

WHAT DOES SCALABLE MEAN?

Operationally:

- In the past: **“Works even if data doesn’t fit in main memory”**
- Now: **“Can make use of 1000s of cheap computers”**

Formally:

- In the past: **If you have N data items, you must do no more than N^k operations -- “polynomial time algorithms”**
- Soon: **If you have N data items, you must do no more than $N * \log(N)$ operations -- “logarithmic time algorithms”**
- As data sizes go up, you’ll only get **one pass** at the data
- So you better make that one pass count

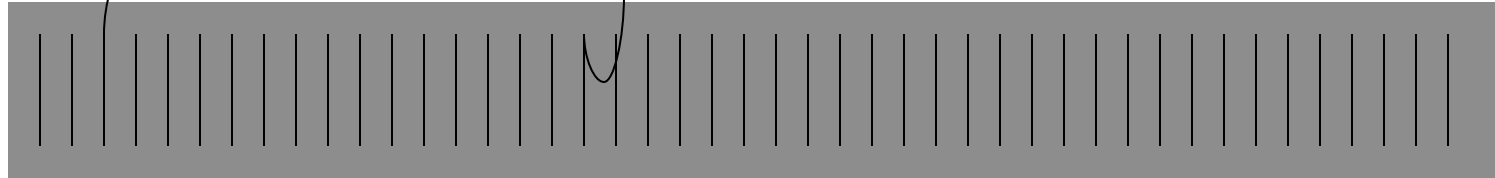
EXAMPLE: FIND MATCHING DNA SEQUENCES

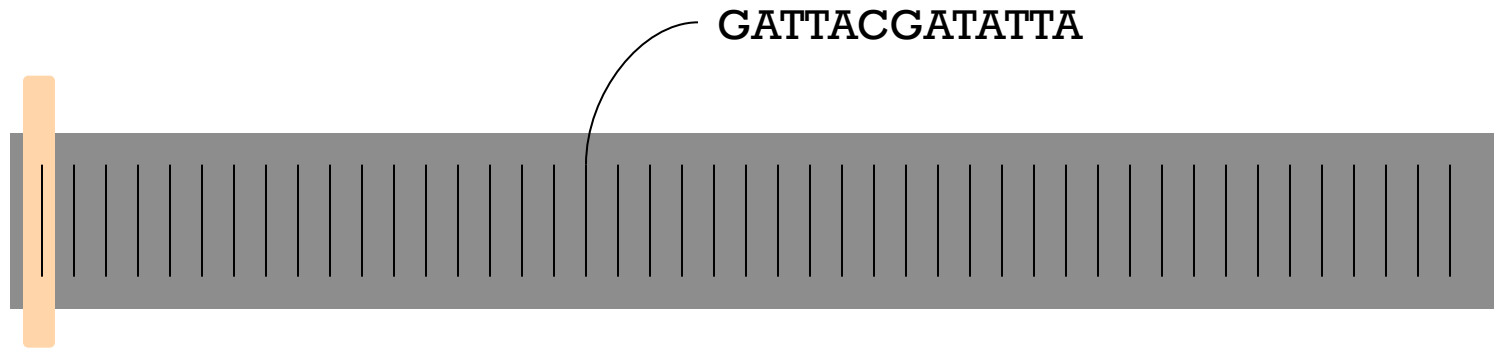
Given a set of sequences

Find all sequences equal to “GATTACGATATTA”

TACCTGCCGTAA

GATTACGATATTA

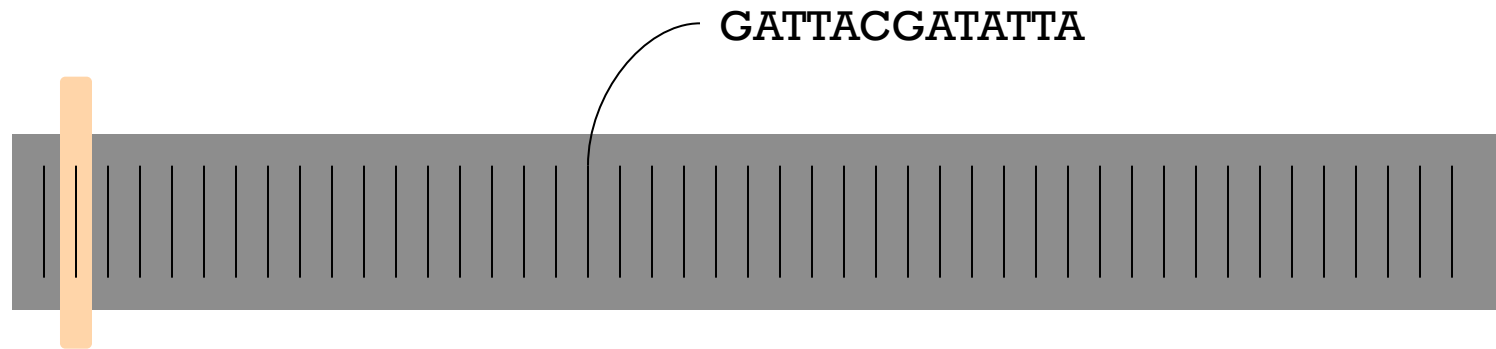




TACCTGCCGTAA = GATTACGATATTA?

No.

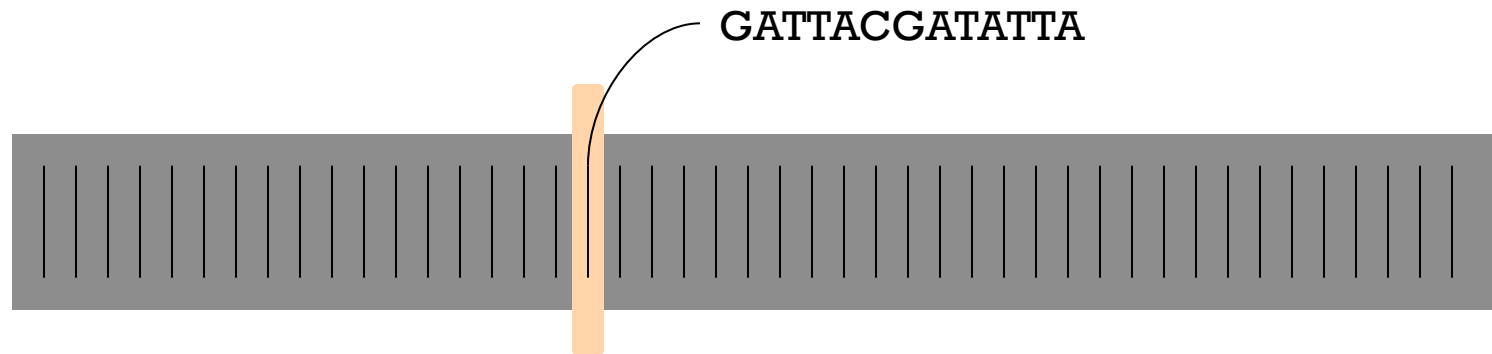
time = 0



CCCCCAATGAC = GATTACGATATTA?

No.

time = 1

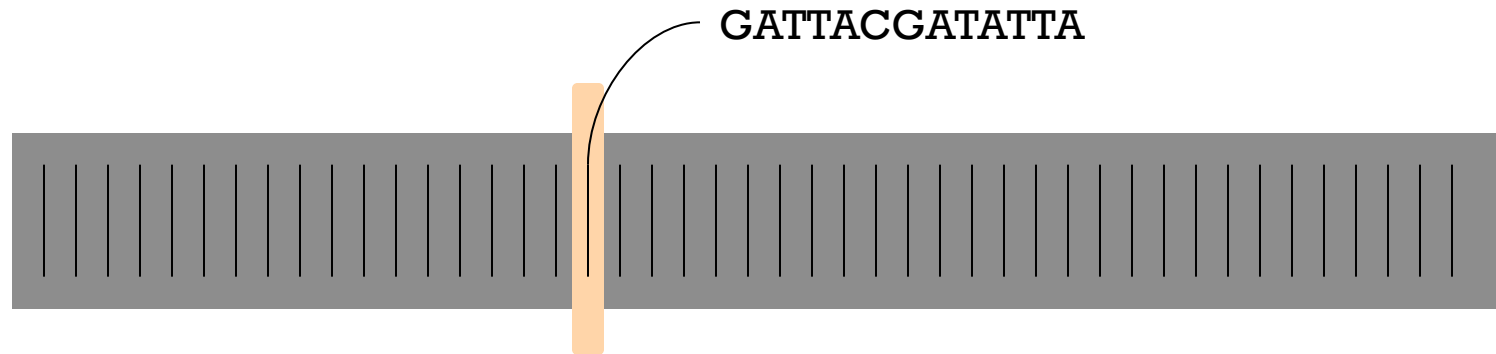


GATTACGATATTA contains GATTACGATATTA?

Yes!

Send it to the output.

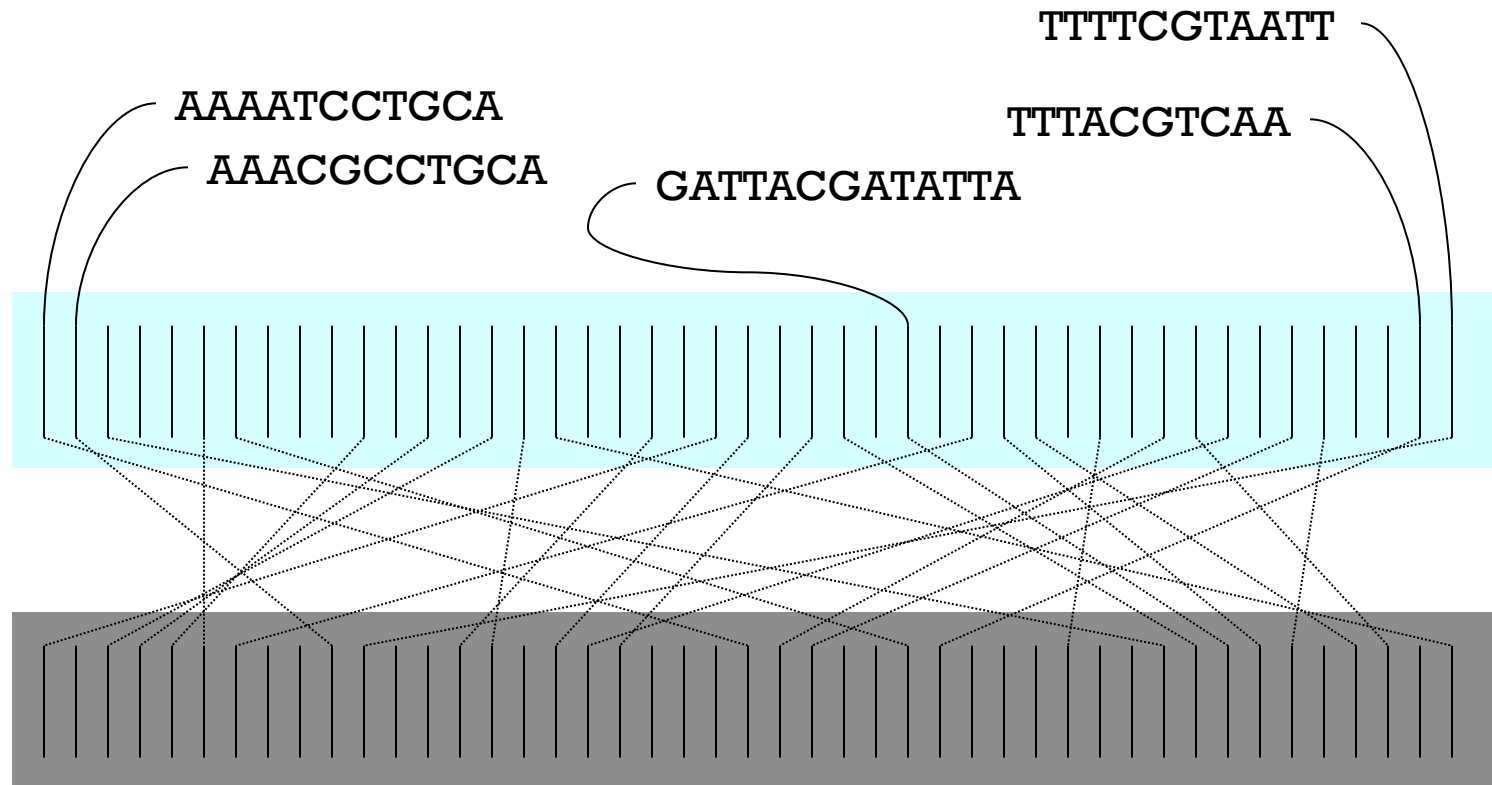
time = 17



40 records, 40 comparisons

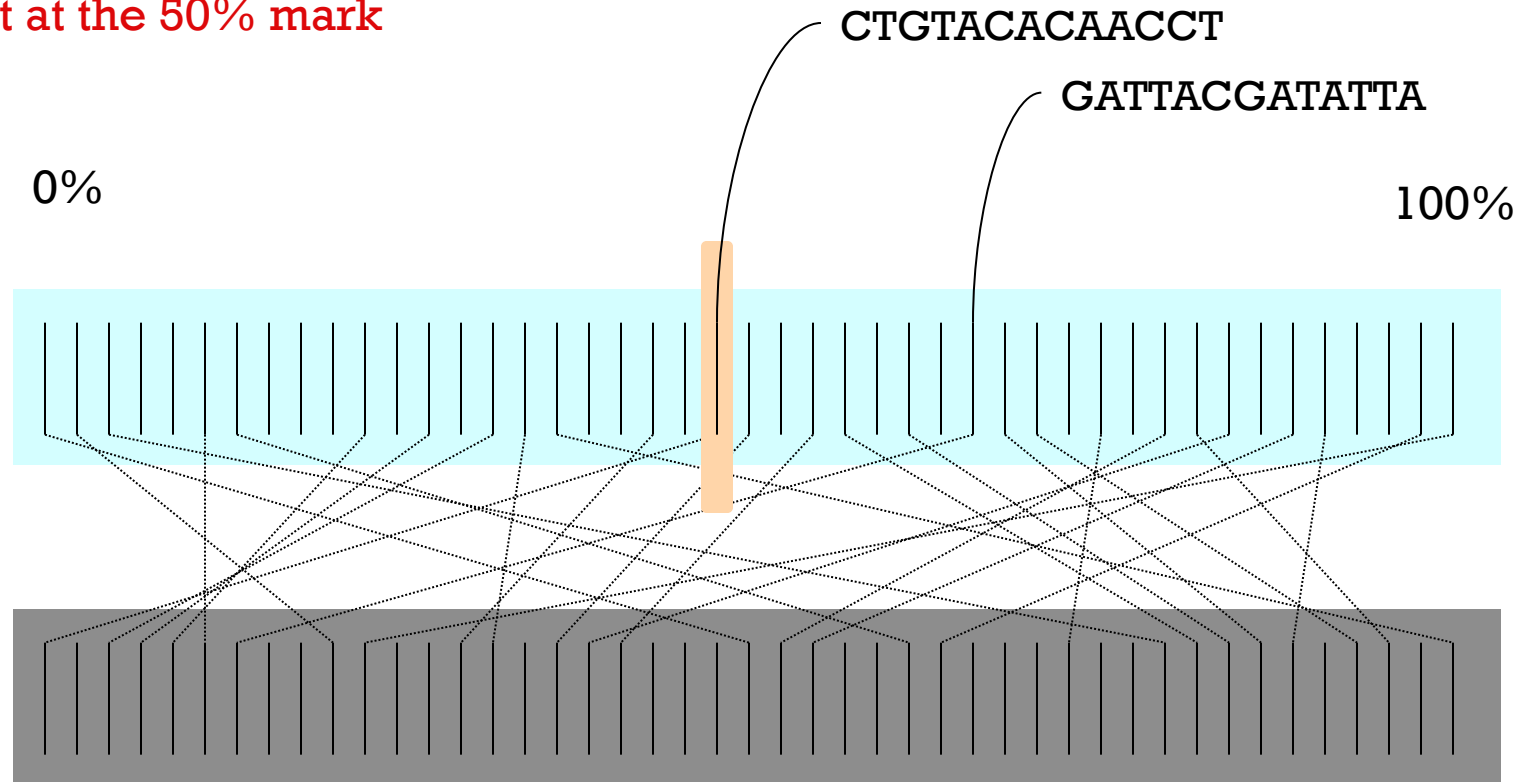
N records, N comparisons

The algorithmic complexity is order N : $O(N)$



What if we sort the sequences?

Start at the 50% mark

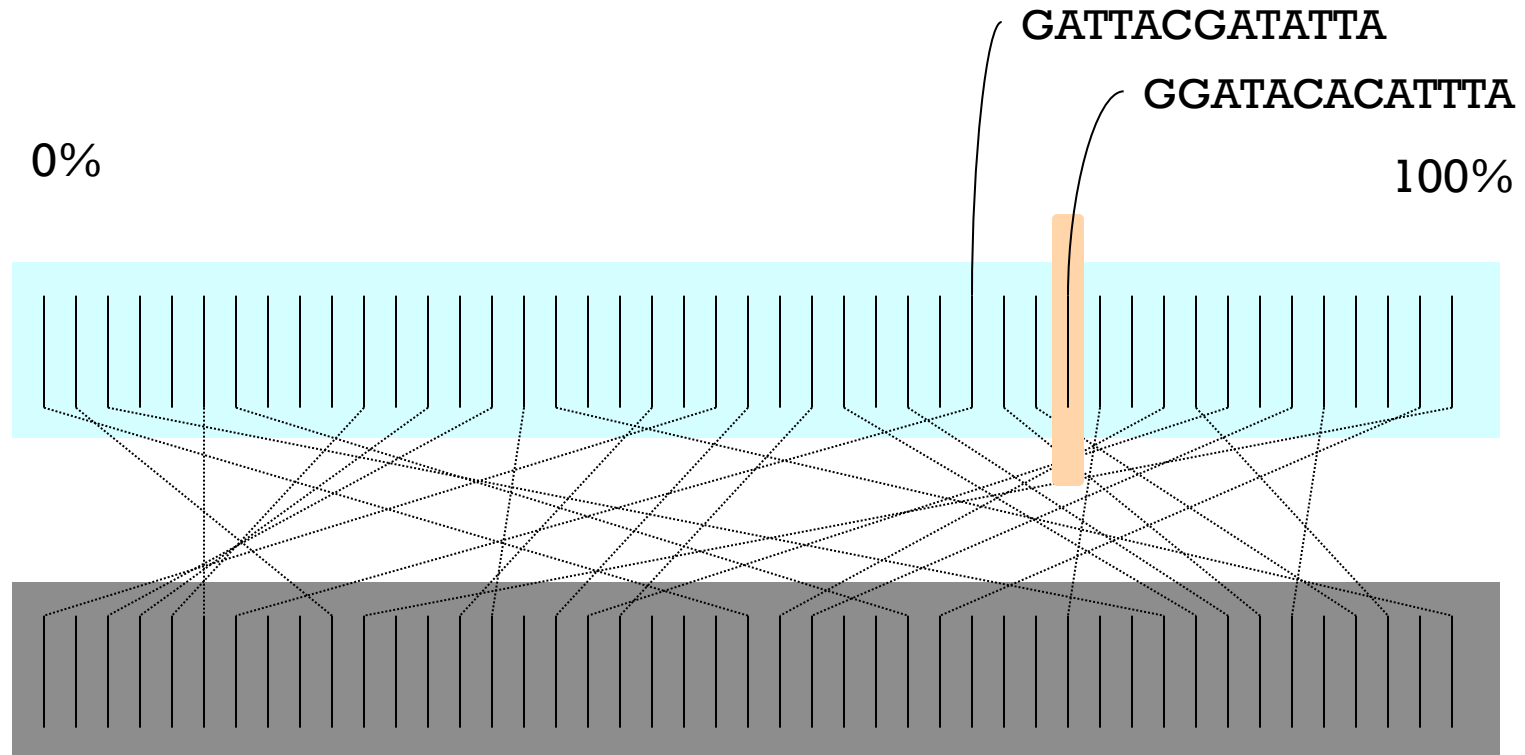


CTGTACACAACCT < GATTACGATATTA

time = 0

No match.

Skip to 75% mark

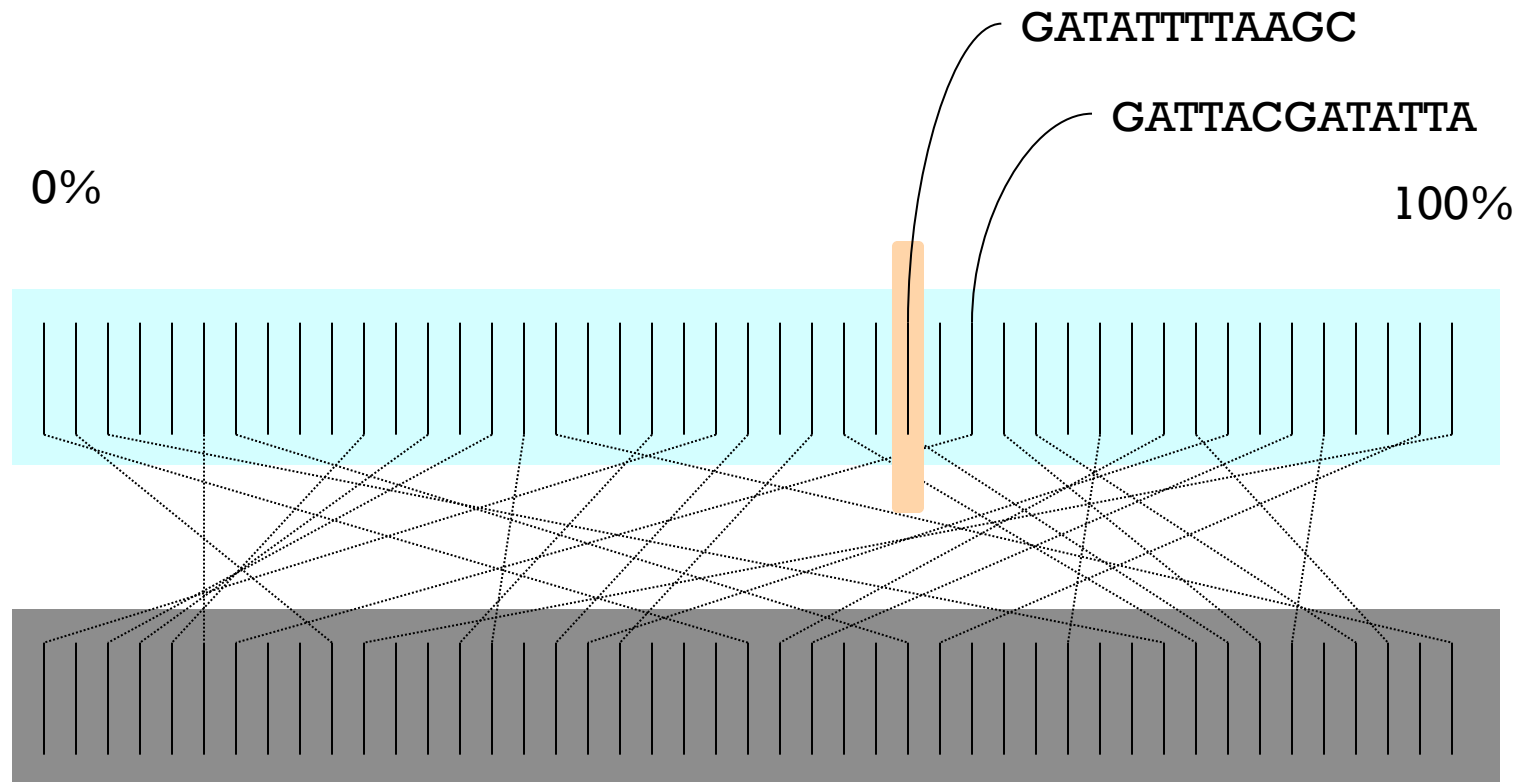


GGATACACATTTA > GATTACGATATTA

time = 1

No match.

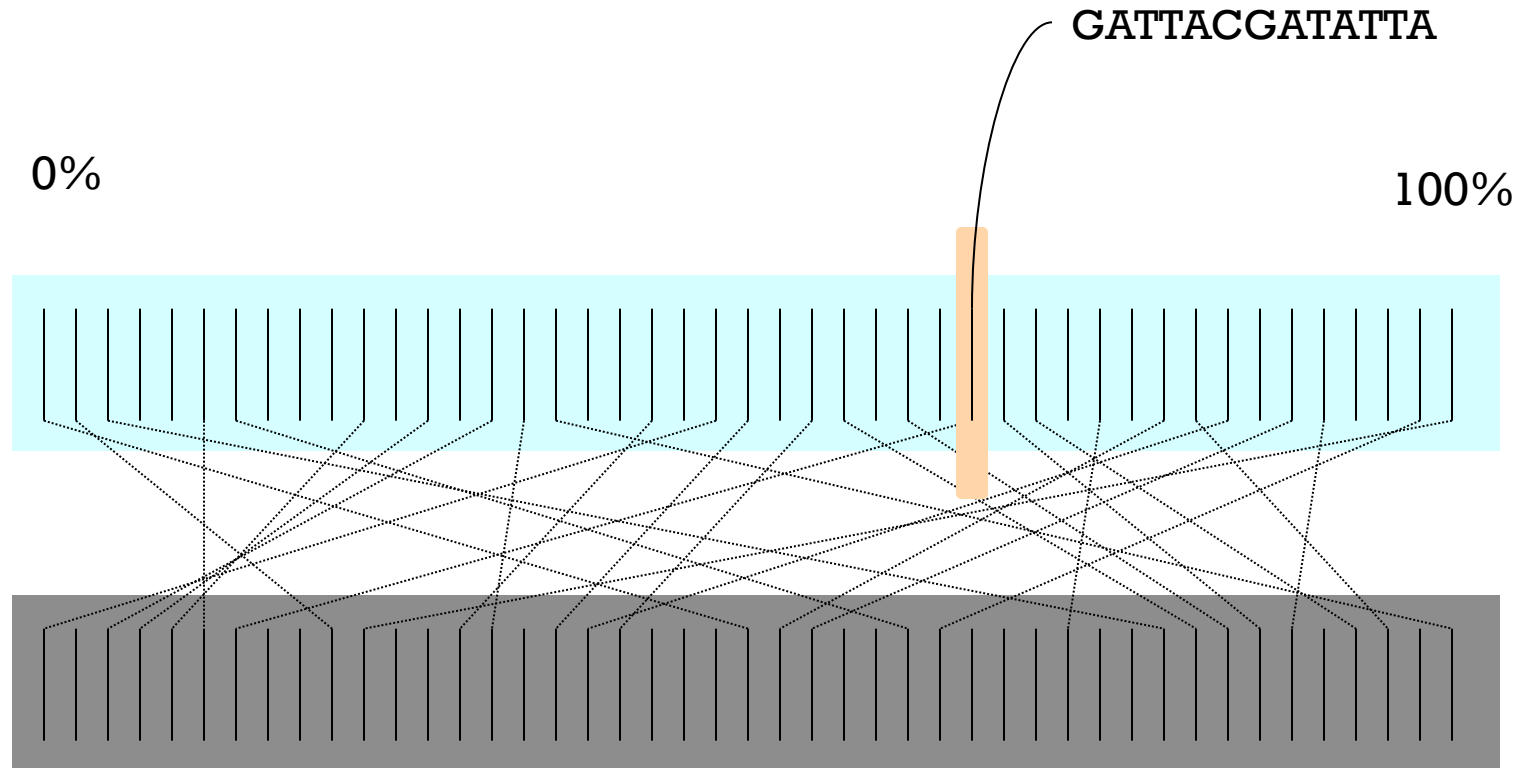
Go back to 62.5% mark



GATATTTTAAGC < GATTACGATATTA

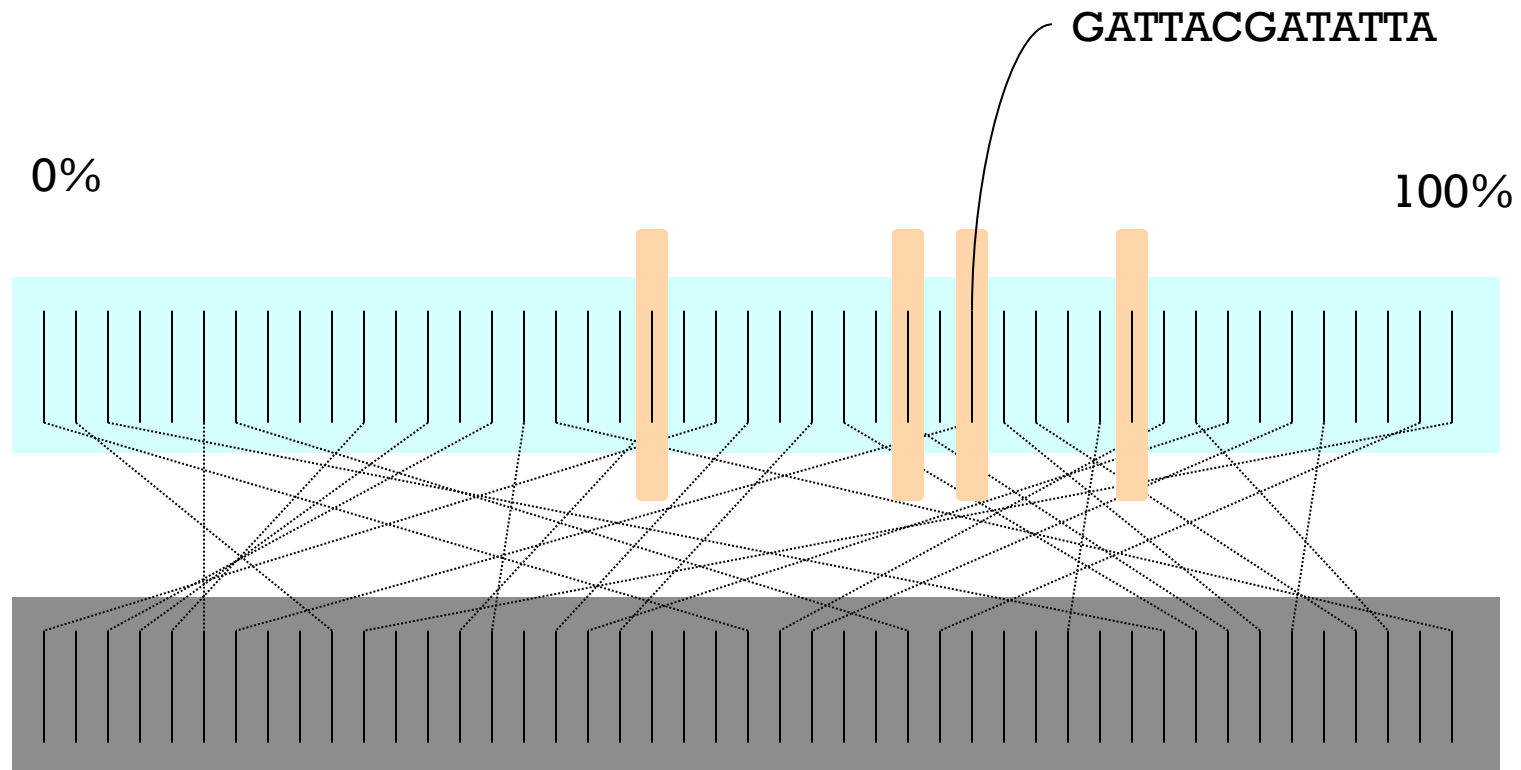
No match.

Skip back to 68.75% mark



Match!

Walk through the records until we fail to match.



How many comparisons did we do?

40 records, only 4 comparisons

N records, $\log(N)$ comparisons

This algorithm is $O(\log(N))$ Far better scalability

RELATIONAL DATABASES

Databases are good at “Needle in Haystack” problems:

- Extracting small results from big datasets
- Transparently provide “old style” scalability
- Your query will **always*** finish, regardless of dataset size.
- Indexes are easily built and automatically used when appropriate

```
CREATE INDEX seq_idx ON sequence(seq) ;  
SELECT seq  
  FROM sequence  
 WHERE seq = 'GATTACGATATTA' ;
```

***almost**

NEW TASK: READ TRIMMING

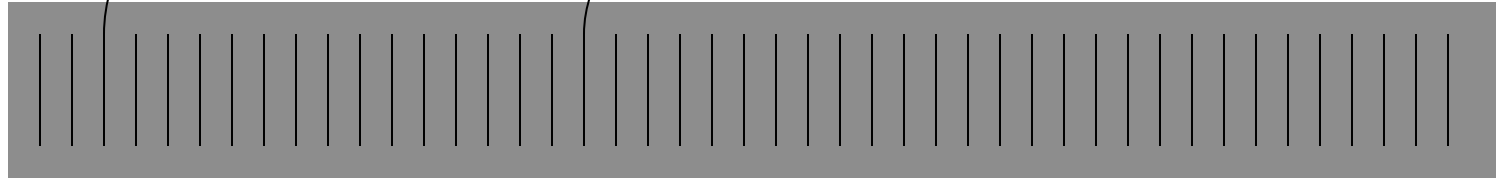
Given a set of DNA sequences

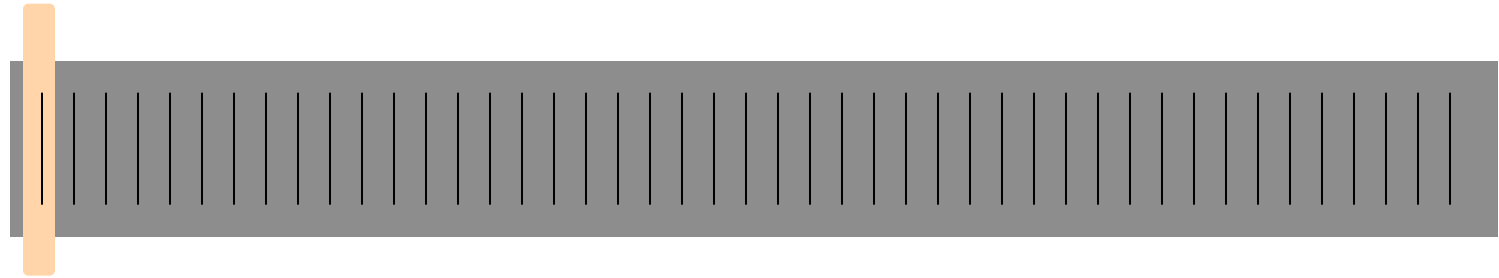
Trim the final n bps of each sequence

Generate a new dataset

TACCTGCCGTAA

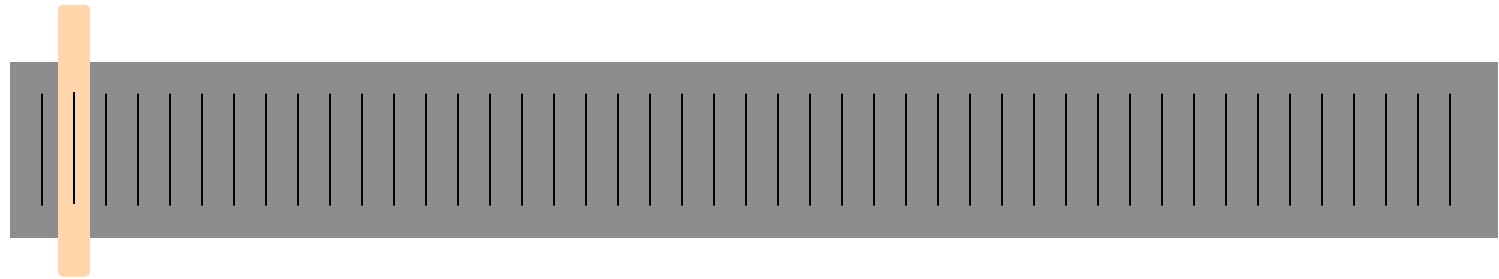
GATTACGATATTA





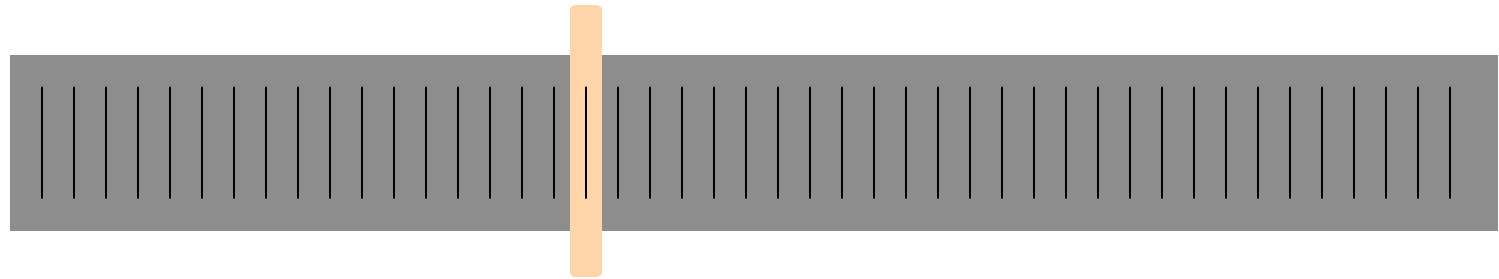
TACCTGCCGTAA becomes TACCT

time = 0



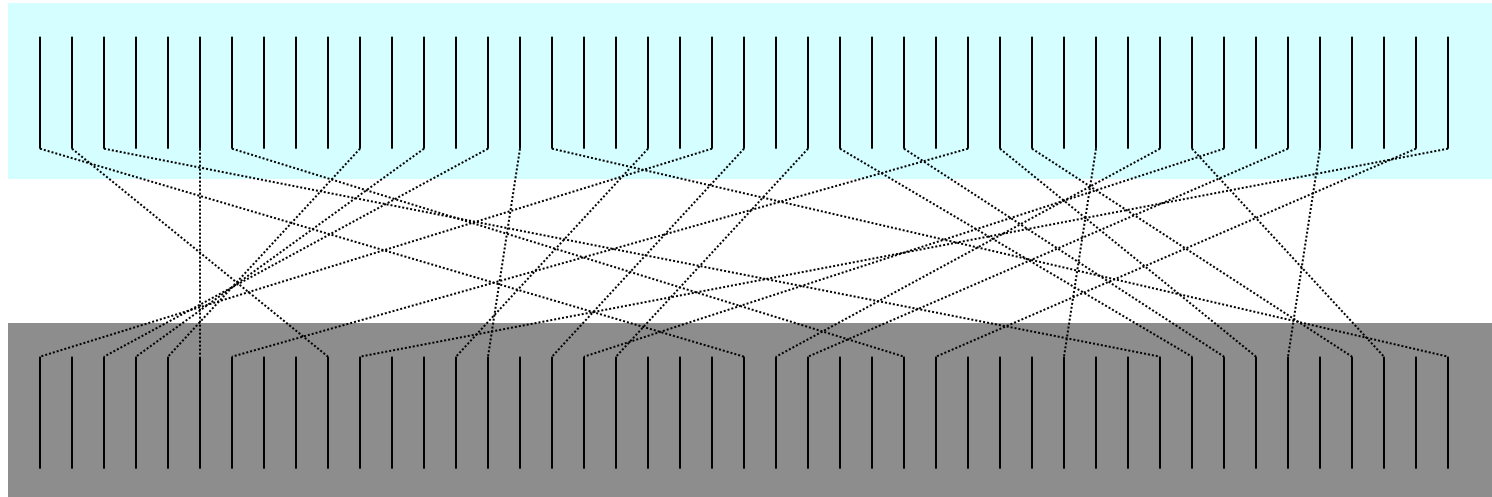
CCCCCAATGAC becomes CCCCC

time = 1



GATTACGATATTA becomes GATTA

time = 17

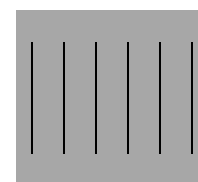
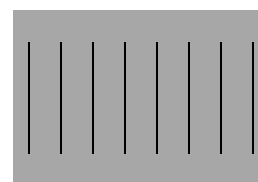
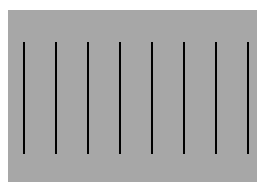
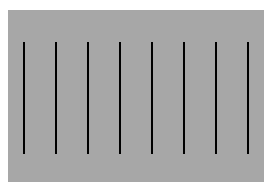
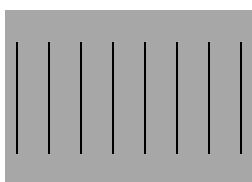
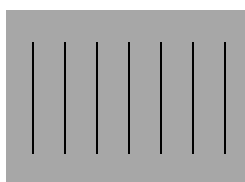
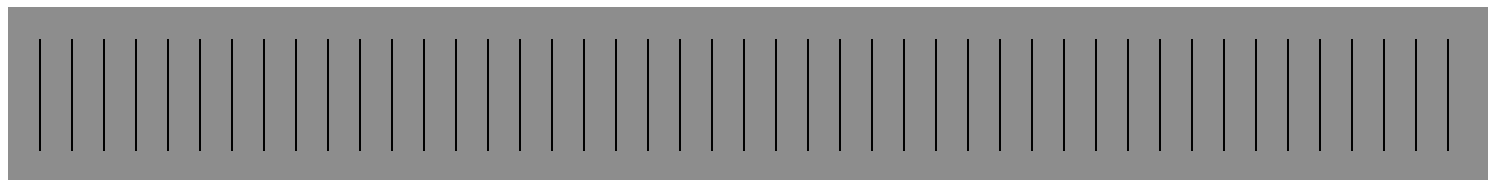


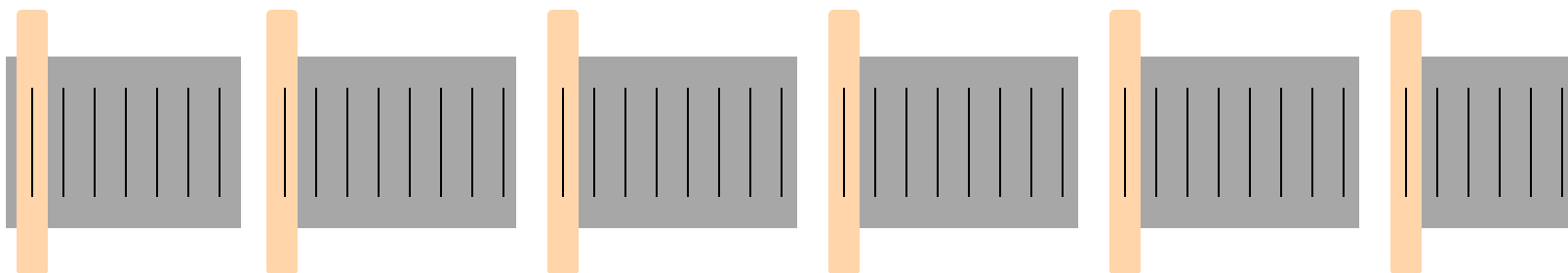
Can we use an index?

No. We have to touch every record no matter what.

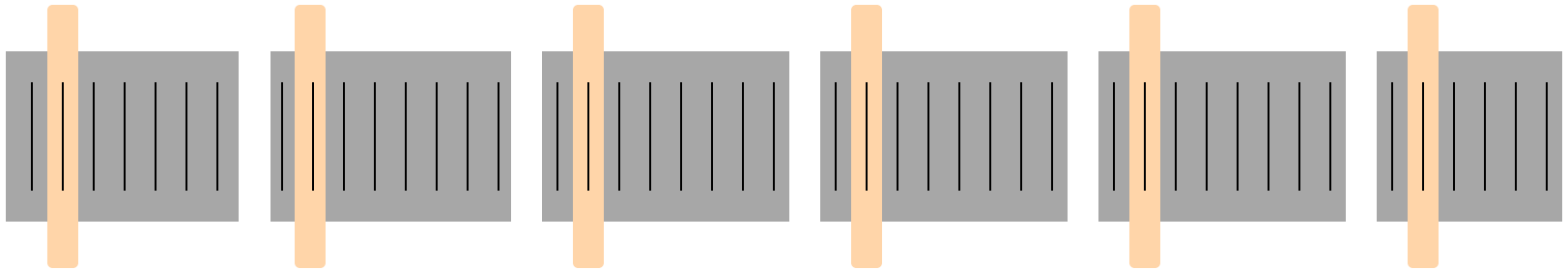
The task is fundamentally $O(N)$

Can we do any better?

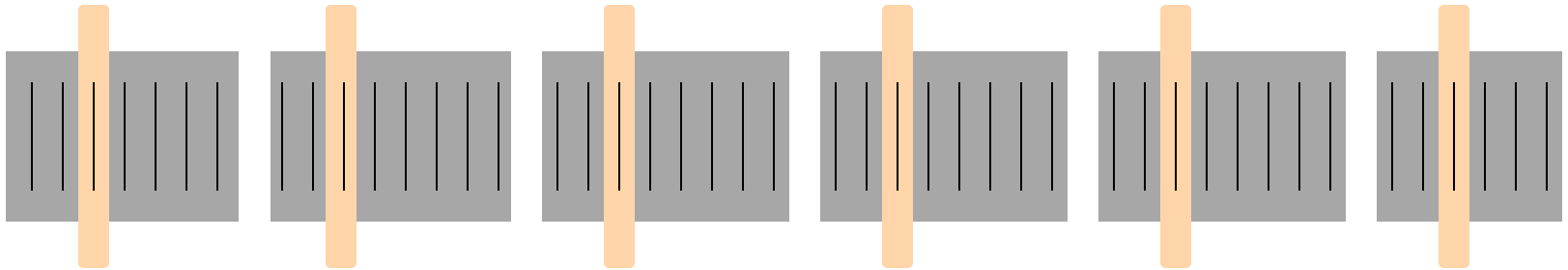




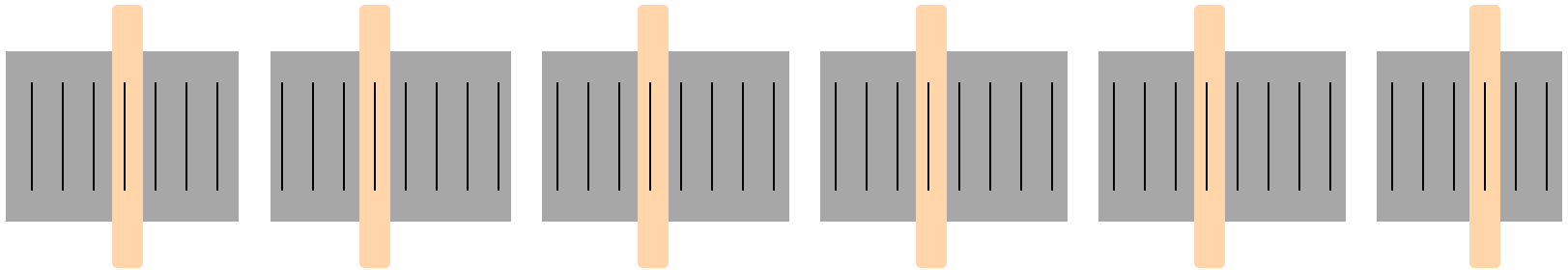
time = 0



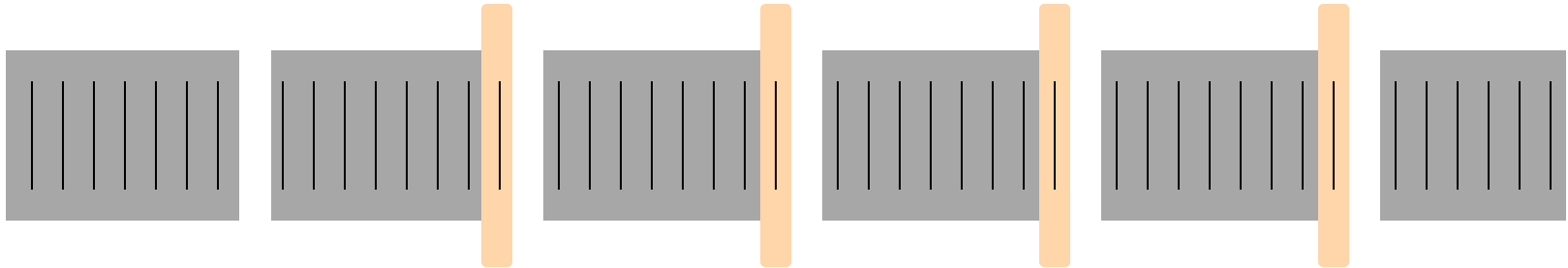
time = 1



time = 2



time = 3



How much time did this take?

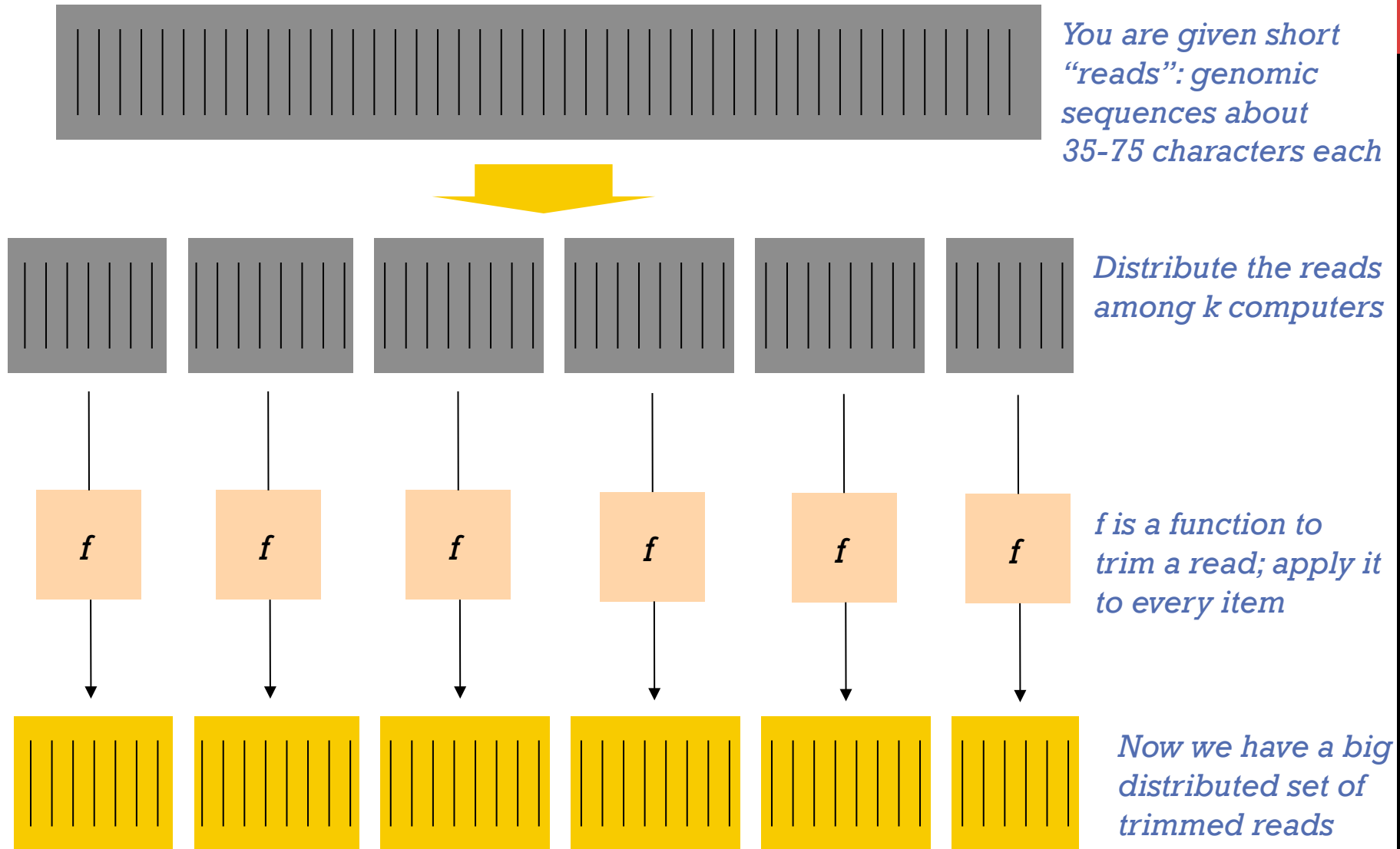
7 cycles

40 records, 6 workers

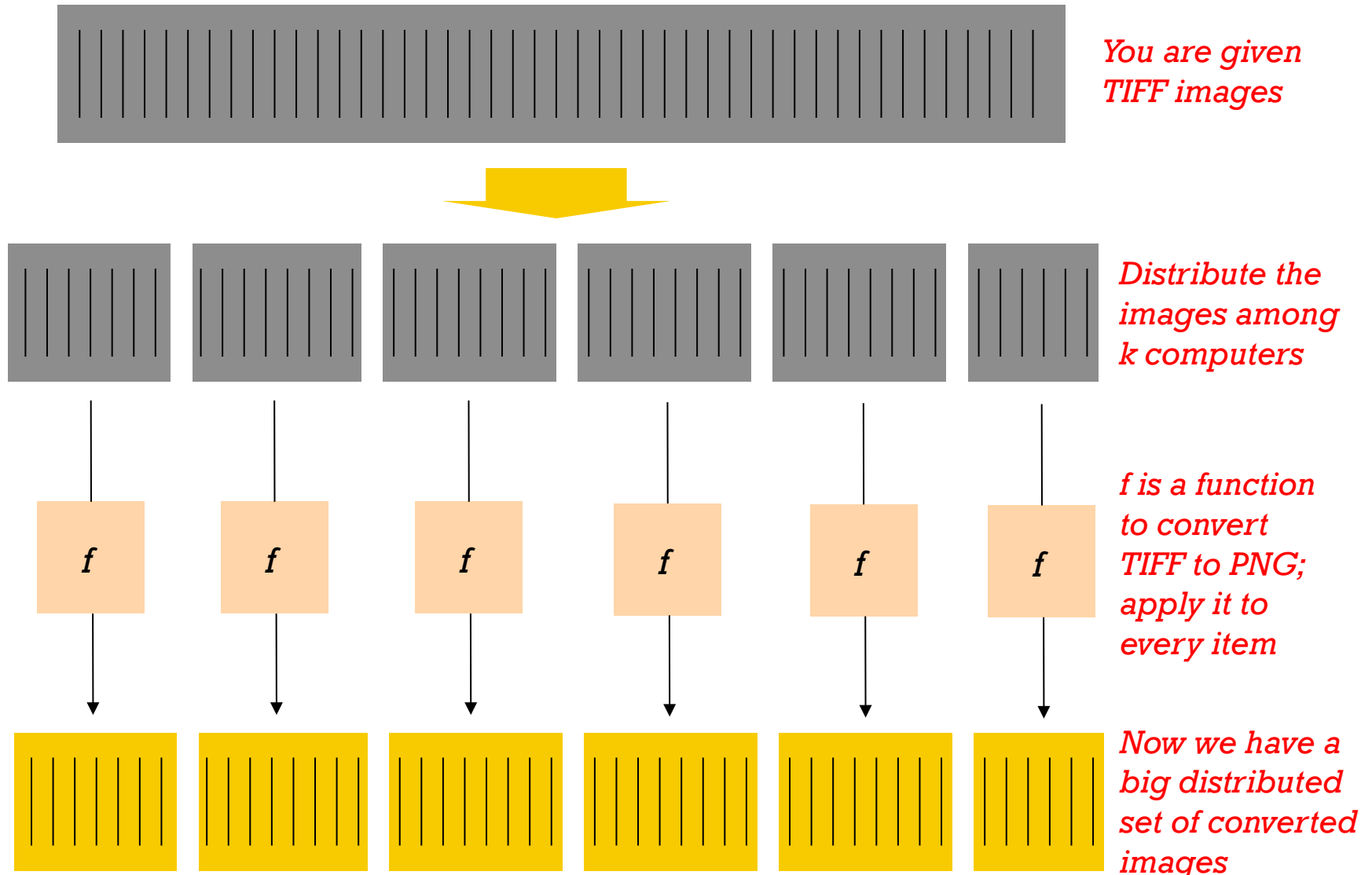
time = 7

$$O\left(\frac{N}{k}\right)$$

SCHEMATIC OF A PARALLEL “READ TRIMMING” TASK

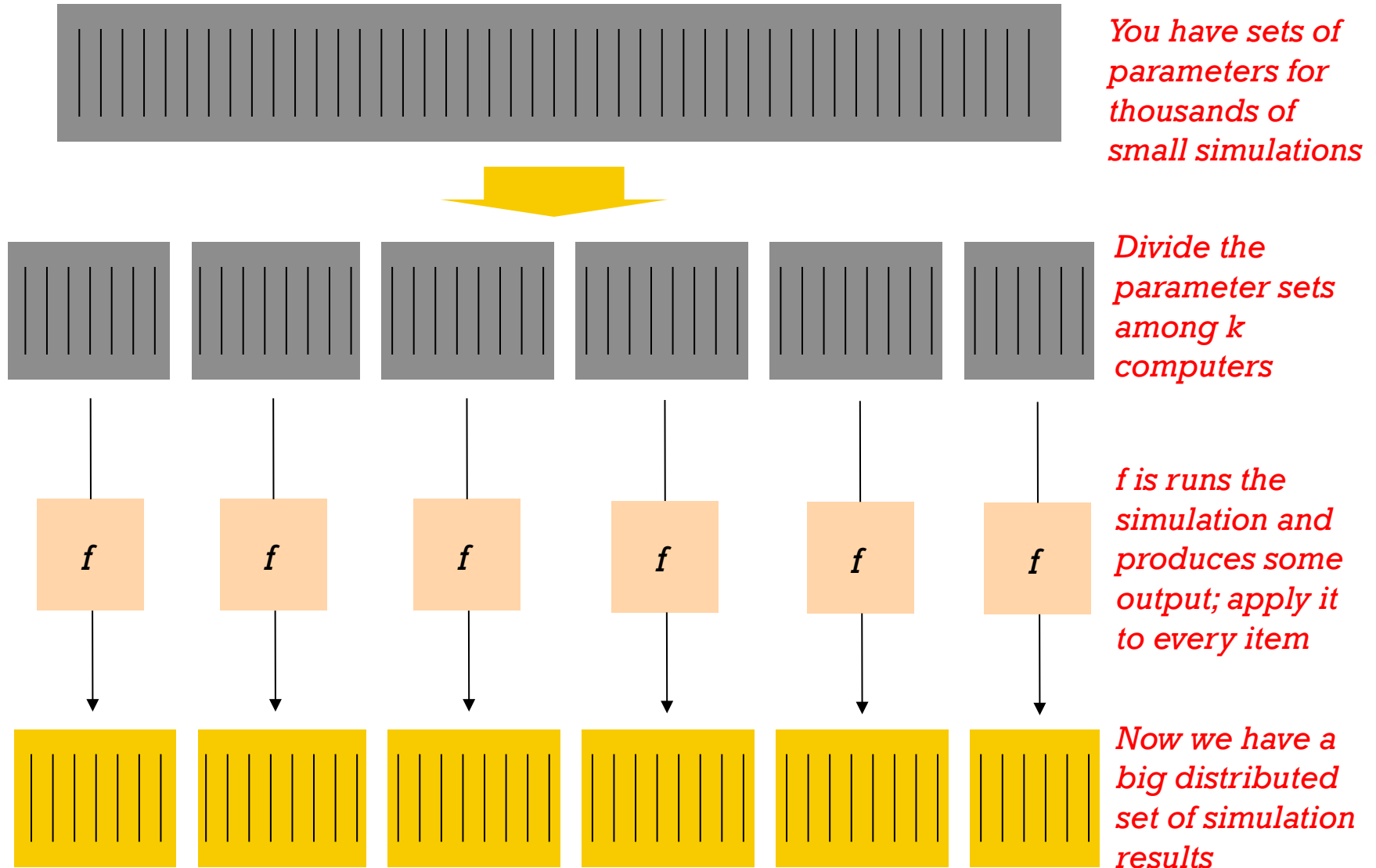


NEW TASK: CONVERT 405K TIFF IMAGES TO PNG

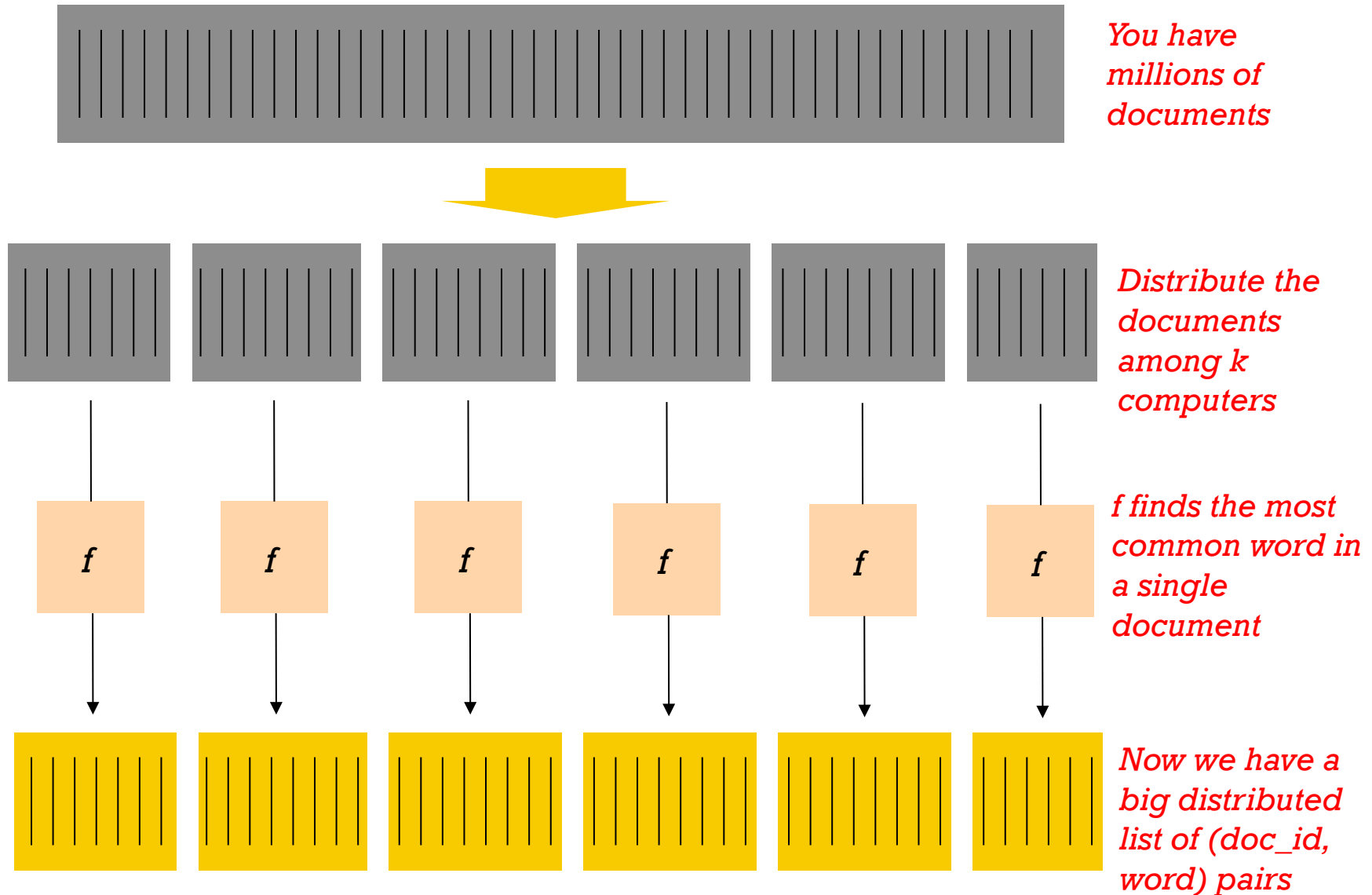


<http://open.blogs.nytimes.com/2008/05/21/the-new-york-times-archives-amazon-web-services-timesmachine/>

NEW TASK: RUN THOUSANDS OF SIMULATIONS



FIND THE MOST COMMON WORD IN EACH DOCUMENT



Consider a slightly more general program to compute the word frequency of every word in a single document

Abridged Declaration of Independence

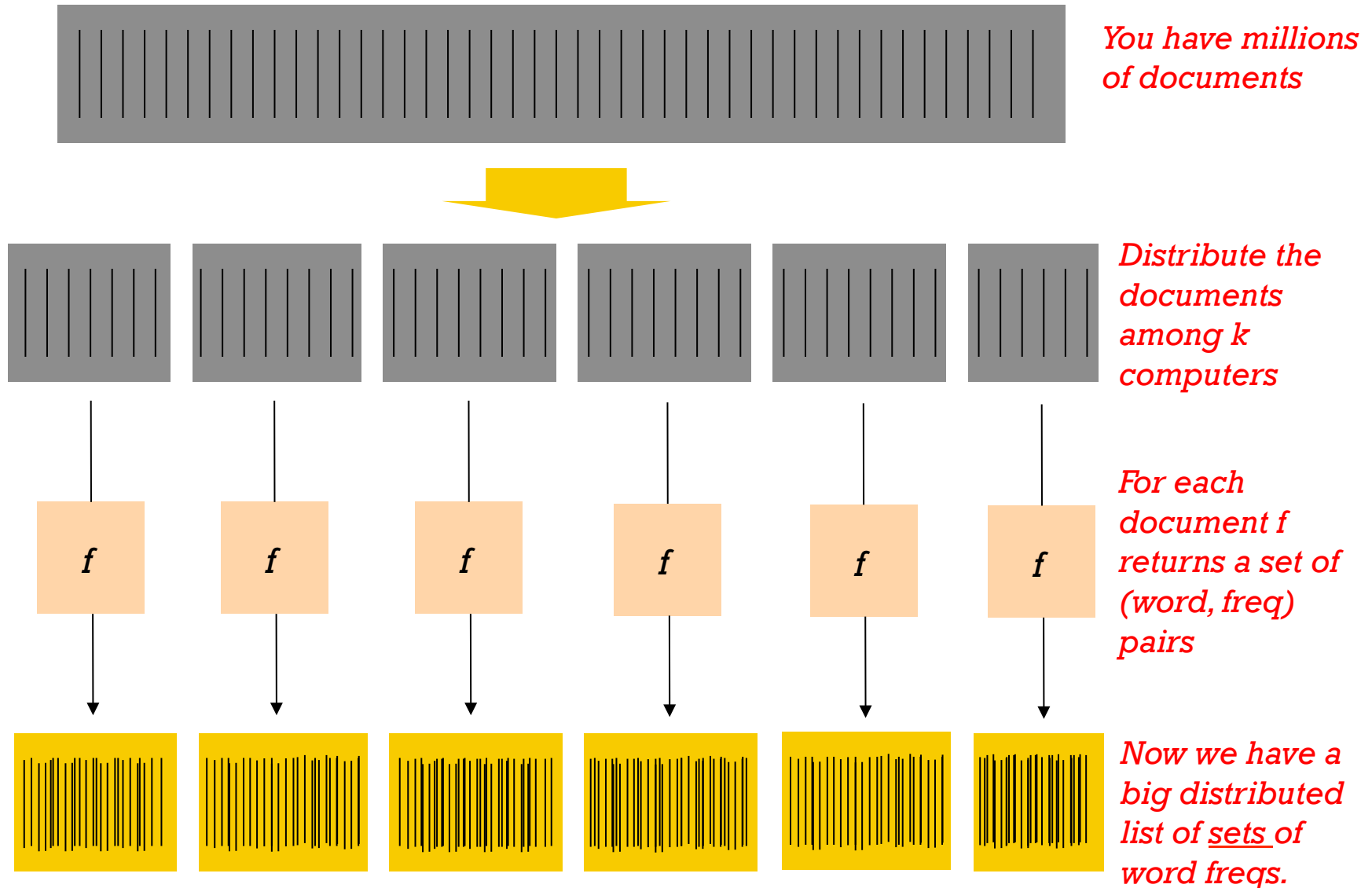
A Declaration By the Representatives of the United States of America, in General Congress Assembled. When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent 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.



(people, 2)
(government, 6)
(assume, 1)
(history, 2)
...

COMPUTE THE WORD FREQUENCY OF 5M DOCUMENTS



THERE'S A PATTERN HERE....

A function that maps a read to a trimmed read

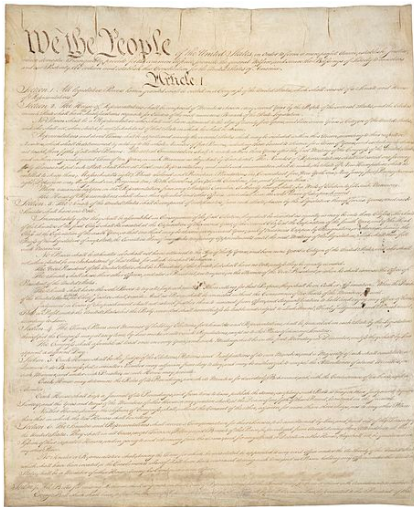
A function that maps a TIFF image to a PNG image

A function that maps a set of parameters to a simulation result

A function that maps a document to its most common word

A function that maps a document to a histogram of word frequencies

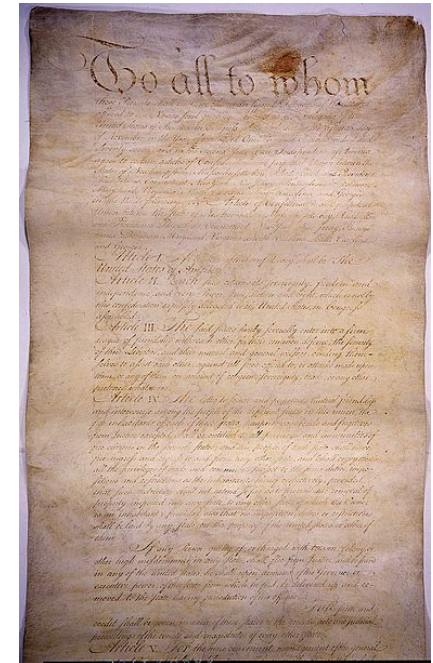
WHAT IF WE WANT TO COMPUTE THE WORD FREQUENCY ACROSS ALL DOCUMENTS?



US Constitution



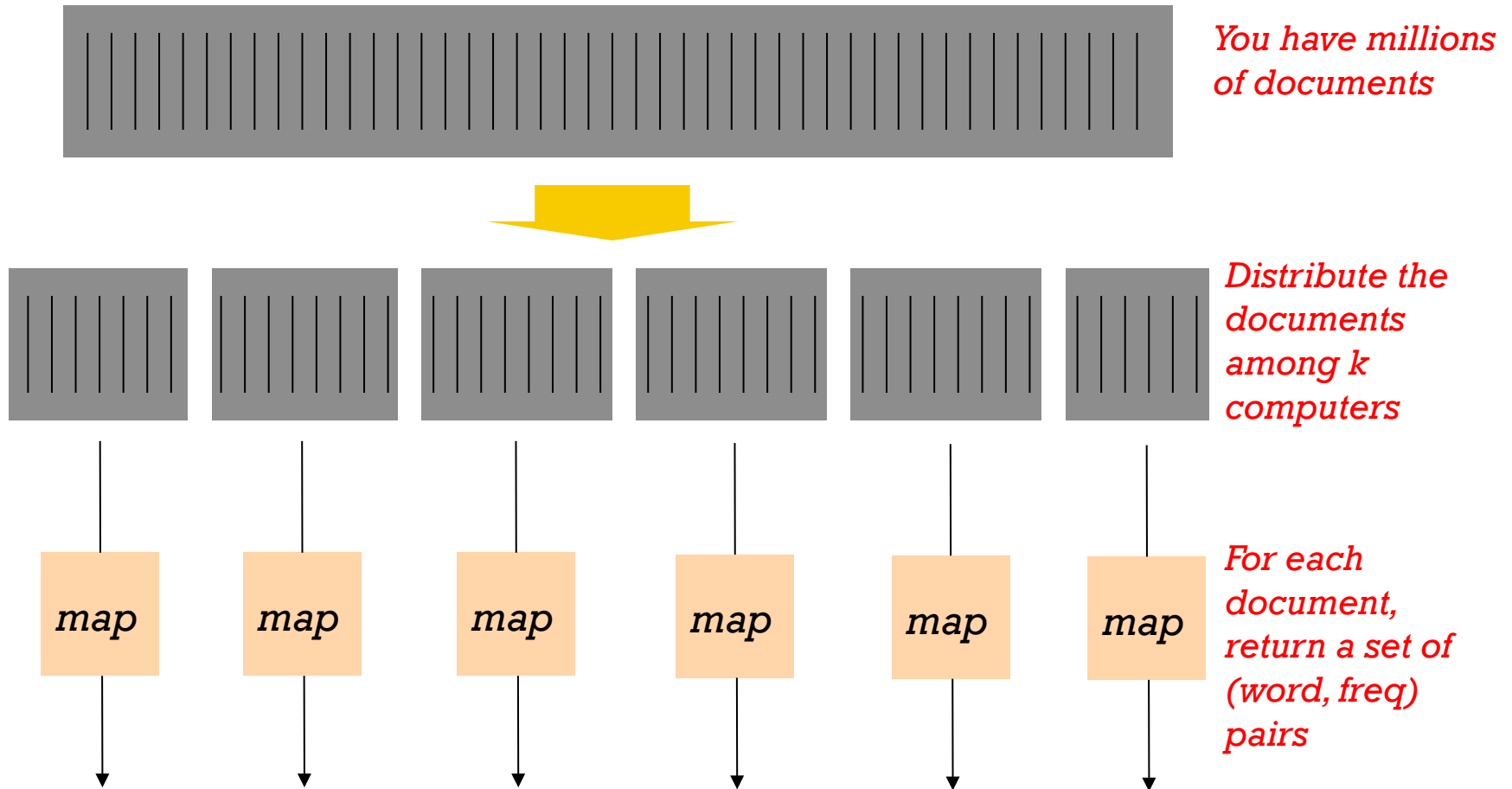
Declaration of Independence



Articles of Confederation

(people, 78)
(government, 123)
(assume, 23)
(history, 38)
...

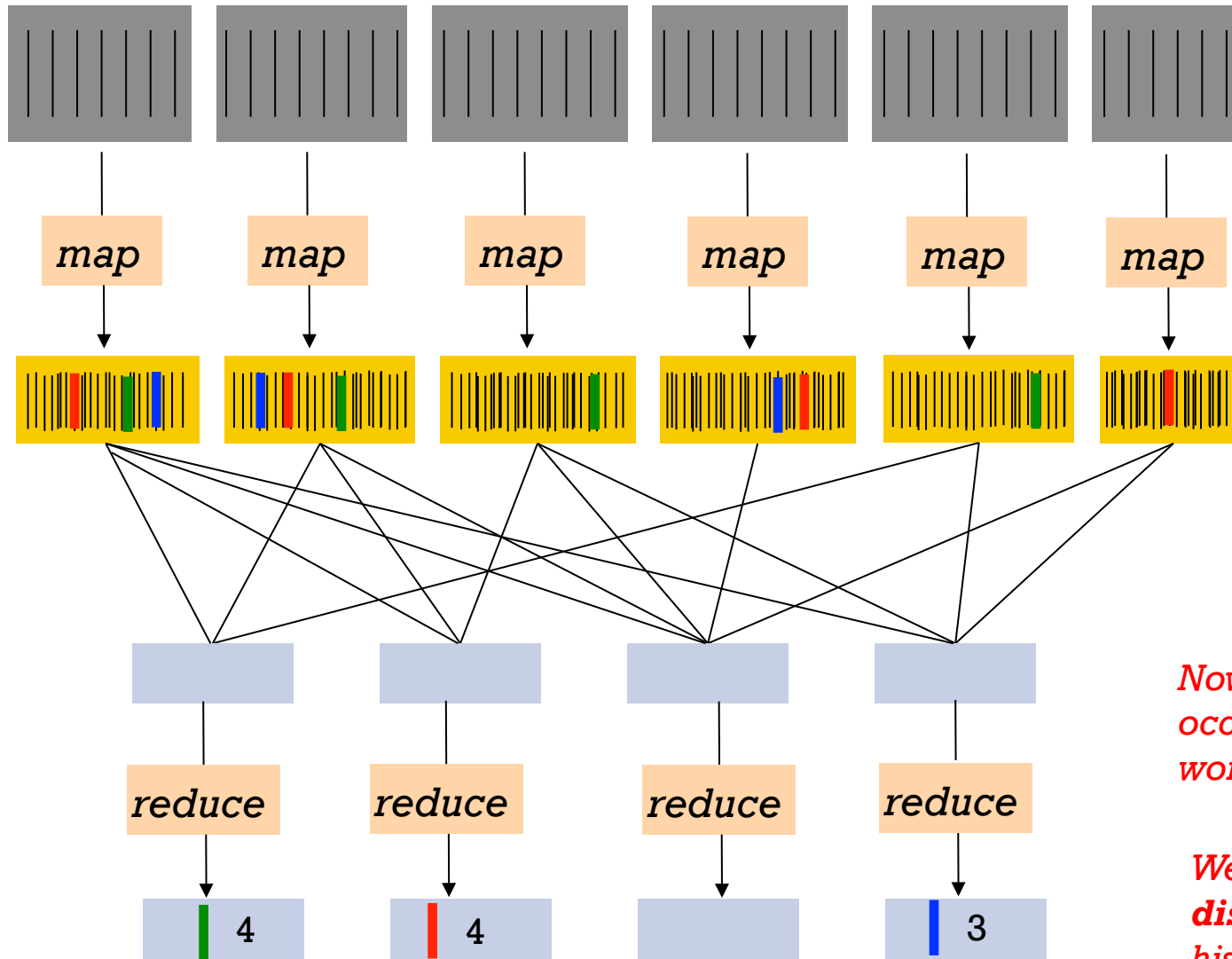
COMPUTE THE WORD FREQUENCY ACROSS 5M DOCUMENTS



Now what?

- But we don't want a bunch of little histograms – we want **one big histogram**.
- How can we make sure that a single computer has access to every occurrence of a given word regardless of which document it appeared in?

COMPUTE THE WORD FREQUENCY ACROSS 5M DOCUMENTS



Distribute the documents among k computers

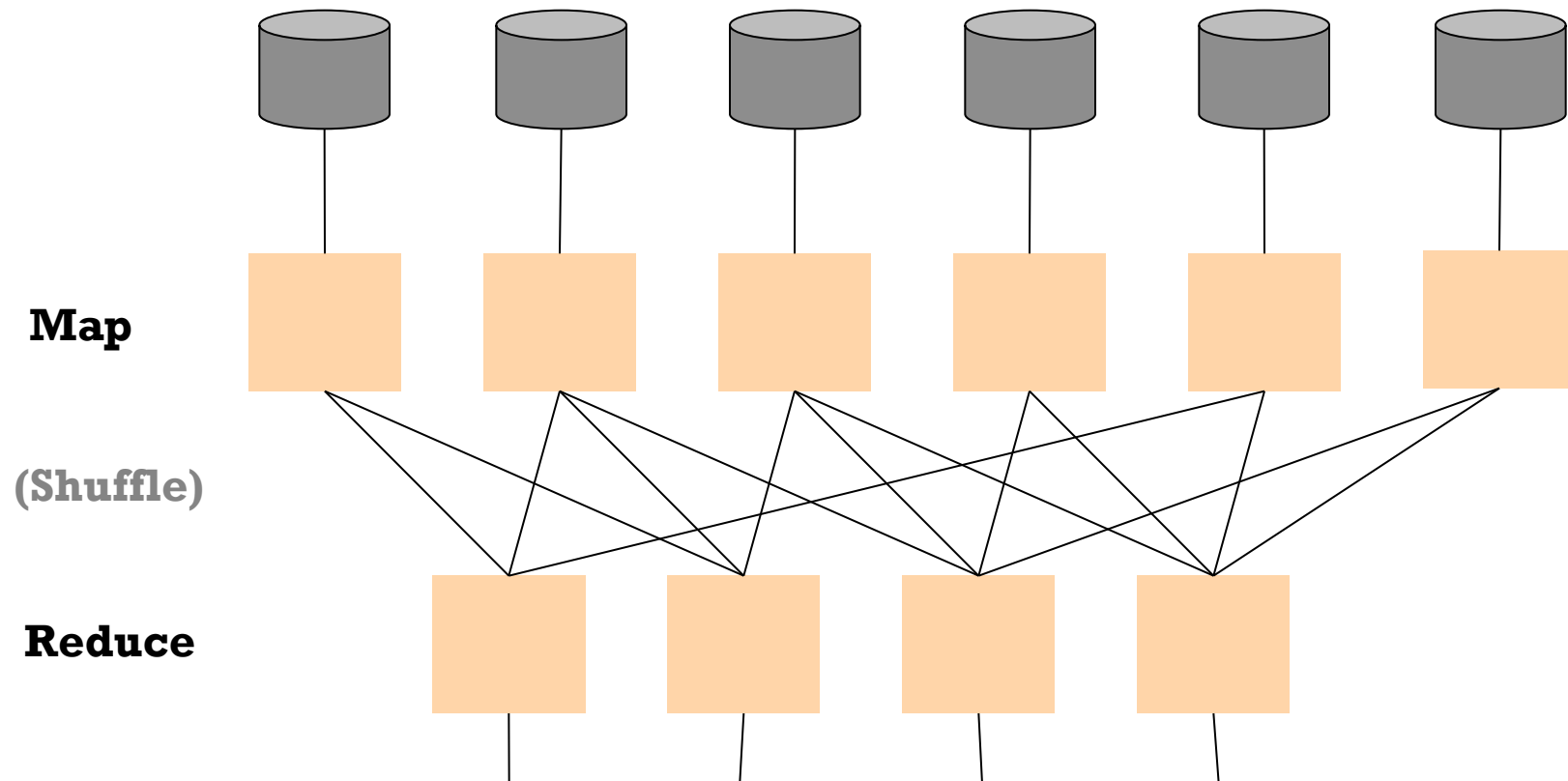
For each document, return a set of (word, freq) pairs

Now we have a big distributed list of sets of word freqs.

Now count the occurrences of each word

We have our distributed histogram

SOME DISTRIBUTED ALGORITHM...



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 output_key, Iterator
intermediate_values):**

// output_key: word

// output_values: ????

int result = 0;

for each v in intermediate_values:

result += v;

Emit(result);

EXAMPLE: DOCUMENT PROCESSING

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

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” , and “small” 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

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

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

Chunk 1

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

Chunk 2

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

Abridged Declaration of Independence

Map Task 1
(204 words)

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

“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 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

(yellow, 17)

(red, 77)

(blue, 107)

(pink, 3)

Map task 2

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)

Reduce tasks

(yellow, 17)
(yellow, 20) (yellow, 37)

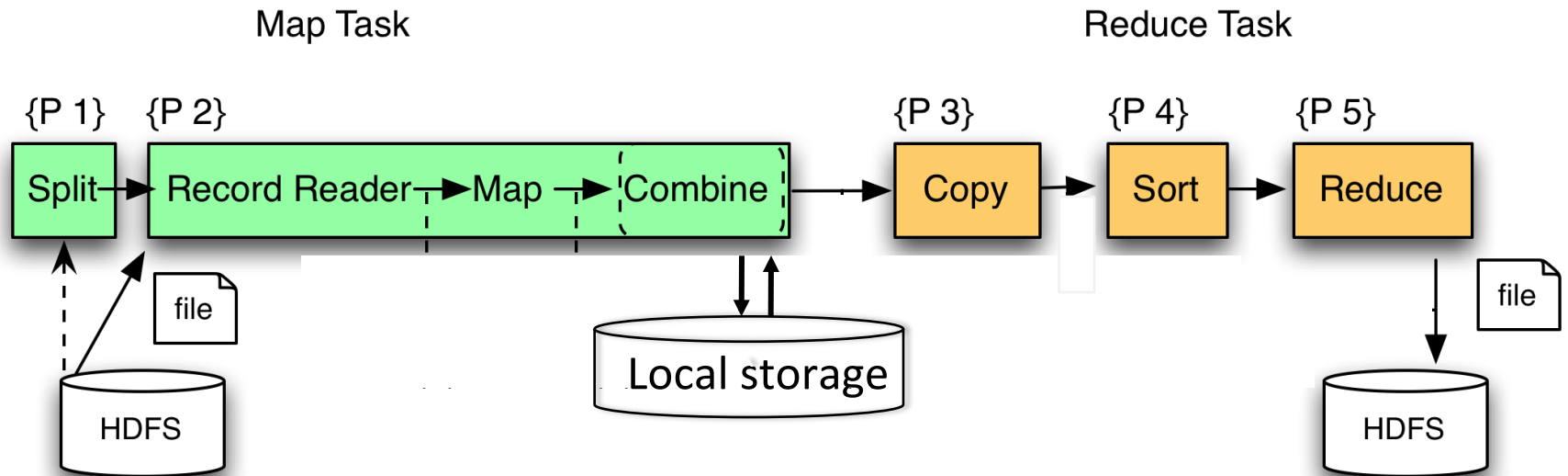
(red, 77)
(red, 71) (red, 148)

(blue, 93)
(blue, 107) (blue, 200)

(pink, 6)
(pink, 3) (pink, 9)

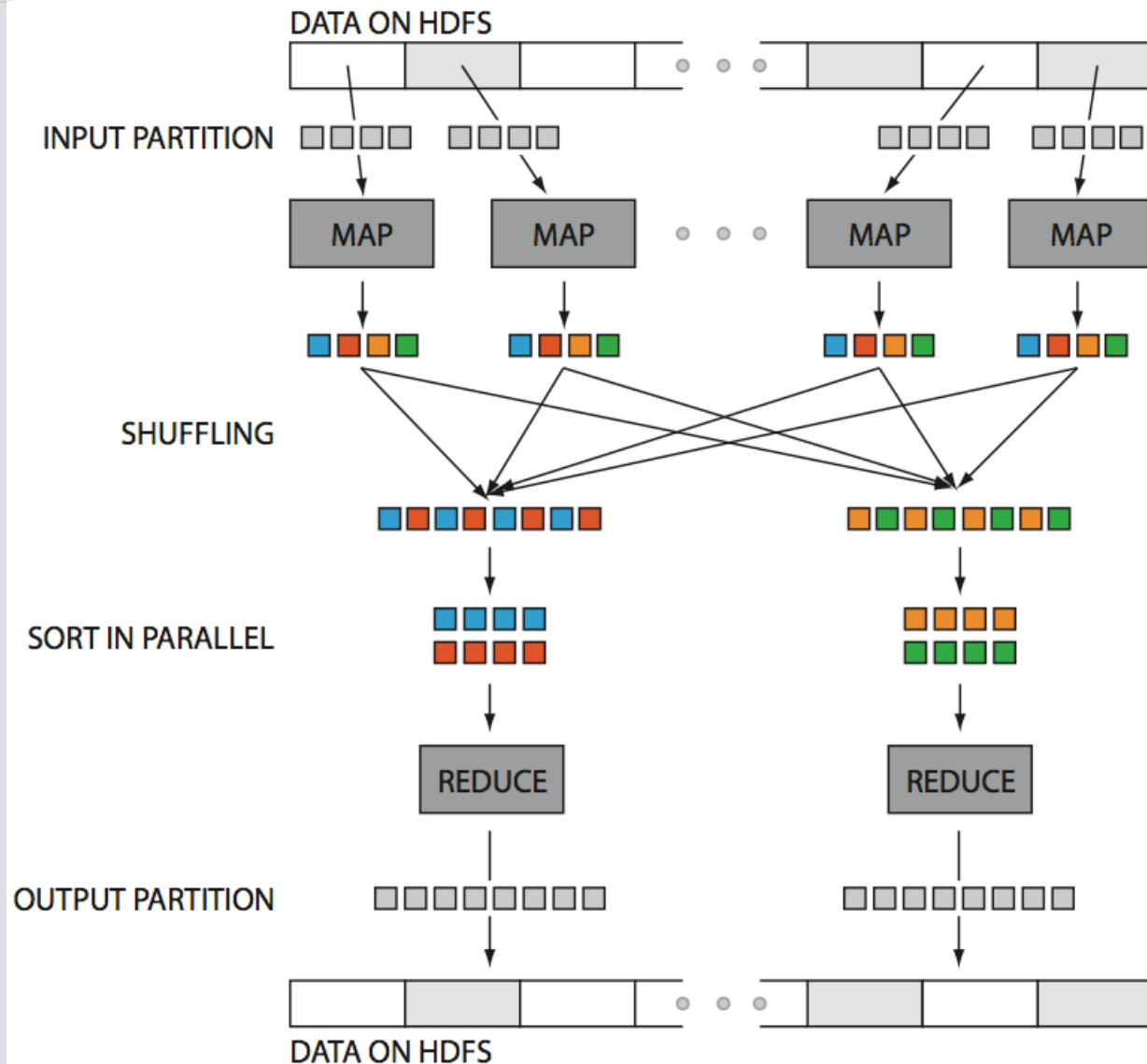
MR PHASES

Each Map and Reduce task has multiple phases:



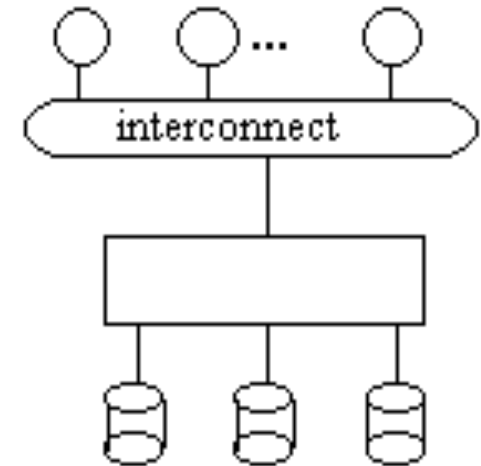
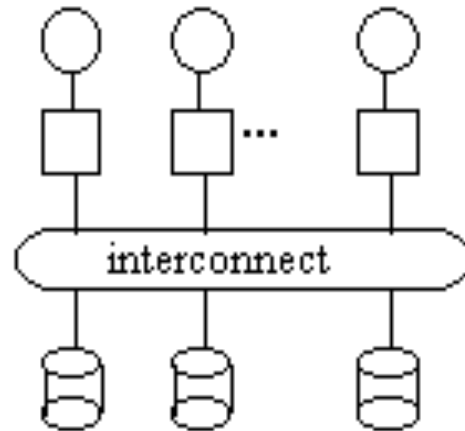
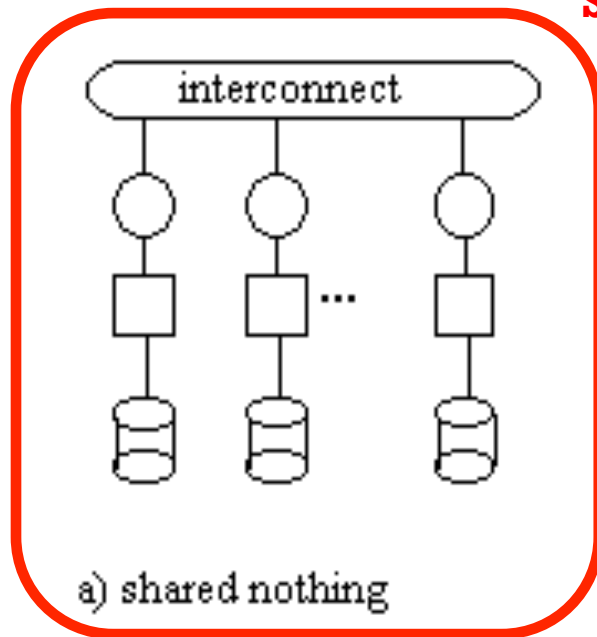
Hadoop in One Slide

src: Huy Vo, NYU Poly






TAXONOMY OF PARALLEL ARCHITECTURES

Scales to 1000s of computers



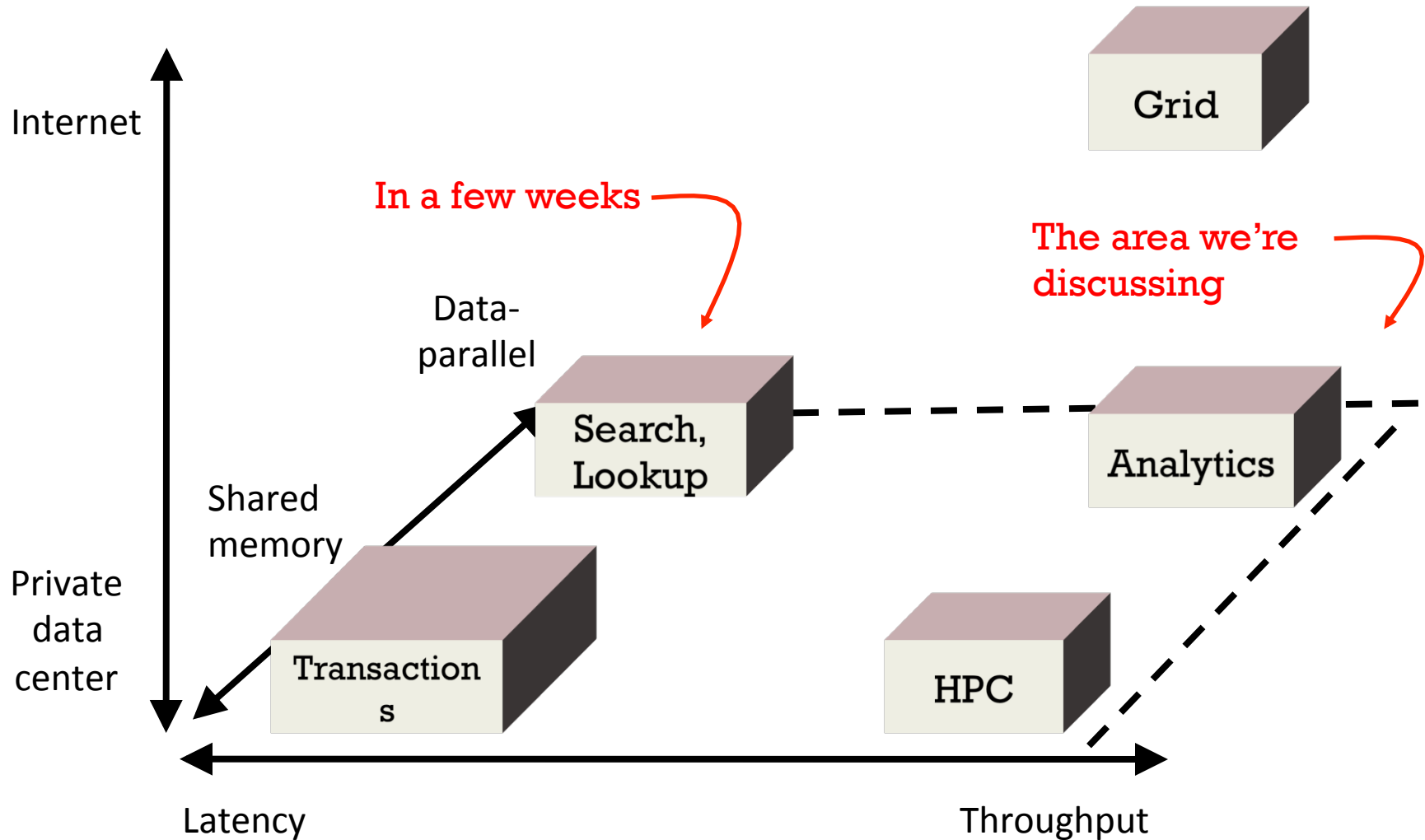
c) shared memory

 = disc  = memory  = processor

Easiest to program,
but \$\$\$\$

Fig. 3.1 Logical multi-processor database designs (diagram after [DEWI92])

DESIGN SPACE



Implementation

- There is one master node
- Master partitions input file into *M splits*, by key
- Master assigns *workers* (=servers) to the *M map tasks*, keeps track of their progress
- Workers write their output to local disk, partition into *R regions*
- Master assigns workers to the *R reduce tasks*
- Reduce workers read regions from the map workers' local disks

LARGE-SCALE DATA PROCESSING

Many tasks process big data, produce big data

Want to use hundreds or thousands of CPUs

- ... but this needs to be easy
- **Parallel databases** exist, but they are expensive, difficult to set up, and do not necessarily scale to hundreds of nodes.

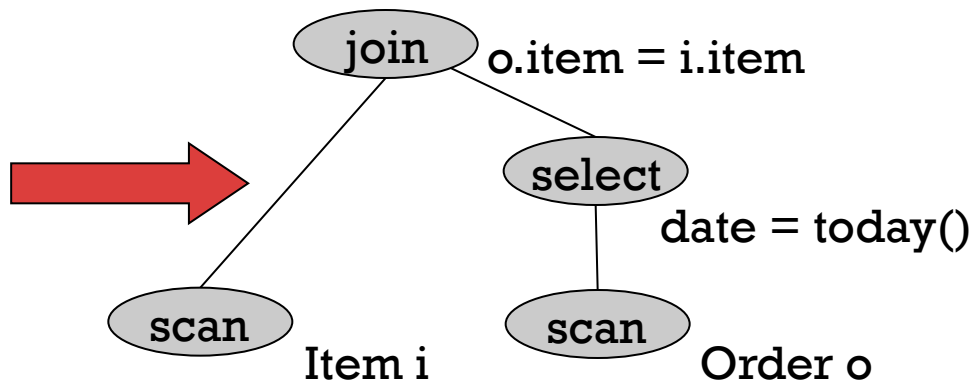
MapReduce is a *lightweight* framework, providing:

- **Automatic parallelization and distribution**
- **Fault-tolerance**
- **I/O scheduling**
- **Status and monitoring**

KEY IDEA: DECLARATIVE LANGUAGES

Find all orders from today, along with the items ordered

SELECT *
FROM Order o, Item i
WHERE o.item = i.item
AND o.date = today()



TWO NOTIONS OF PARALLEL QUERY PROCESSING

“Distributed Query”

- Rewrite the query as a union of subqueries
- Workers communicate through standard interfaces, so compatible with federated, heterogeneous, or distributed databases

“Parallel Query”

- Each operator is implemented with a parallel algorithm

DISTRIBUTED QUERY EXAMPLE

```
CREATE VIEW Sales AS
```

```
SELECT * FROM JanSales  
UNION ALL
```

```
SELECT * FROM FebSales  
UNION ALL
```

```
SELECT * FROM MarSales
```

```
CREATE TABLE MarSales(  
    OrderID          INT,  
    CustomerID       INT          NOT NULL,  
    OrderDate        DATETIME     NULL  
        CHECK (DATEPART(mm, OrderDate) = 3),  
    CONSTRAINT OrderIDMonth PRIMARY KEY(OrderID)  
)
```

DISTRIBUTED FILE SYSTEM (DFS)

For very large files: TBs, PBs

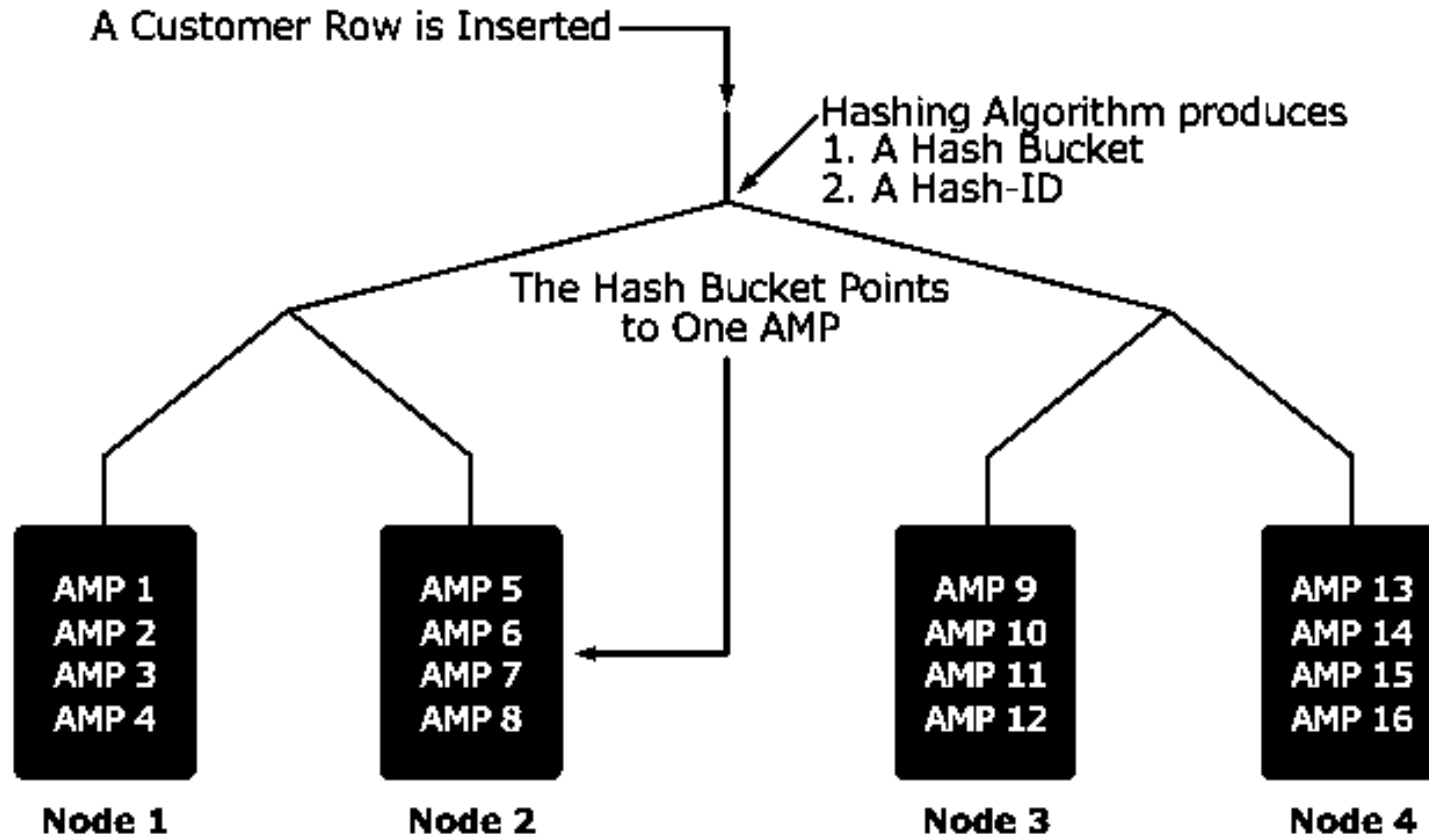
Each file is partitioned into *chunks*, typically 64MB

Each chunk is replicated several times (≥ 3), on different racks, for fault tolerance

Implementations:

- Google's DFS: **GFS**, proprietary
- Hadoop's DFS: **HDFS**, open source

PARALLEL QUERY EXAMPLE: TERADATA

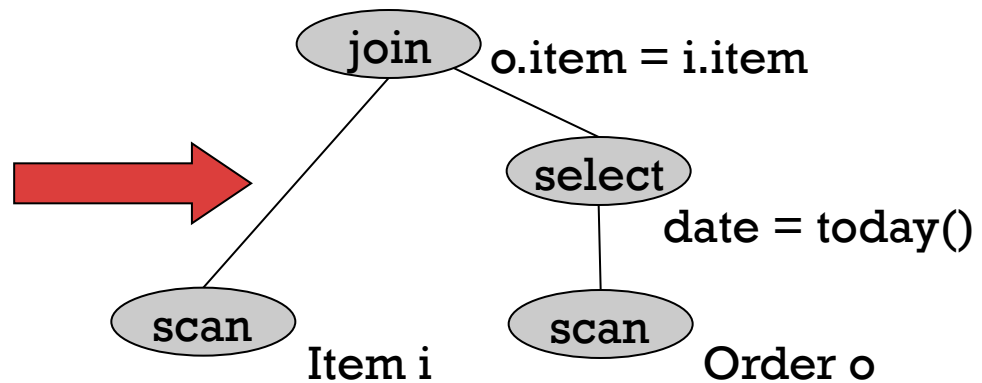


AMP = unit of parallelism

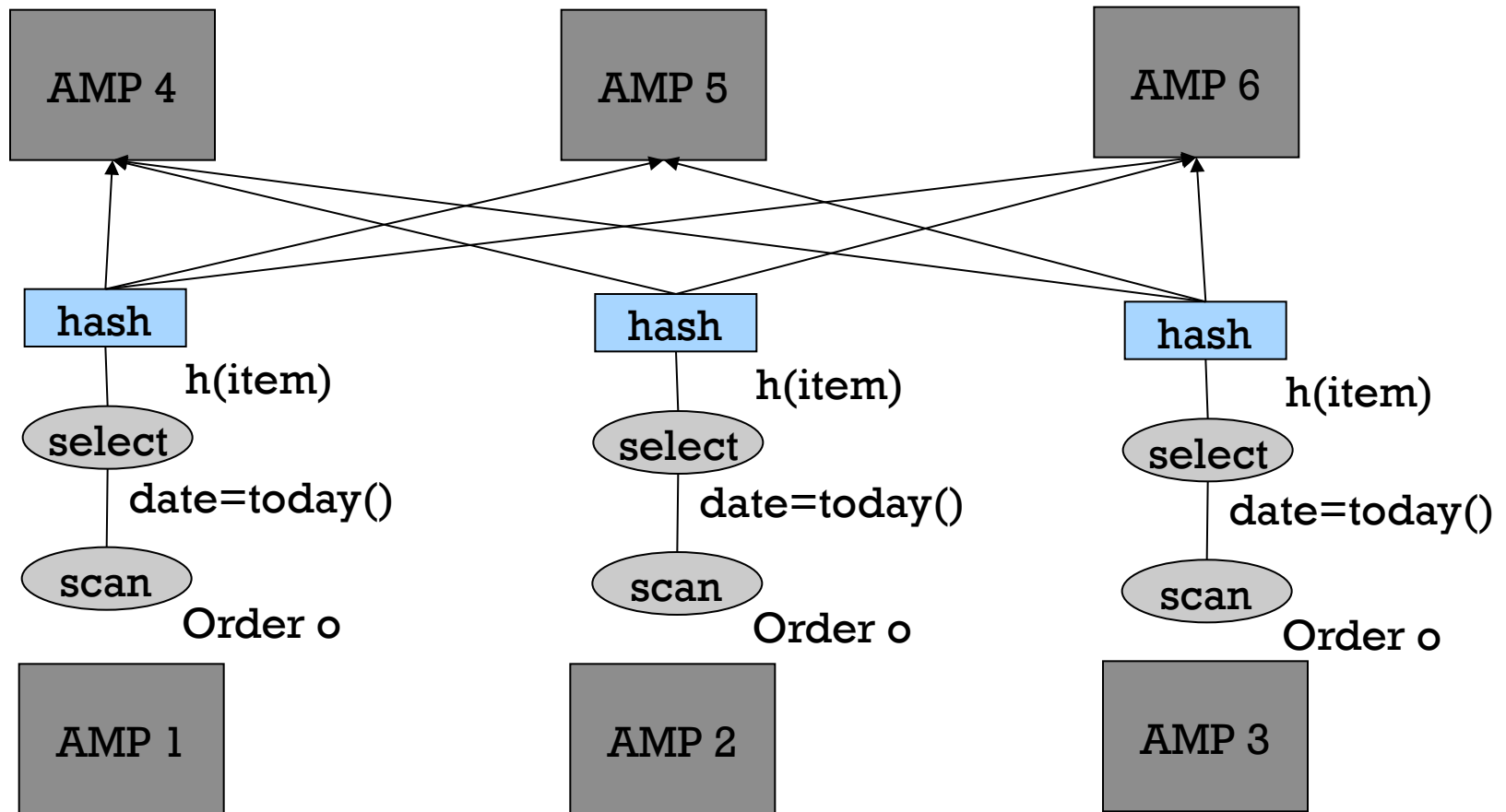
EXAMPLE SYSTEM: TERADATA

Find all orders from today, along with the items ordered

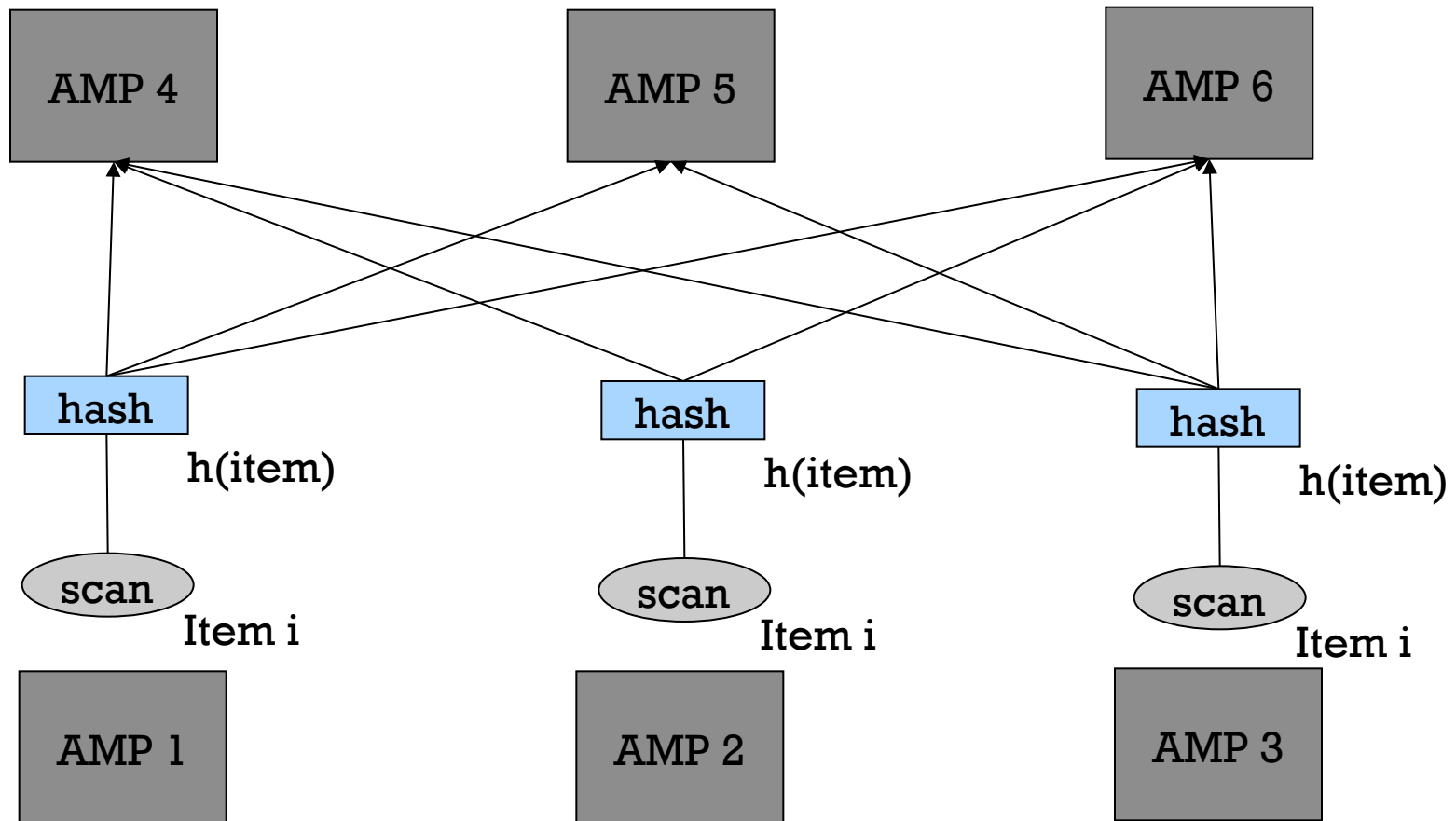
SELECT *
FROM Orders o, Lines i
WHERE o.item = i.item
AND o.date = today()



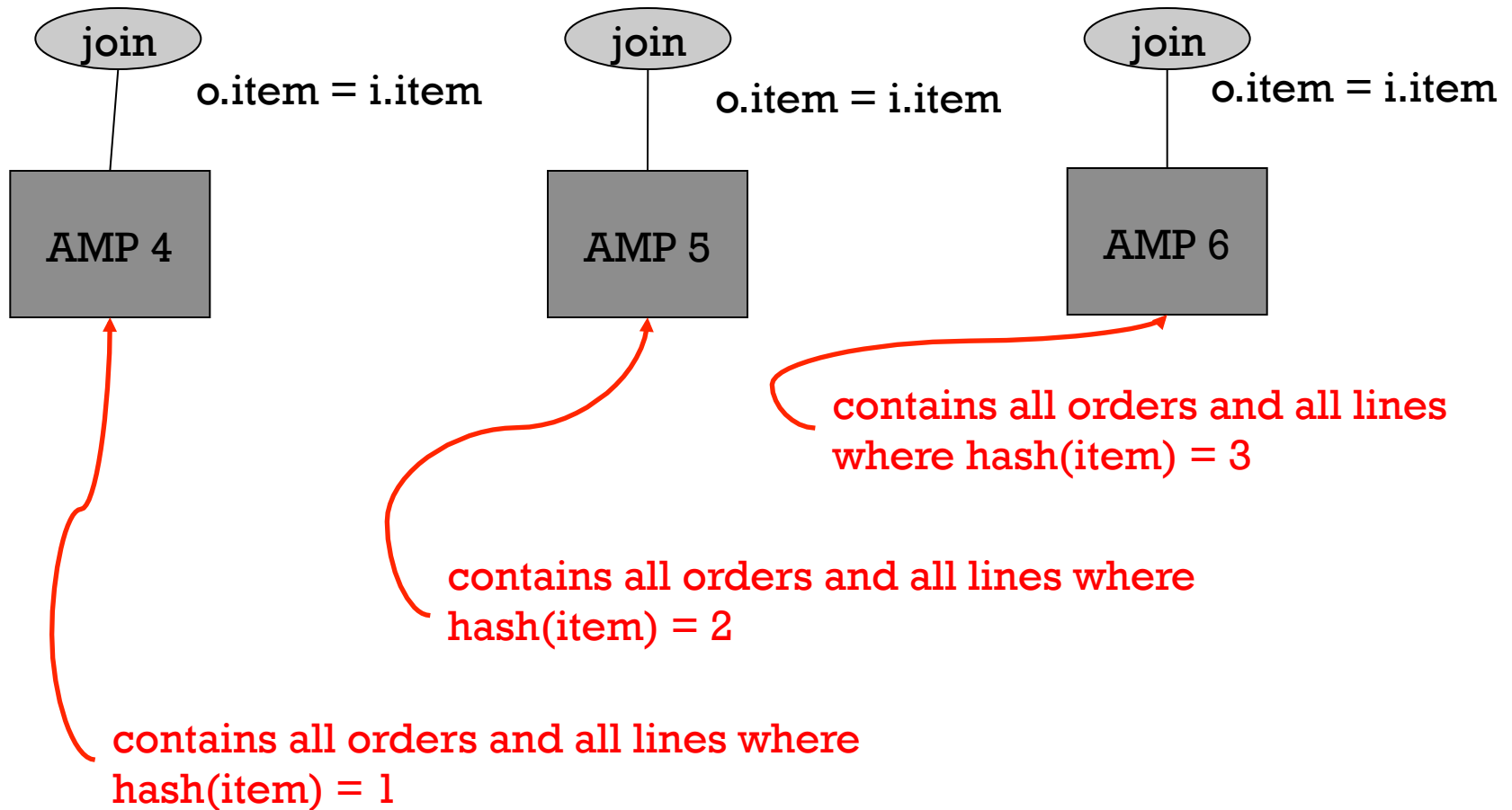
EXAMPLE SYSTEM: TERADATA



EXAMPLE SYSTEM: TERADATA



EXAMPLE SYSTEM: TERADATA



MAPREDUCE CONTEMPORARIES

Dryad (Microsoft)

- Relational Algebra

Pig (Yahoo)

- Near Relational Algebra over MapReduce

HIVE (Facebook)

- SQL over MapReduce

Cascading

- Relational Algebra

Clustera

- U of Wisconsin

Hbase

- Indexing on HDFS

MAPREDUCE VS RDBMS

RDBMS

- Declarative query languages
- Schemas
- Logical Data Independence
- Indexing
- Algebraic Optimization
- Caching/Materialized Views
- *ACID/Transactions*

DryadLINQ, Pig, HIVE

HIVE, Pig, DryadLINQ

Hbase

Pig, (Dryad, HIVE)

MapReduce

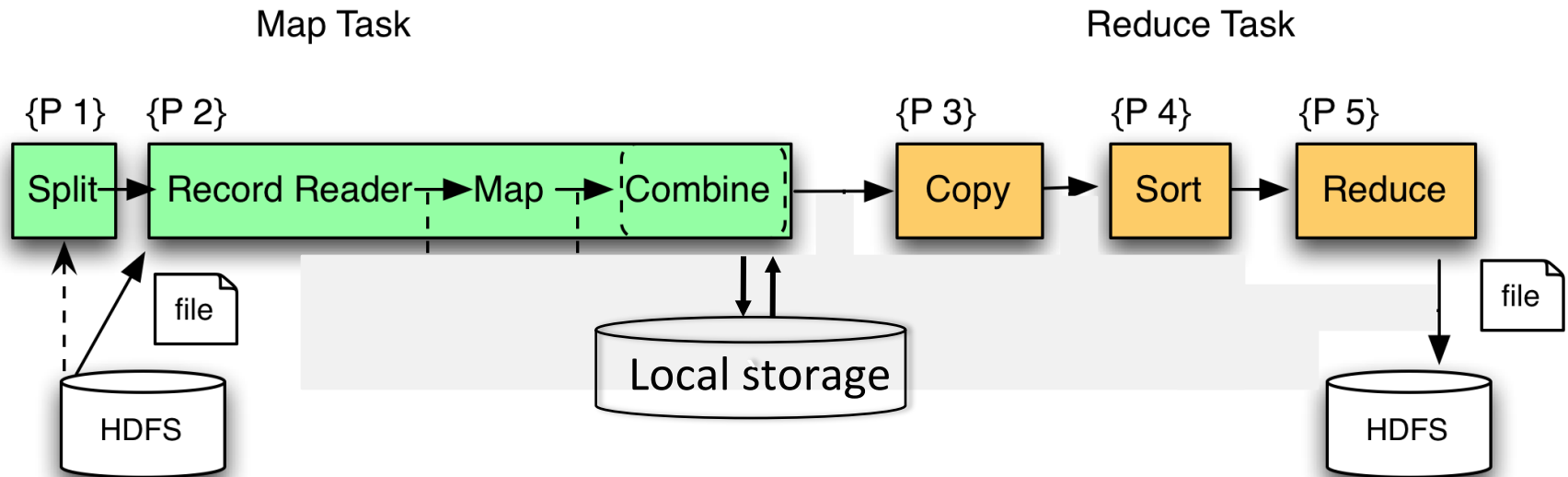
- High Scalability
- Fault-tolerance
- “One-person deployment”

COMPARISON

	Data Model	Prog. Model	Services
GPL	*	*	Typing (maybe)
Workflow	*	dataflow	typing, provenance, scheduling, caching, task parallelism, reuse
Relational Algebra	Relations	Select, Project, Join, Aggregate, ...	optimization, physical data independence, data parallelism
MapReduce	[(key,value)]	Map, Reduce	massive data parallelism, fault tolerance
MS Dryad	IQueryable, IEnumerable	RA + Apply + Partitioning	typing, massive data parallelism, fault tolerance
MPI	Arrays/ Matrices	70+ ops	data parallelism, full control

MAP REDUCE PHASES

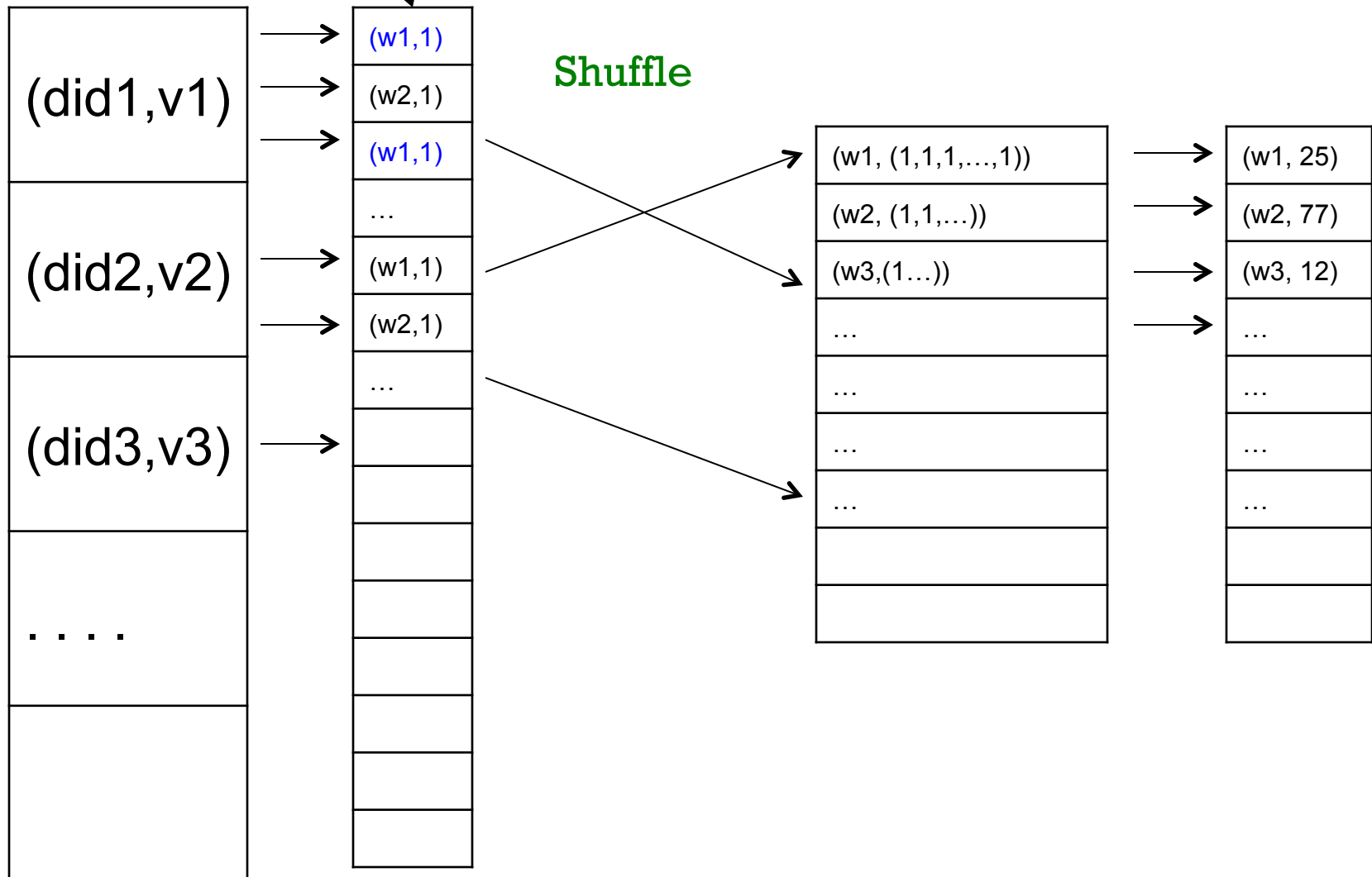
Each Map and Reduce task has multiple phases:



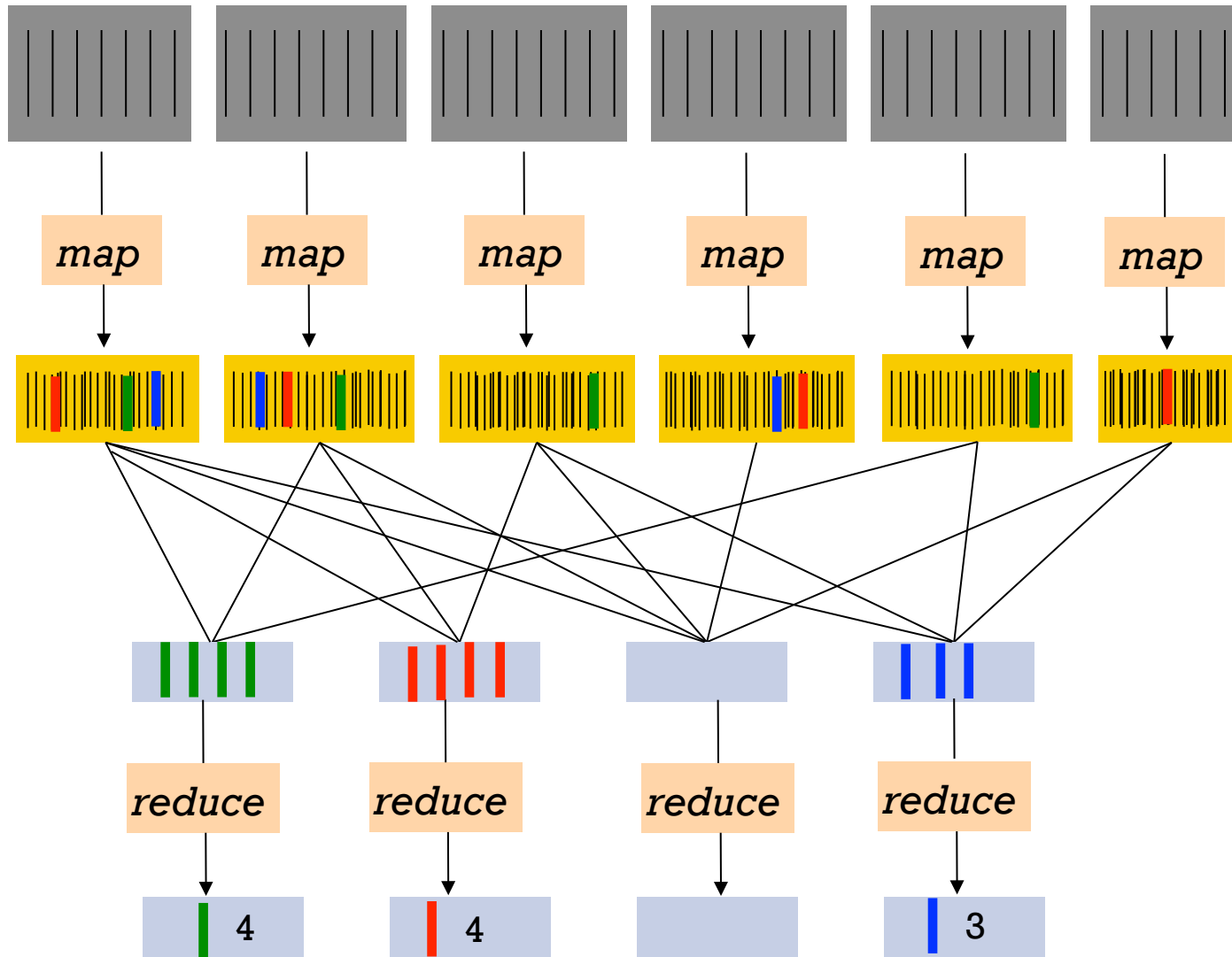
MAP

*Same word appears
twice. Why not just send
(w1, 2)?*

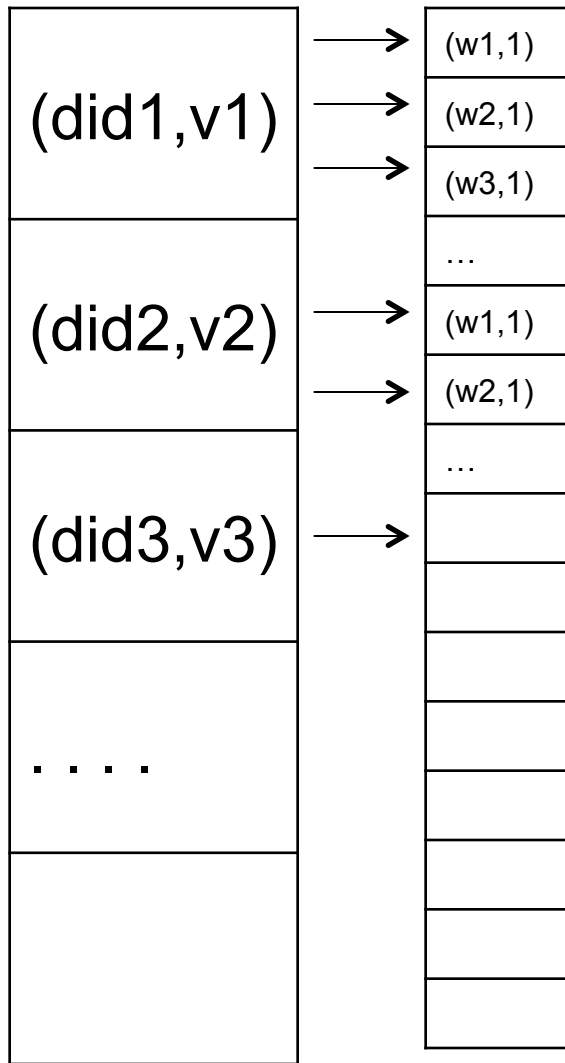
REDUCE



COUNT WORD OCCURRENCES ACROSS ALL DOCUMENTS

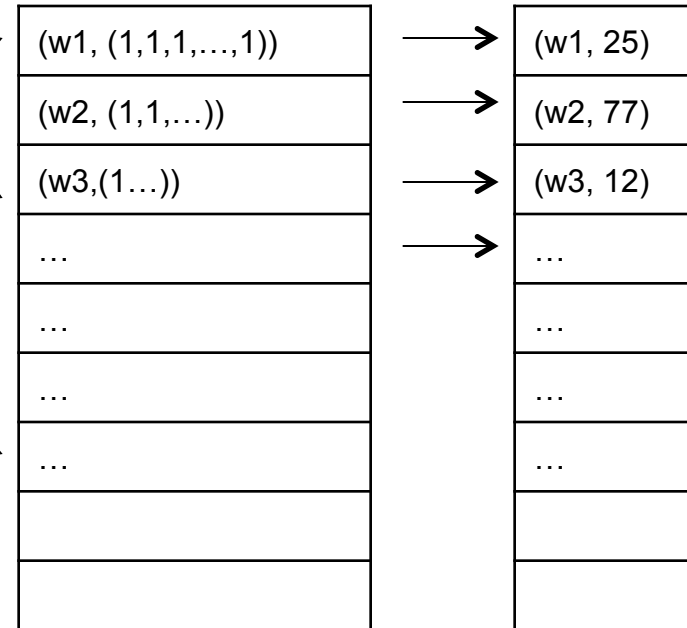


MAP



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 output_key, Iterator, intermediate_values):
```

```
// output_key: word
```

```
// output_values: ????
```

```
  int result = 0;
```

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

```
    result += v;
```

```
  Emit(result);
```

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)
...

MORE EXAMPLES: RELATIONAL JOIN

Employee

Name	SSN
Sue	9999999999
Tony	7777777777

Assigned Departments

EmpSSN	DepName
9999999999	Accounts
7777777777	Sales
7777777777	Marketing

Employee ⋈ Assigned Departments

Name	SSN	EmpSSN	DepName
Sue	9999999999	9999999999	Accounts
Tony	7777777777	7777777777	Sales
Tony	7777777777	7777777777	Marketing

RELATIONAL JOIN IN MAPREDUCE: BEFORE MAP PHASE

Employee

Name	SSN
Sue	999999999
Tony	777777777

Key idea: Lump all the tuples together into one dataset

Assigned Departments

EmpSSN	DepName
999999999	Accounts
777777777	Sales
777777777	Marketing



Employee, Sue, 999999999
Employee, Tony, 777777777
Department, 999999999, Accounts
Department, 777777777, Sales
Department, 777777777, Marketing

What is this for?

RELATIONAL JOIN IN MAPREDUCE: MAP PHASE

Employee, Sue, 999999999

Employee, Tony, 777777777

Department, 999999999, Accounts

Department, 777777777, Sales

Department, 777777777, Marketing



key=999999999, value=(Employee, Sue, 999999999)

key=777777777, value=(Employee, Tony, 777777777)

key=999999999, value=(Department, 999999999, Accounts)

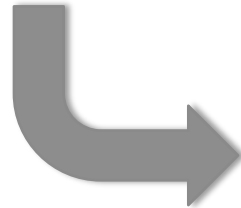
key=777777777, value=(Department, 777777777, Sales)

key=777777777, value=(Department, 777777777, Marketing)

why do we use this as the key?

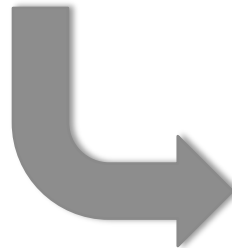
RELATIONAL JOIN IN MAPREDUCE: REDUCE PHASE

key=999999999, **values**=[(Employee, Sue, 999999999),
(Department, 999999999, Accounts)]



Sue, 999999999, 999999999, Accounts

key=777777777, **values**=[(Employee, Tony, 777777777),
(Department, 777777777, Sales),
(Department, 777777777, Marketing)]



Tony, 777777777, 777777777, Sales
Tony, 777777777, 777777777, Marketing

RELATIONAL JOIN IN MAPREDUCE, AGAIN

Order(orderid, account, date)


1, aaa, d1
2, aaa, d2
3, bbb, d3

LineItem(orderid, itemid, qty)

1, 10, 1
1, 20, 3
2, 10, 5
2, 50, 100
3, 20, 1

Map

tagged with
relation
name



Order

1, aaa, d1 → 1 : "Order", (1,aaa,d1)
2, aaa, d2 → 2 : "Order", (2,aaa,d2)
3, bbb, d3 → 3 : "Order", (3,bbb,d3)

Line

1, 10, 1 → 1 : "Line", (1, 10, 1)
1, 20, 3 → 1 : "Line", (1, 20, 3)
2, 10, 5 → 2 : "Line", (2, 10, 5)
2, 50, 100 → 2 : "Line", (2, 50, 100)
3, 20, 1 → 3 : "Line", (3, 20, 1)

Reducer for key 1

"Order", (1,aaa,d1)
"Line", (1, 10, 1)
"Line", (1, 20, 3)



(1, aaa, d1, 1, 10, 1)
(1, aaa, d1, 1, 20, 3)

SIMPLE SOCIAL NETWORK ANALYSIS: COUNT FRIENDS

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

MATRIX MULTIPLY IN MAPREDUCE

$$\mathbf{C} = \mathbf{A} \times \mathbf{B}$$

A has dimensions **L,M**

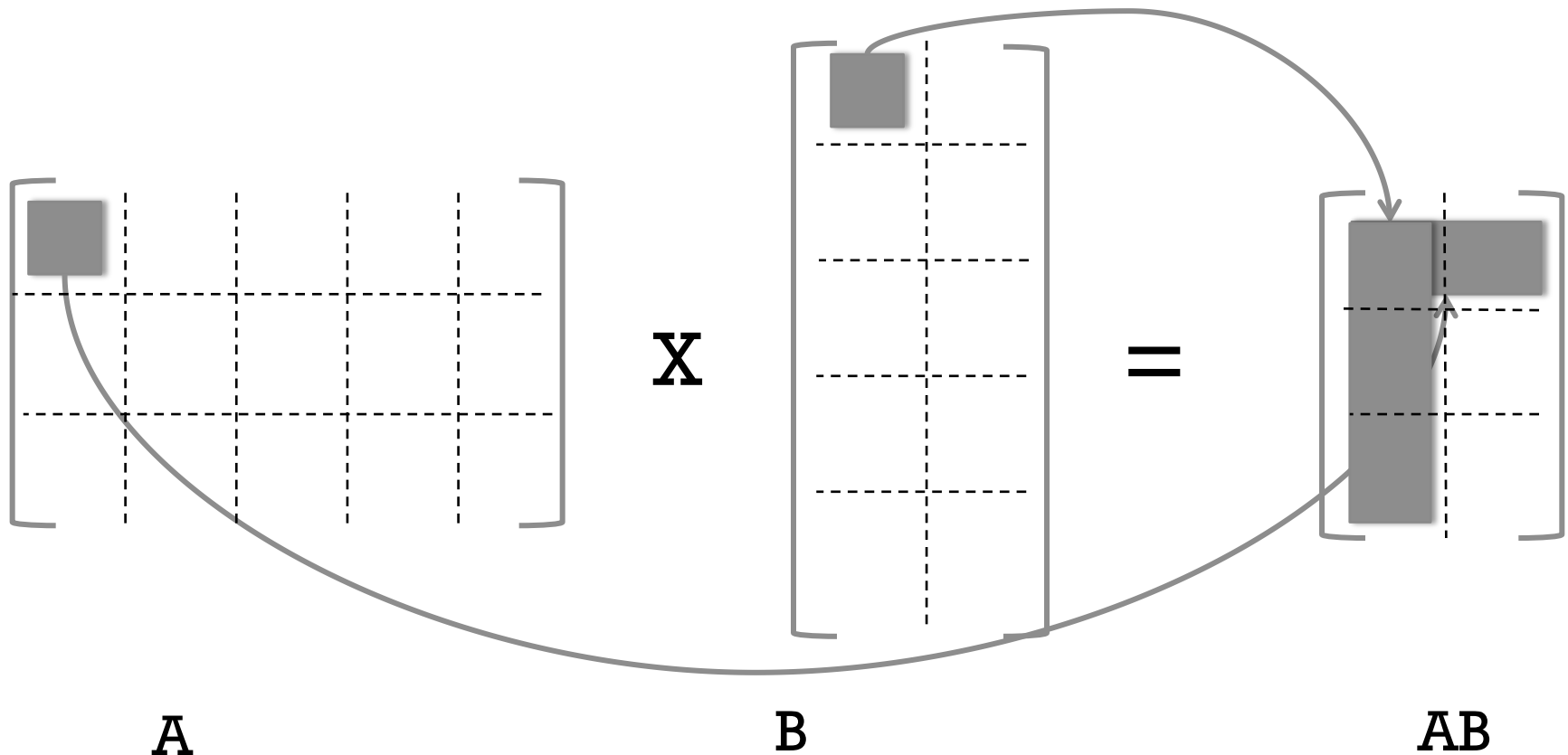
B has dimensions **M,N**

In the map phase:

- for each element (i,j) of **A**, emit $((i,k), A[i,j])$ for k in $1..N$
- for each element (j,k) of **B**, emit $((i,k), B[j,k])$ for i in $1..L$

In the reduce phase, emit

- key = (i,k)
- value = $\text{Sum}_j (A[i,j] * B[j,k])$



- One reducer per output cell
-
- Each reducer computes: $\sum_j A_{i,j} B_{j,k}$

CLUSTER COMPUTING

Large number of commodity servers, connected by high speed, commodity network

Rack: holds a small number of servers

Data center: holds many racks

CLUSTER COMPUTING

Massive parallelism:

- 100s, or 1000s, or 10000s servers
- Many hours

Failure:

- If medium-time-between-failure is 1 year
- Then 10000 servers have one failure / hour

DISTRIBUTED FILE SYSTEM (DFS)

For very large files: TBs, PBs

Each file is partitioned into *chunks*, typically 64MB

Each chunk is replicated several times (≥ 3), on different racks, for fault tolerance

Implementations:

- Google's DFS: **GFS**, proprietary
- Hadoop's DFS: **HDFS**, open source

MAP REDUCE VS. DATABASES

HADOOP VS. RDBMS

Comparison of 3 systems

- Hadoop
- Vertica (a column-oriented database)
- DBMS-X (a row-oriented database)
 - rhymes with “schmoracle”

Qualitative

- Programming model, ease of setup, features, etc.

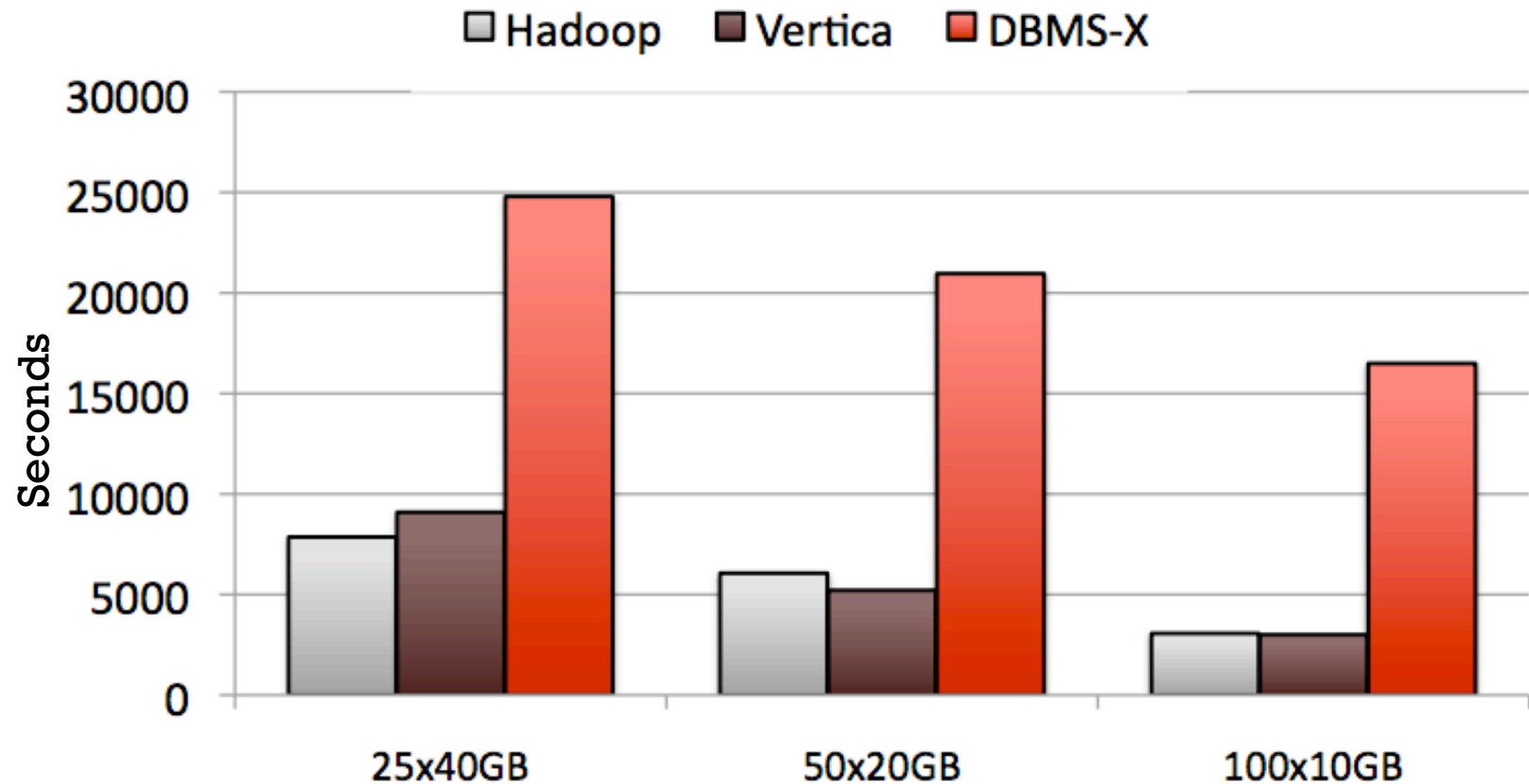
Quantitative

- Data loading, different types of queries

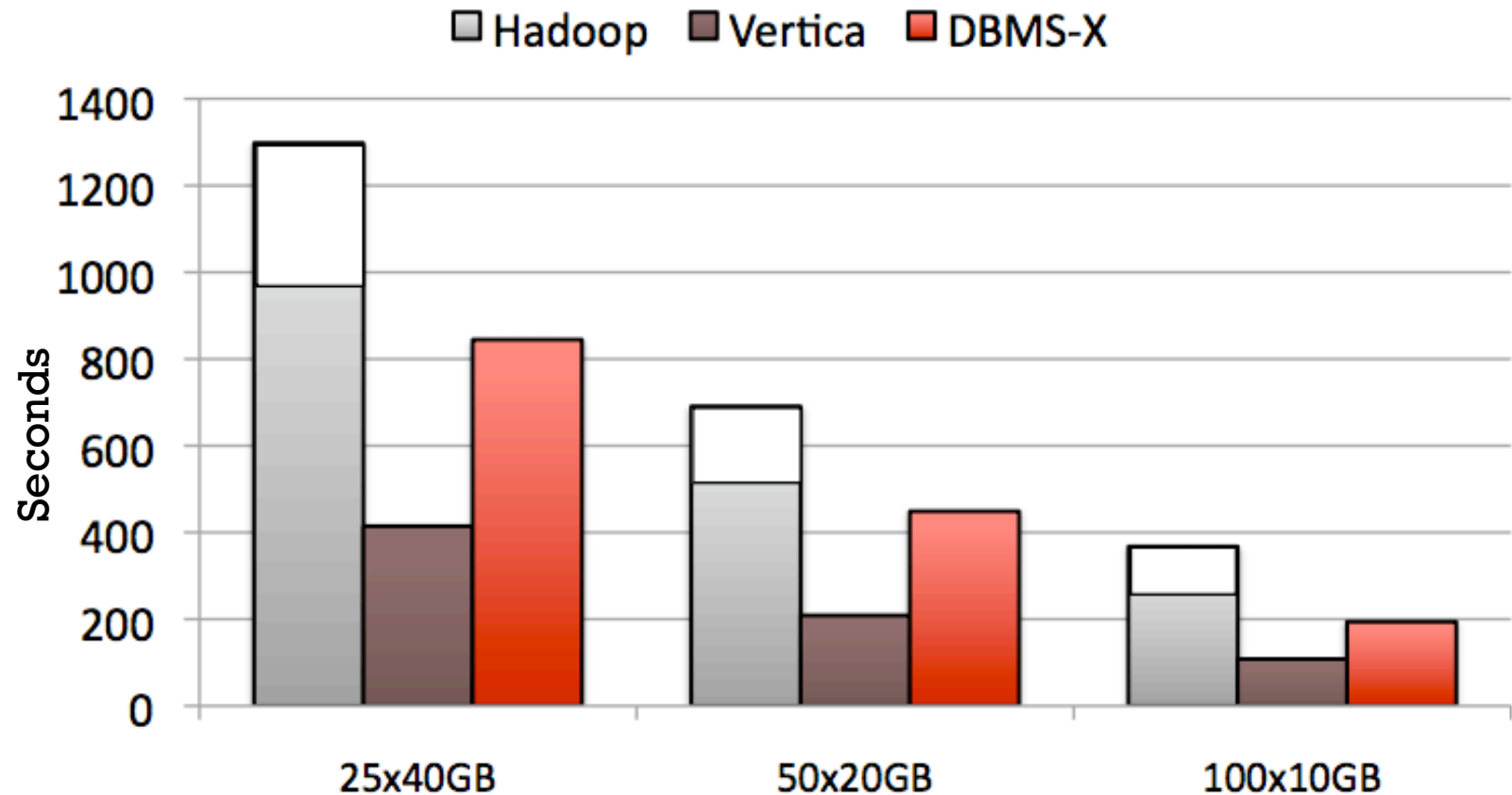
GREP TASK

- **Find 3-byte pattern in 100-byte record**
 - 1 match per 10,000 records
- **Data set:**
 - 10-byte unique key, 90-byte value
 - 1TB spread across 25, 50, or 100 nodes
 - 10 billion records
- **Original MR Paper (Dean et al. 2004)**

GREP TASK LOADING RESULTS

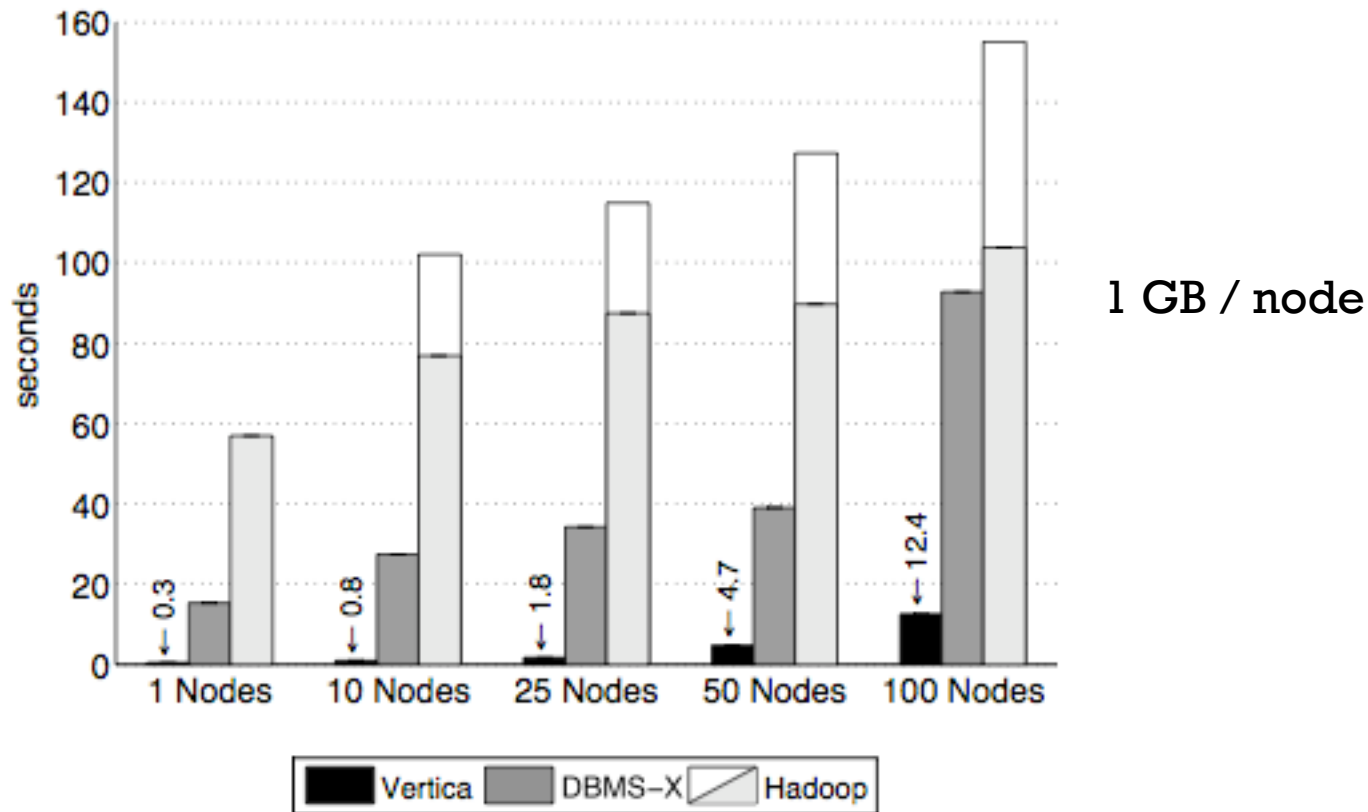


GREP TASK EXECUTION RESULTS



SELECTION TASK

```
SELECT pageURL, pageRank  
FROM Rankings WHERE pageRank > X
```



ANALYTICAL TASKS

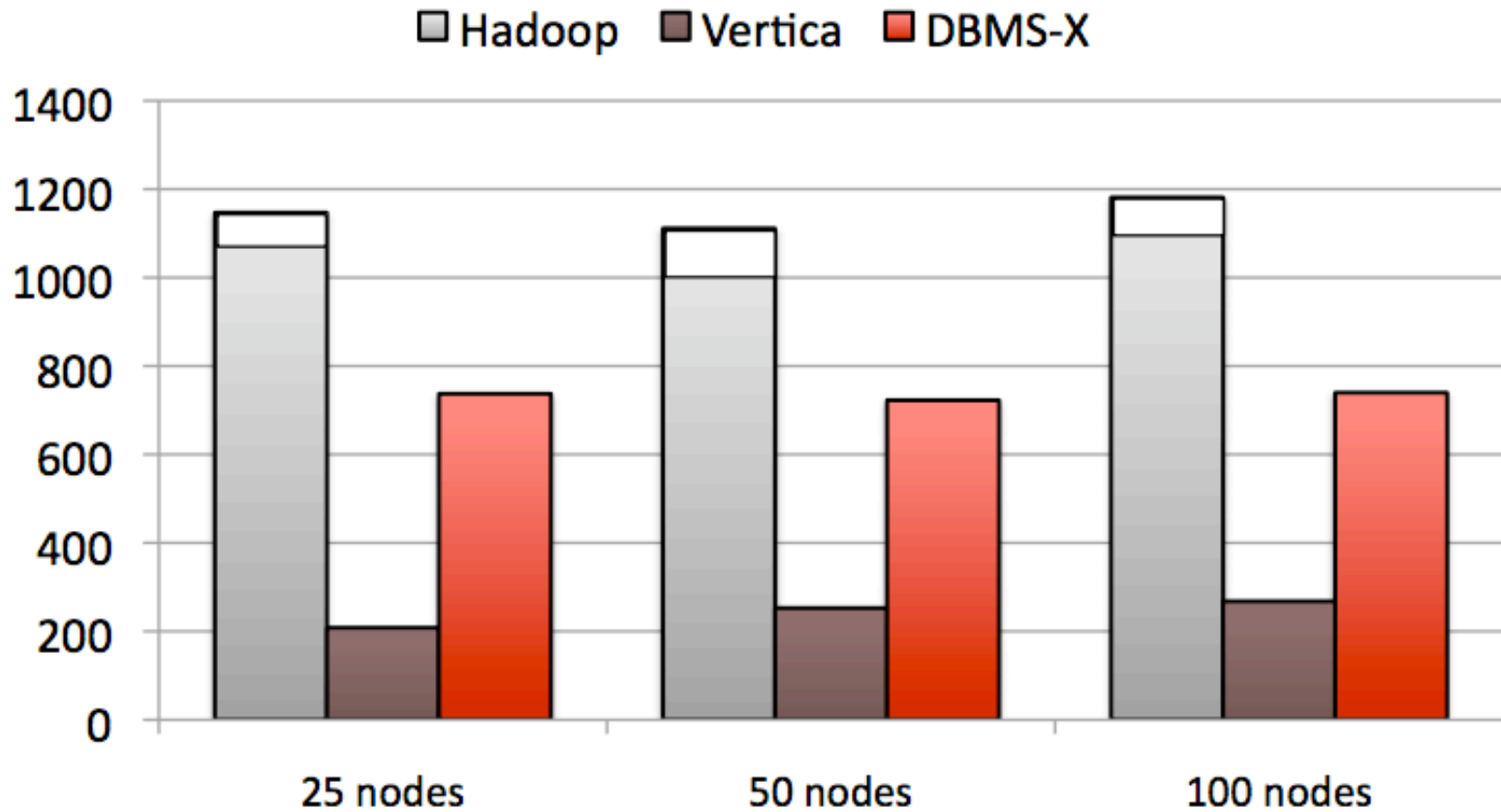
- **Simple web programming schema**
- **Data set**
 - 600k HTML Documents (6GB/node)
 - 155 million UserVisit records (20GB/node)
 - 18 million Rankings records (1GB/node)

AGGREGATE TASK

- **Simple query to find adRevenue by IP prefix**

```
SELECT SUBSTR(sourceIP, 1, 7),  
       SUM (adRevenue)  
FROM userVisits  
GROUP BY SUBSTR (sourceIP, 1, 7)
```

AGGREGATE TASK RESULTS



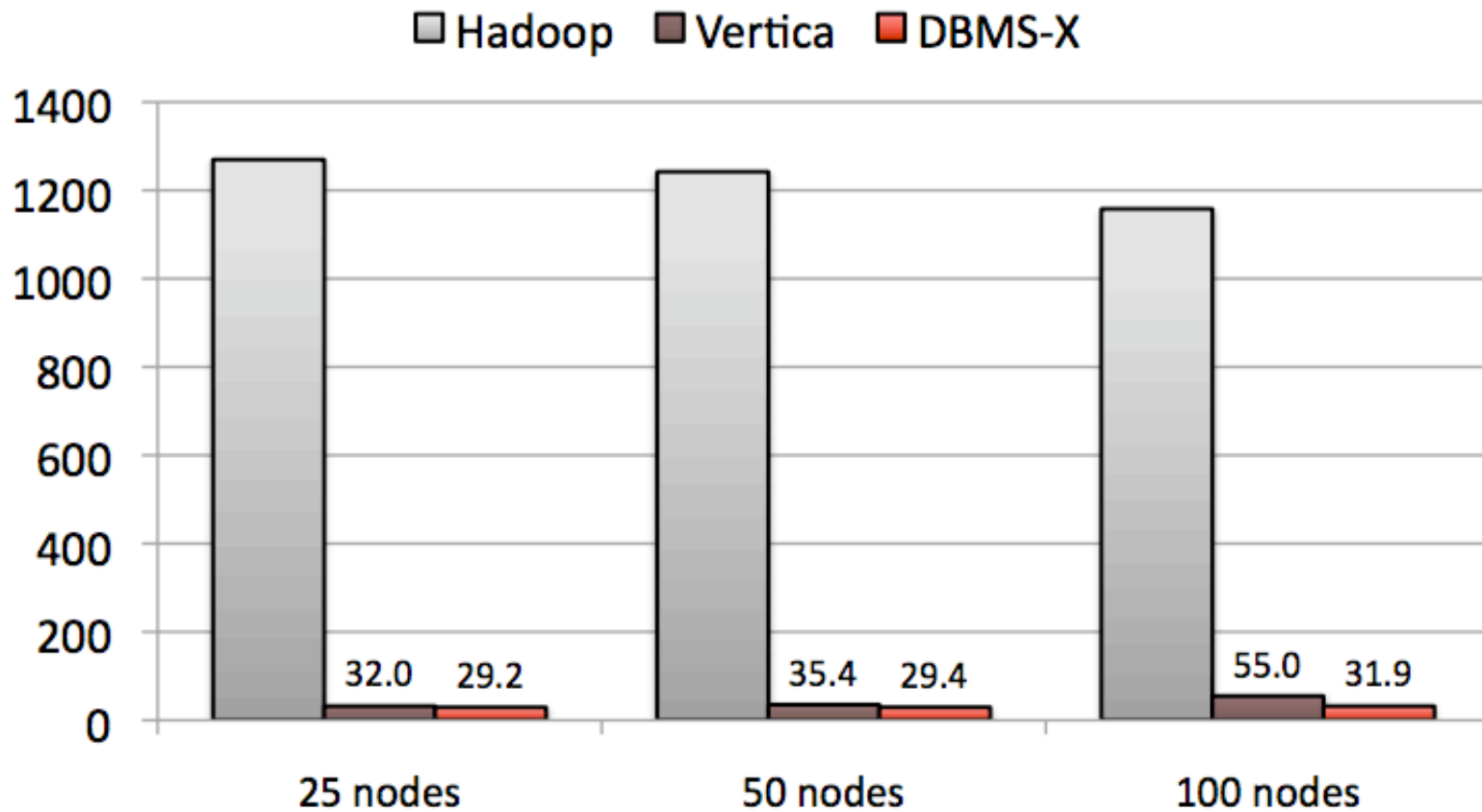
JOIN TASK

- **Find the sourceIP that generated the most adRevenue along with its average pageRank**
- **Implementations:**
 - DMPSs – complex SQL using temporary table
 - MapReduce – Three separate MR programs

JOIN TASK

```
SELECT INTO TempsourceIP,  
           AVG(pageRank)as avgPageRank,  
           SUM(adRevenue)as totalRevenue  
FROM RankingsAS R  
   , UserVisitsAS UV  
WHERE R.pageURL = UV.destURL  
AND UV.visitDate  
     BETWEEN '2000-01-15'  
     AND '2000-01-22'  
GROUP BY UV.sourceIP;  
  
SELECT sourceIP,  
       totalRevenue,  
       avgPageRank  
FROM Temp  
ORDER BY totalRevenueDESC  
LIMIT 1;
```

JOIN TASK RESULTS



PROBLEMS WITH THIS ANALYSIS?

Other ways to avoid sequential scans?

Fault-tolerance in large clusters?

Tasks that cannot be expressed as queries?

GOOGLE'S RESPONSE: CLUSTER SIZE

- **Largest known database installations**
 - Greenplum – 96 nodes – 4.5 PB (eBay)[1]
 - Teradata – 72 nodes – 2+PB (eBay)[1]
- **Largest known MR installations:**
 - Hadoop – 3658 nodes – 1 PB (Yahoo)[2]
 - Hive – 600+ nodes – 2.5 PB (Facebook)[3]

[1] eBay's two enormous data warehouses – April 30th, 2009

<http://www.dbms2.com/2009/04/30/ebays-two-enormous-data-warehouses/>

[2] Hadoop sorts a petabyte in 16.25 hours and a terabyte in 62 seconds – May 11th, 2009

<https://developer.yahoo.com/blogs/hadoop/hadoop-sorts-petabyte-16-25-hours-terabyte-62-422.html>

[3] Hive – A Petabyte Scale Data Warehouse using Hadoop – June 10th, 2009

<https://www.facebook.com/notes/facebook-engineering/hive-a-petabyte-scale-data-warehouse-using-hadoop/89508453919>

CONCLUDING REMARKS

- **What can *MapReduce* learn from *Databases*?**
 - Declarative languages are a good thing
 - Schemas are important
- **What can *Databases* learn from *MapReduce*?**
 - Query fault-tolerance
 - Support for in situ data
 - Embrace open-source

SLIDES CAN BE FOUND AT:
TEACHINGDATASCIENCE.ORG