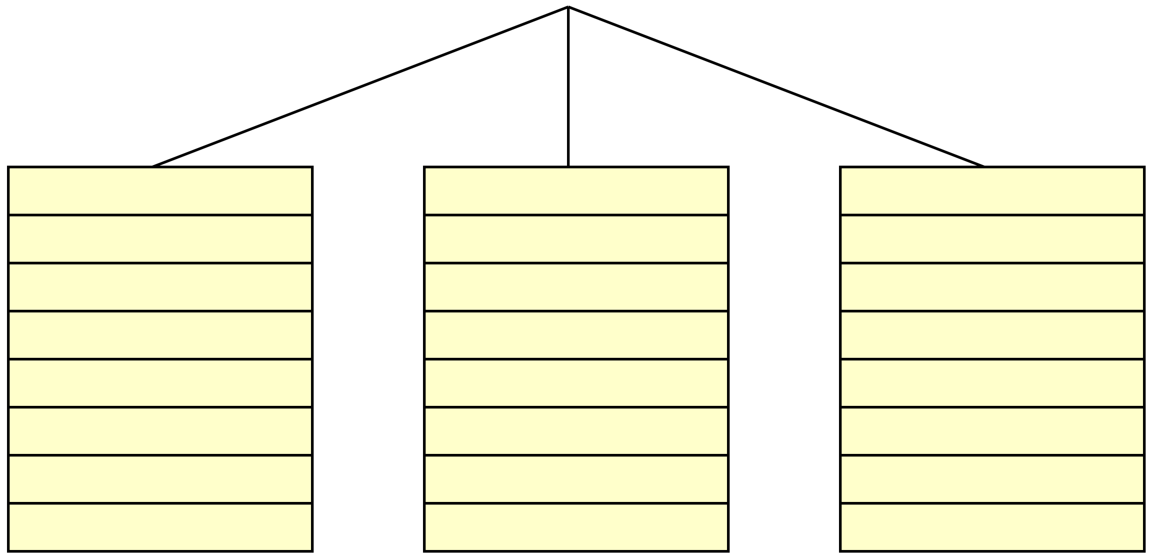
- **Issue: Copying data over a network takes time**
- **Idea:**
  - Bring **computation** to data
  - Store files multiple times for reliability
- **MapReduce** addresses these problems
  - **Storage Infrastructure – File system**
    - Google: GFS. Hadoop: HDFS
  - **Programming model**
    - MapReduce.

Storage Infrastructure

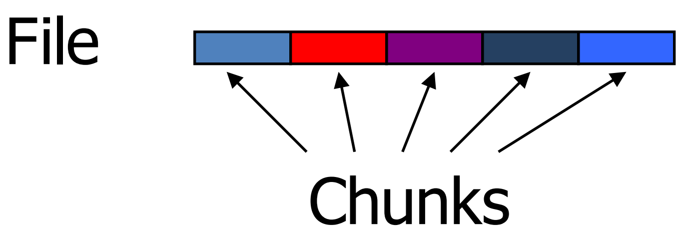
- **Problem:**
  - If nodes fail, how to store data persistently?
- **Answer:**
  - **Distributed File System:**
    - Provides global file namespace,
- **Typical usage pattern**
  - Huge files (100s of Gigabytes to Terabytes)
  - Data are rarely updated in place
  - Reads and appends are common.

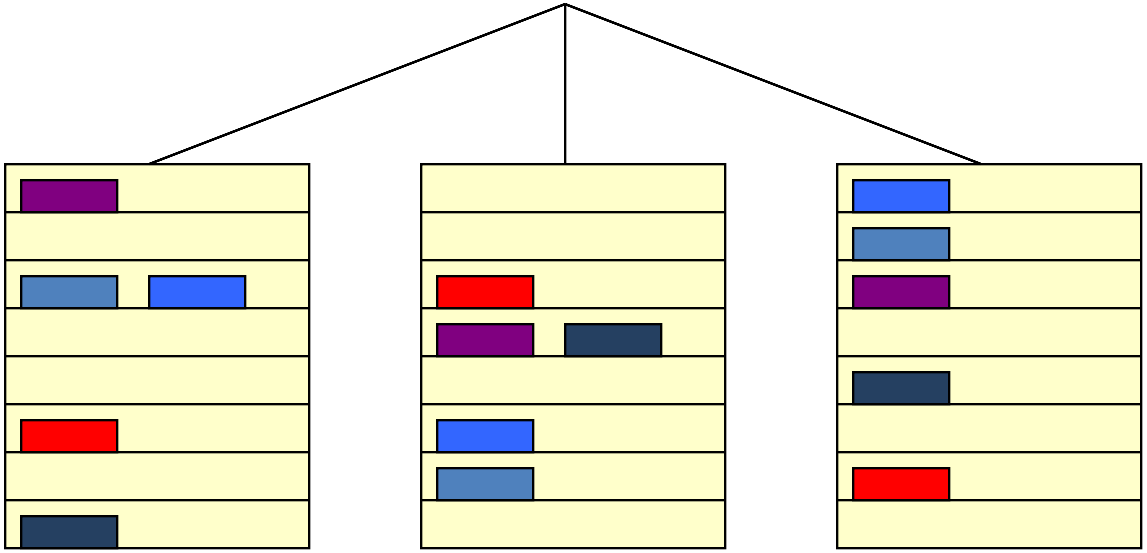
Distributed File System

- **Chunk servers**
  - File is split into contiguous ***chunks***
  - Typically each chunk is 64MB
  - Each chunk replicated (usually 2x to 3x)
  - Try to keep replicas in **different racks**,
- **Master node**
  - also known as **Name Node** in Hadoop's HDFS
  - Stores **metadata** about where files are stored
  - Master node might be **replicated**,
- **Client library for file access**
  - Talks to master to find chunk servers
  - Connects directly to chunk servers to access data.



Racks of Compute Nodes

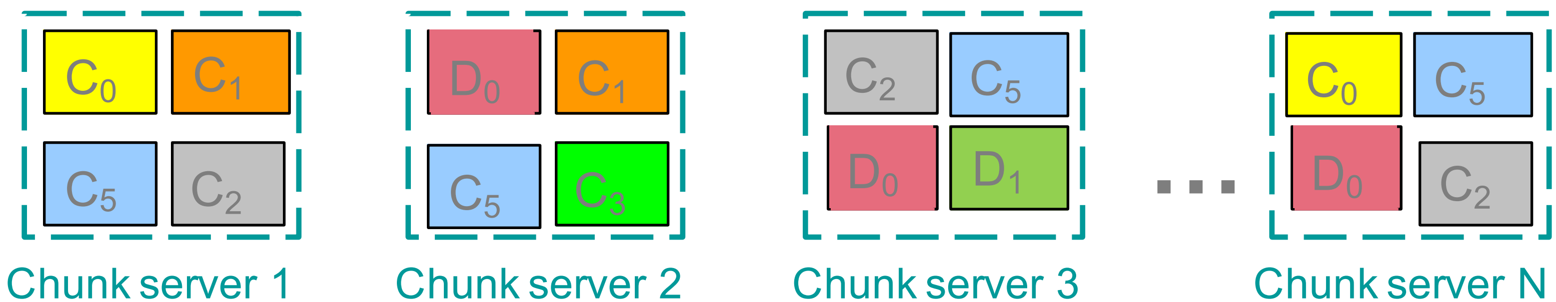




3-way replication of files, with copies on different racks.

Distributed File System

- **Reliable distributed file system**
- Data kept in “chunks” spread across machines
- Each chunk **replicated** on different machines
  - Seamless recovery from disk or machine failure



Bring computation directly to the data!

Chunk servers also serve as compute servers

Programming Model: MapReduce

Warm-up task:

- We have a huge text document
- Count the number of times each distinct word appears in the file,
- **Sample application:**
  - Analyze **web server logs** to find **popular URLs**.

Task: Word Count

Case 1:

- File too large for memory, but all <word, count> pairs fit in memory
  - Hashtable Word -> Count

Case 2:

- Even the <word, count> pairs don't fit in memory,
- Count occurrences of words:
  - **words(doc.txt) | sort | uniq -c**
    - where **words** takes a file and outputs the words in it, one per a line
- Case 2 captures the essence of **MapReduce**
  - Great thing is that it is naturally parallelizable.

MapReduce: Overview

3 steps of MapReduce

- Sequentially read a lot of data

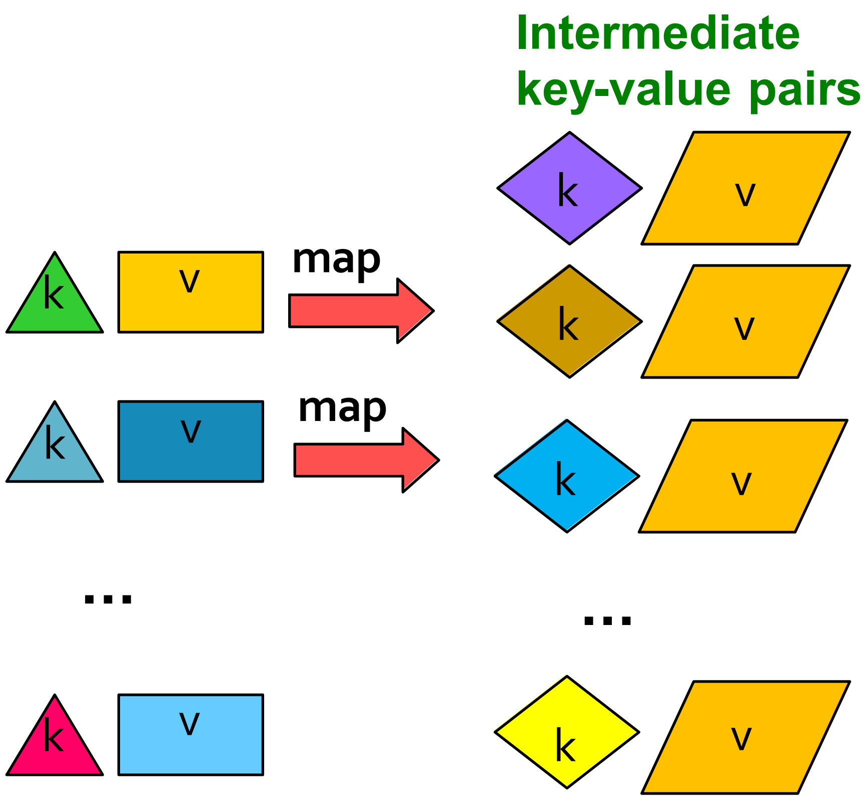
  `words(doc.txt)| sort| uniq -c`
- **Map:**
  - Extract something you care about (keys),
- **Group by key:** Sort and shuffle,
- **Reduce:**
  - Aggregate, summarize, filter or transform
- Output the result

Outline stays the same, **Map** and **Reduce** change to fit the problem

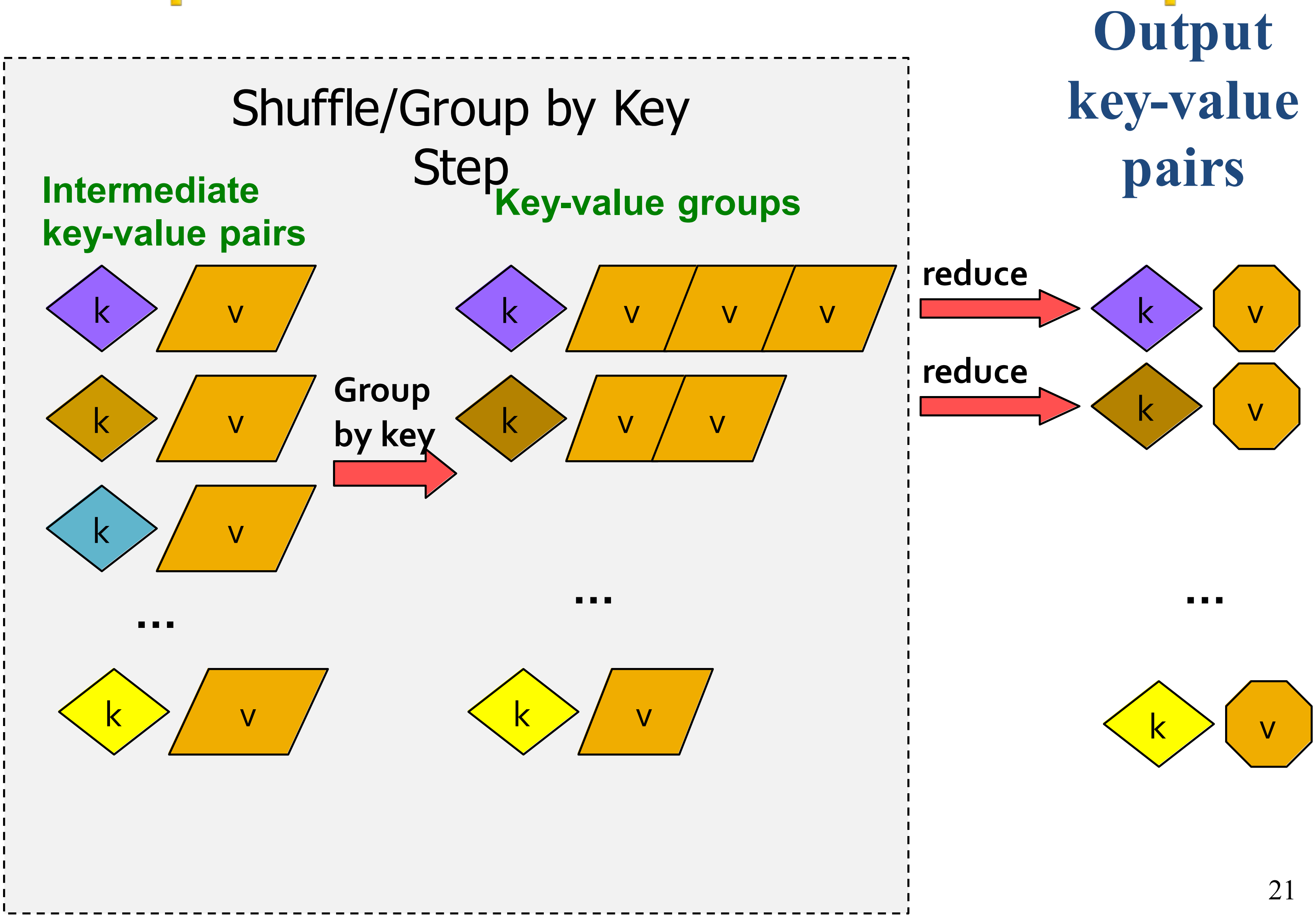
MapReduce: The Map Step

Input: key-value pairs

Output: (intermediate) key-value pairs



MapReduce: The Reduce Step



More Specifically

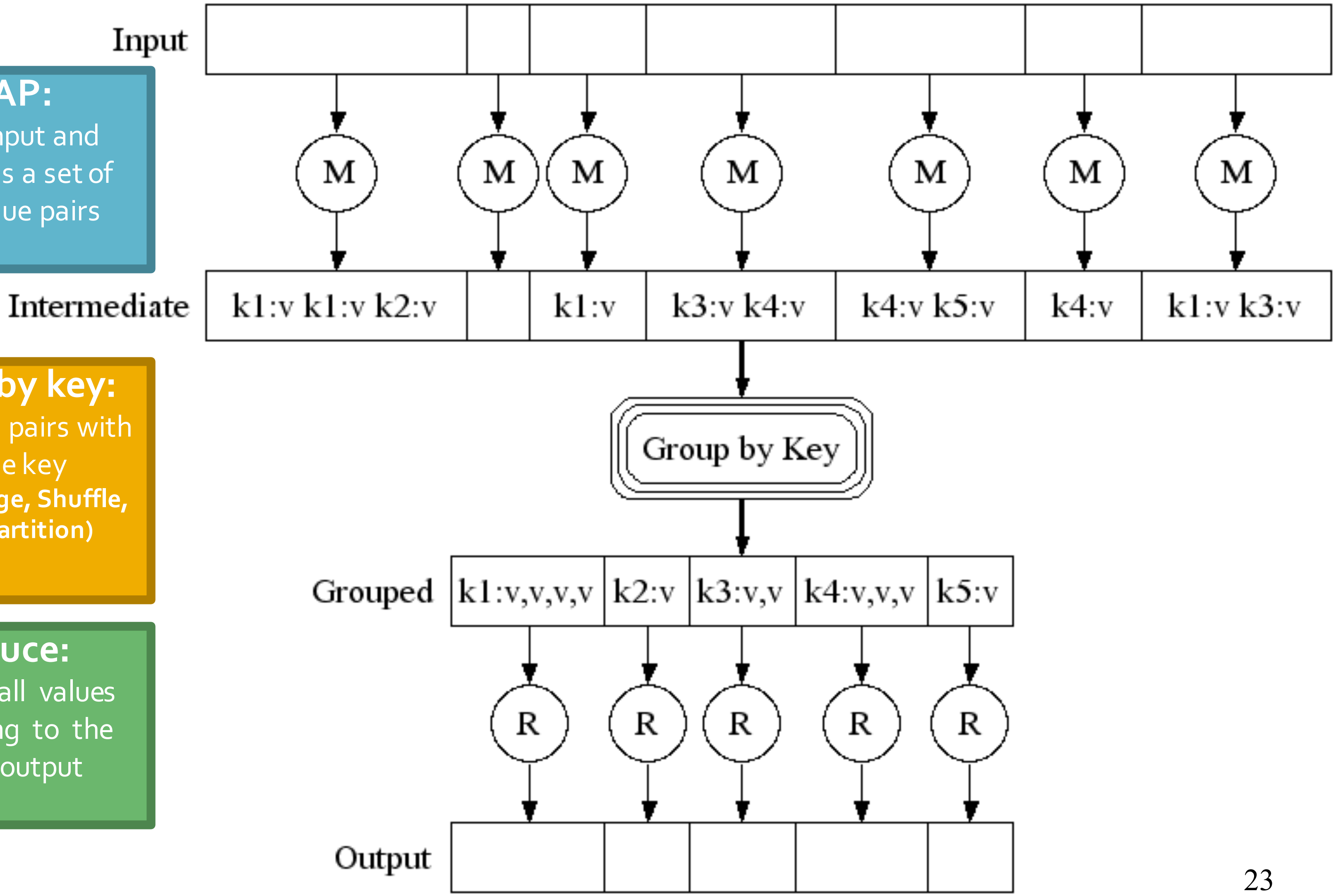
- **Input:** a set of key-value pairs
- **Programmer specifies two methods:**
  - **Map(k, v)** → <k', v'>*
    - **Takes a key-value pair and outputs a set of key-value pairs**
      - E.g., key is the filename, value is a single line in the file
    - **There is one Map call for every *(k,v)* input pair**
  - **Reduce(k', <v'>*)** → <k', v''>*
    - **Takes a key-value group as input, outputs key-value pairs**
    - **All values *v'* with same key *k'* are reduced together and processed in *v'* order**
    - **There is one Reduce function call per unique key *k'*.**

Map-Reduce: A diagram

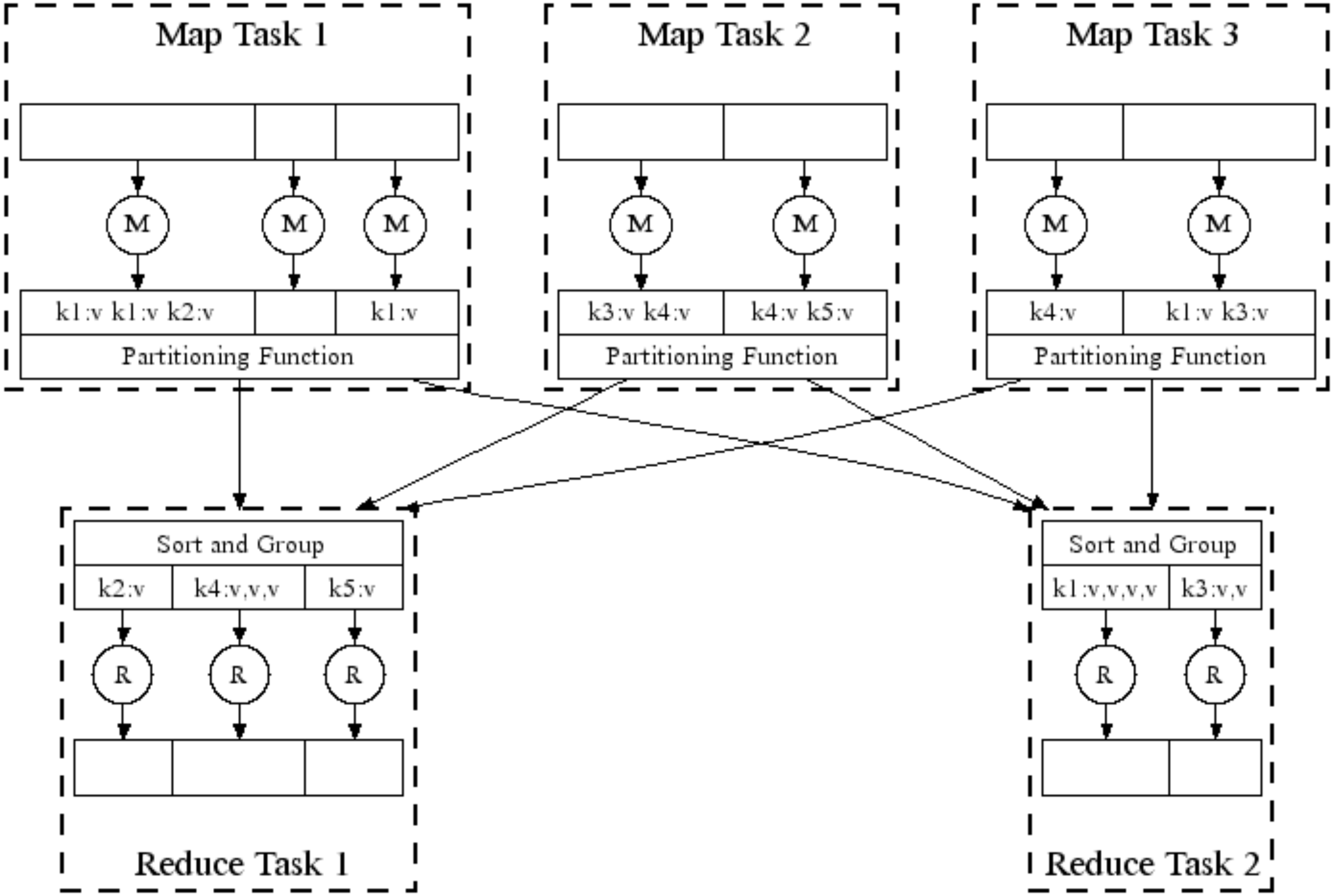
MAP:
Read input and produces a set of key-value pairs

Group by key:
Collect all pairs with same key
(Hash merge, Shuffle, Sort, Partition)

Reduce:
Collect all values belonging to the key and output



Map-Reduce: In Parallel

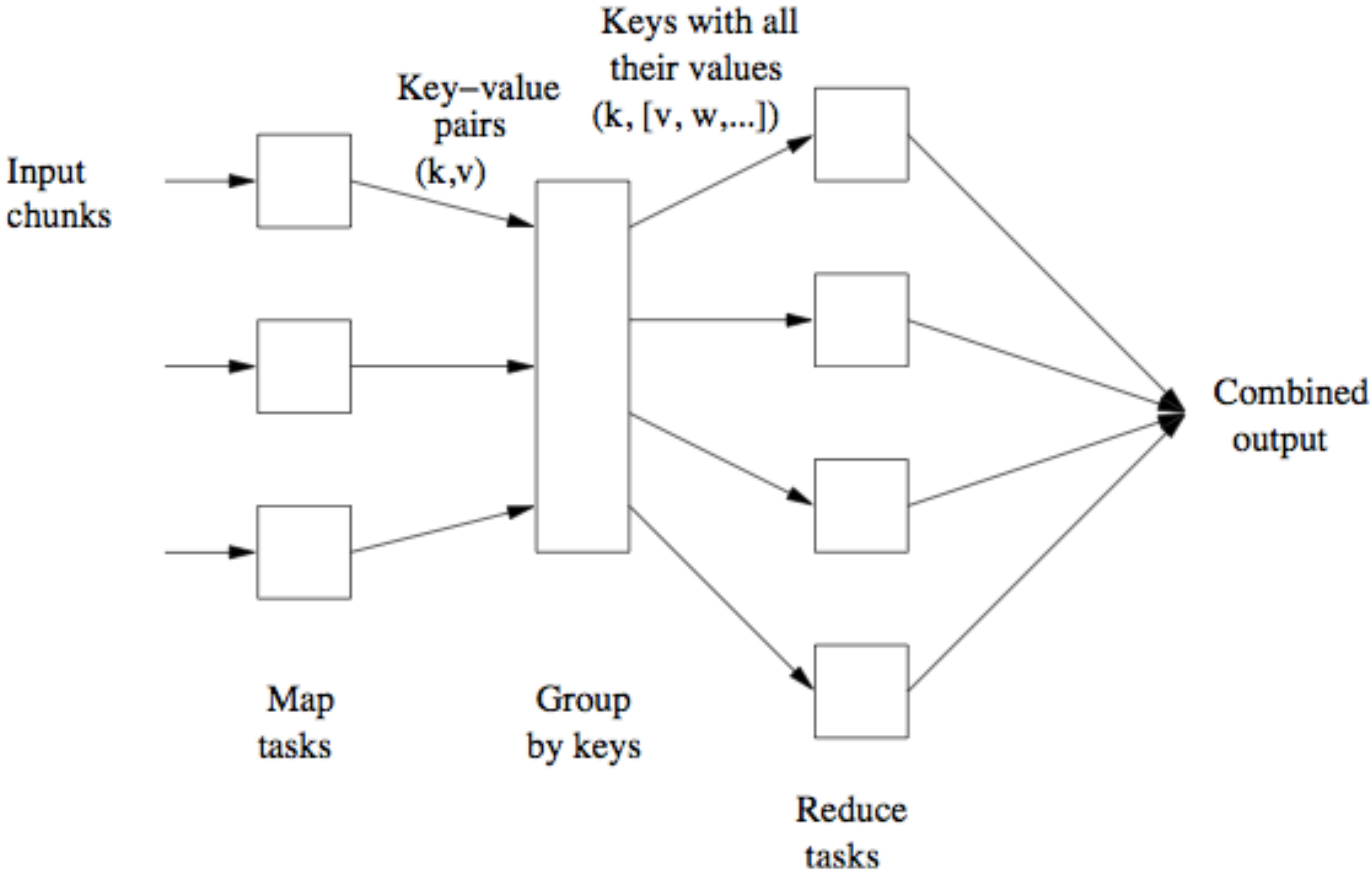


All phases are distributed with many tasks doing the work

Map-Reduce Summary

- **Map tasks**
  - Some number of **Map tasks** are given one or more **chunks** from a distributed file system
  - Map code written by the **user**
  - Processes chunks and produces sequence of **key-value pairs,**
- **Master controller: Group by Key/Shuffle**
  - Collects **key-value** pairs **from each Map task**
  - Divides keys among all **Reduce** tasks
  - All key-value pairs with **same key** go to **same Reduce task,**
- **Reduce task**
  - Works on **one key** at a time
  - Reduce code written by **user**: combines all values associated with that key in some way
  - Produces output key-value pairs.

Map-Reduce Summary



Grouping by Key: the Master Controller

After all Map tasks have all completed successfully:

- Key-value pairs are **grouped by key**
- Values associated with each key are formed into a **list of values**

Grouping is performed by the system, regardless of what the Map and Reduce tasks do

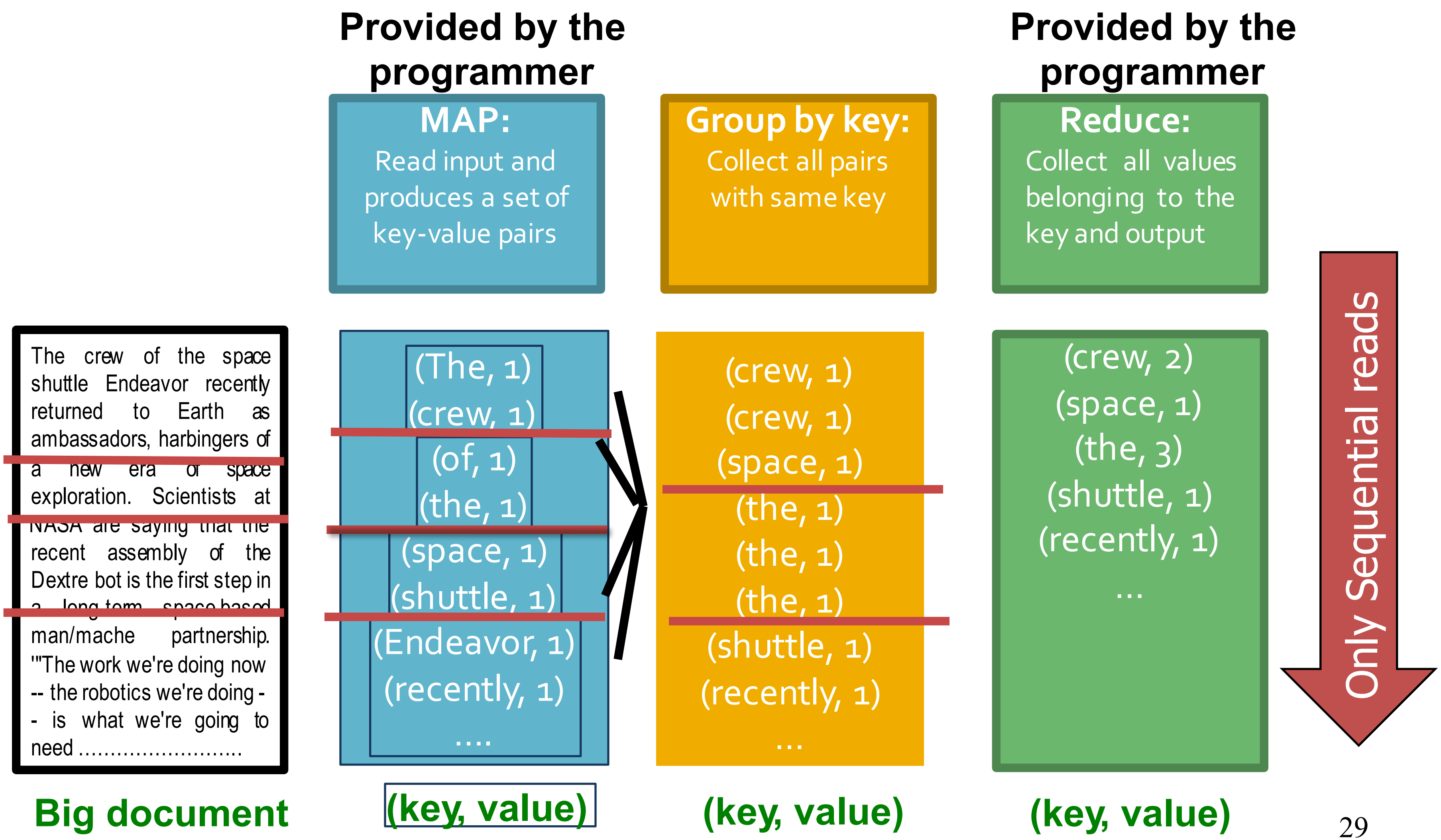
- r Reduce tasks (specified by user)
- The **master controller** typically **applies a hash function** to keys and produces a bucket number from 0 to r-1
- **Each key that is output by a Map task is hashed and its key-value pair is put in one of r local files**
- Each file is sent to one of the Reduce tasks.

More on Map Tasks

- **Map task input**
  - Input is in key-value form but normally the **keys of input** elements are **not relevant**
  - Use of this form **allows composition of several MapReduce processes**
    - E.g., in PageRank calculations

- **Keys produced by map tasks**
  - Not "keys" in usual sense: do **not have to be unique**
  - **A Map task can produce several key-value pairs with same key.**

MapReduce: Word Counting



Word Count Using MapReduce

Pseudocode

```
map(key, value):
// key: document name; value: text of the document
  for each word w in value:
     emit(w, 1)



reduce(key, values):
// key: a word; value: an iterator over counts
      result = 0
      for each count v in values:
            result += v
      emit(key, result)
```

Combiners

- **Sometimes, a Reduce function is associative and commutative**
  - Can be combined in any order, with the same result
- **Can push some of what Reducers do to Map task**
- Reduces **network traffic** for data sent to Reduce task

- Example: word count
- Instead of producing many pairs (w, 1), (w, 1), … can sum the *n* occurrences and emit (w, *n*)

- **Still required to do aggregation at the Reduce task for key, value pairs coming from multiple Map tasks.**

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 unsullied by falsehood.

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

Example: Word length histogram

Big = Yellow = 10+ letters

Medium = Red = 5..9 letters

Small = Blue = 2..4 letters

Tiny = Pink = 1 letter

Map task:

Reduce task:

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 unsullied by falsehood.

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 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 unsullied by falsehood.

Example: Word length histogram

Map Task 1
(204 words)

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

(key, value)

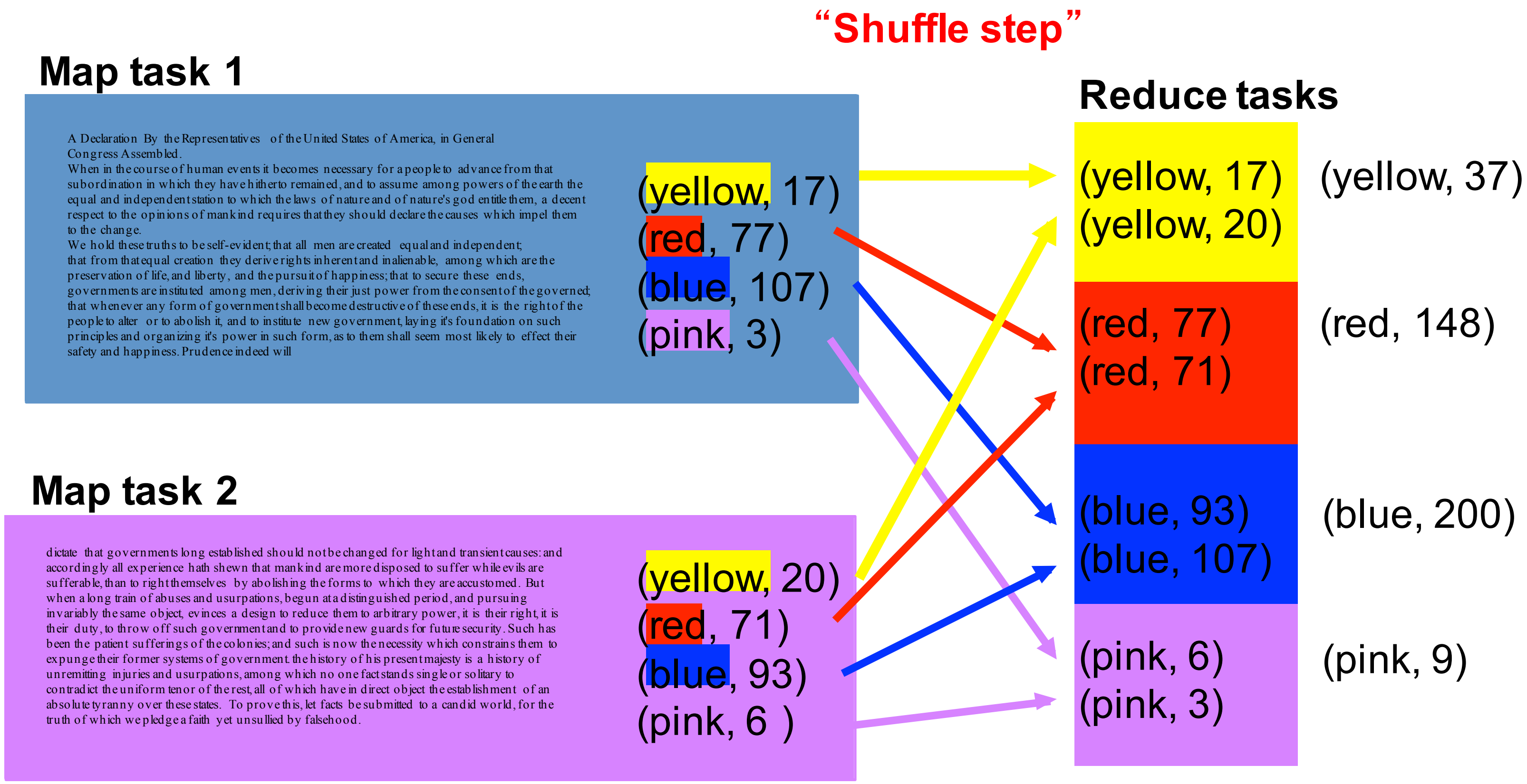
(yellow, 17)
(red, 77)
(blue, 107)
(pink, 3)

Map Task 2
(190 words)

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.

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

Example: Word length histogram



Does this Example use a Combiner?

Example: Host size

- **Suppose we have a large web corpus**
- Look at the metadata file
  - Lines of the form: (URL, size, date, …)
- **For each host, find the total number of bytes**
  - That is, the sum of the page **sizes for all URLs** from that particular **host**
- **Map ?**
  - For each record, output **hostname(URL, size)**
- **Reduce ?**
  - Sum the sizes for each host
- **Other examples:**
  - Link analysis and graph processing
  - Machine Learning algorithms.

Example: Language Model

- **Statistical machine translation:**
  - Need to count number of times every 5-word sequence occurs in a large corpus of documents

- **Very easy with MapReduce:**
  - **Map:**
    - Extract (5-word sequence, count) from document
  - **Reduce:**
    - Combine the counts.

More Examples: Build an Inverted Index

- **Data structure that maps content to its location**
- **Used in search engines to support full text searches**

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)

…

Map task:

What intermediate key, value pairs produced?

Reduce task: ?

Parse the tweet to identify **individual words**

For each **word, emit (word, tweetID)** as intermediate key-value pair

Groups key-value pairs by **key (word)**

Reduce function then just emits **key and list of tweetIDs** associated with that key.

Simple Social Network Analysis: Count Friends

- **In a social network (Facebook, Instagram, etc.), how many friends does each person have?**
- **Massive database spread across a distributed file system**

Input

Jim, Sue
Sue, Jim
Lin, Joe
Joe, Lin
Jim, Kai
Kai, Jim
Jim, Lin
Lin, Jim

MAP

Jim, 1
Sue, 1
Lin, 1
Joe, 1
***Jim**, 1*
Kai, 1
Jim, 1
Lin, 1

SHUFFLE

Jim, (1, 1, 1)
Lin, (1, 1)
Sue, (1)
Kai, (1)
Joe, (1)

REDUCE

Desired Output

Jim, 3
Lin, 2
Sue, 1
Kai, 1
Joe, 1

Map task:

Reduce task:

Could you use a combiner in this case?

Example: Integers divisible by 7

- Design a MapReduce algorithm that takes a **very large file of integers** and produces as output **all unique integers from the original file that are evenly divisible by 7**

- Very large file of integers => can't be handled in one node's main memory
  - Data are distributed across multiple nodes in chunks
- **Each Map task gets a chunk of the file of integers and processes it**

- Map task:?
- Reduce task:?

Divisible by 7 (cont.)

```
map(key, valuelist):
        for v in valuelist:
                if (v % 7) == 0 :
                        emit(v, 1)
reduce(key, values):
        // Eliminate duplicates
        emit (key, 1)
```

Question: Why check whether divisible by 7 in the Map task rather than the Reduce task?

Reduce communication: send less data over network

Perform logic in the Map where possible without introducing errors.

More on Combiners

- Yahoo tutorial on MapReduce
- https://developer.yahoo.com/hadoop/tutorial/module4.html#wordcount

- **Combiner:** this pass runs after the Mapper and before the Reducer
- Usage of the Combiner is optional
- If this pass is suitable for your job, instances of the Combiner class are run on every node that has run map tasks.

More on Combiners

- The Combiner will receive as input all data emitted by **the Mapper** instances on a given node
- The **output** from the Combiner is then sent to the **Reducers**, instead of the output from the Mappers
- The Combiner is **a "mini-reduce" process which operates only on data generated by one machine.**

More on Combiners

- **Word count** is a prime example for where a Combiner is useful
- Map task: Emits a (*word*, 1) pair for every instance of every word it sees
  - E.g., if the same document contains the word "cat" 3 times, the pair ("cat", 1) is emitted three times
  - All of these are then sent to the Reducer
- By using a Combiner, these can be condensed into a single ("cat", 3) pair to be sent to the Reducer
- Now each node only sends a single value to the reducer for each word
- Drastically **reduces the total bandwidth** required for the shuffle process, and speeds up the job.

More on Combiners

- We do not need to write any additional code to take advantage of this
- **If a Reduce function is both *commutative* and *associative*, then it can be used as a Combiner as well**
- You can enable combining in the word count program by adding the following line to the driver:
- `conf.setCombinerClass(Reduce.class);`
- The Combiner should be an instance of the *Reducer* interface
- In some cases, the Reducer itself cannot be used directly as a Combiner because of commutativity or associativity
- You might still be able **to write a third class to use as a Combiner for your job.**

Summary

- Large-scale computing for data mining
- Cluster architecture
- How do you distribute computation?
  - How can we make it easy to write distributed programs?
  - Distributed file system
  - Chunk servers and Master node
- Map-Reduce
  - Map tasks
  - Master controller: Group by Key/Shuffle
  - Reduce task.