Lecture 3: MapReduce and Parallel Algorithms

Tonight

- Review
- "Scalability"
- MapReduce
- Assignment: MapReduce Basics

Review

- MADSkills paper takeaway
  - Even "non-relational" tasks can be expressed in the database
  - Why might this be a good idea?
- LINQ paper takeaway
  - Relational abstractions can be embedded in a programming language – no database required.
 - Why might this be a good idea?

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"**

  - Soon, you'll only get one pass at the data
  - So you better make that one pass count

Example: Find matching sequences

- Given a set of sequences
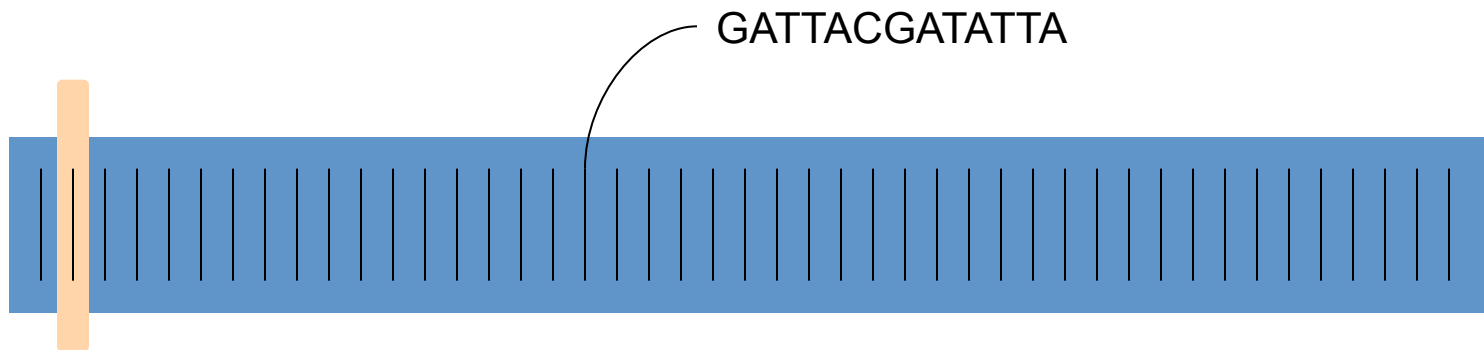- Find all sequences equal to "GATTACGATATTA"

TACCTGCCGTAA GATTACGATATTA
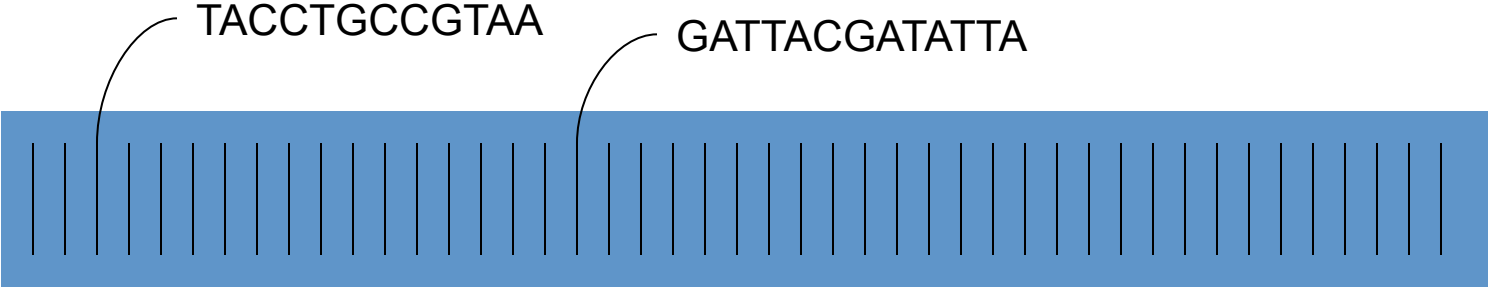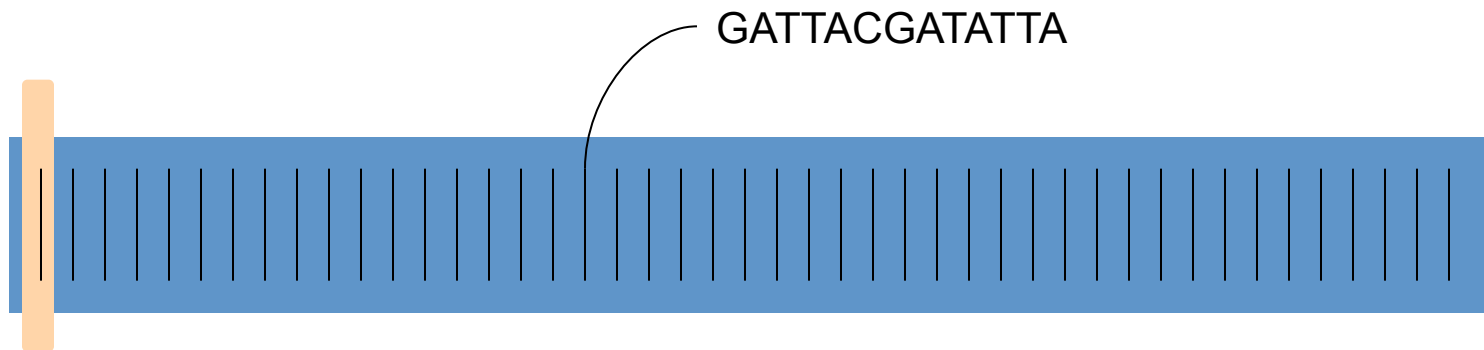
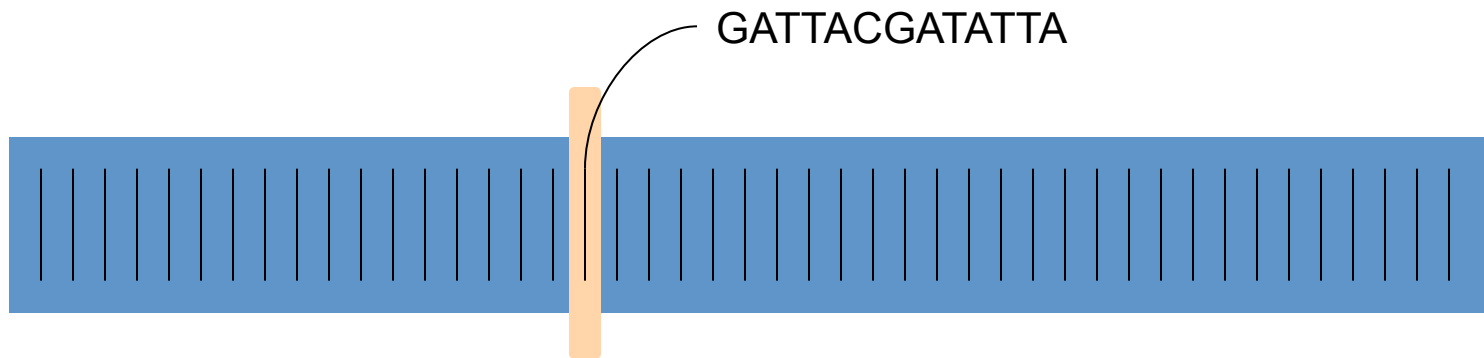GATTACGATATTA

TACCTGCCGTAA = GATTACGATATTA?

No.

time = 0

GATTACGATATTA

CCCCCAATGAC = GATTACGATATTA?

No.

time = 1
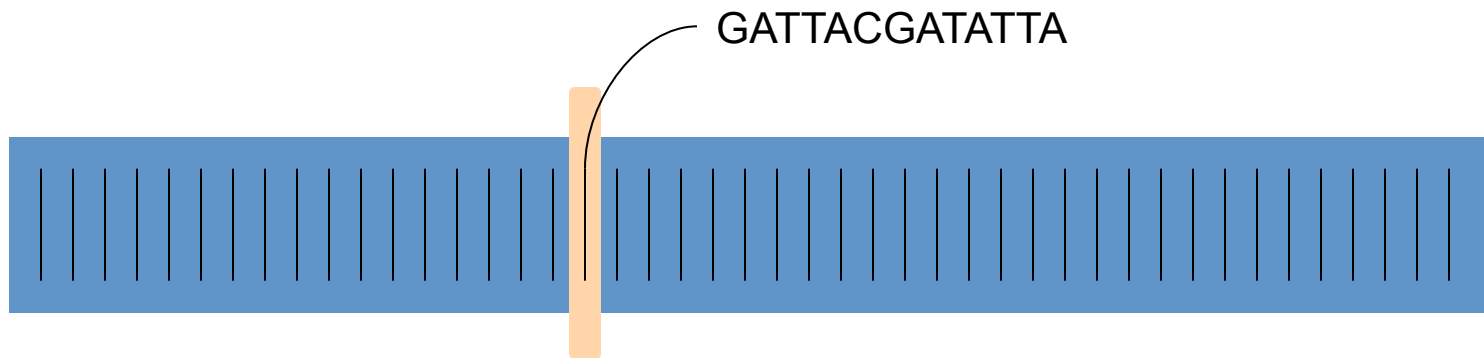
Bill Howe, eScience Institute

GATTACGATATTA

GATTACGATATTA contains GATTACGATATTA?

Yes!

Send it to the output.

time = 17

GATTACGATATTA

40 records, 40 comparions

N records, *N* comparisons

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

AAAATCCTGCA

AAACGCCTGCA

GATTACGATATTA

TTTTCGTAATT

TTTACGTCAA
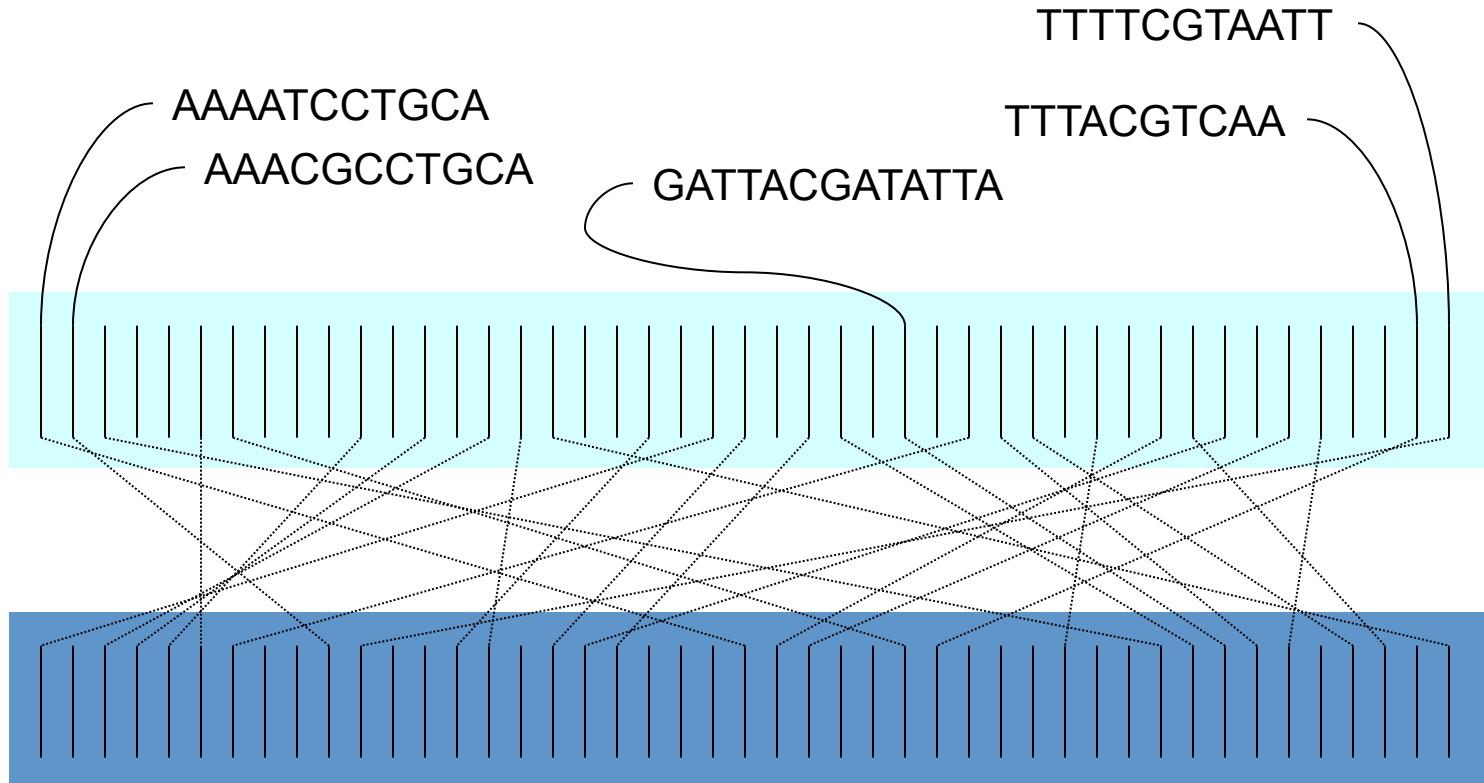
What if we sort the sequences?

CTGTACACAACCT

GATTACGATATTA

0%

100%
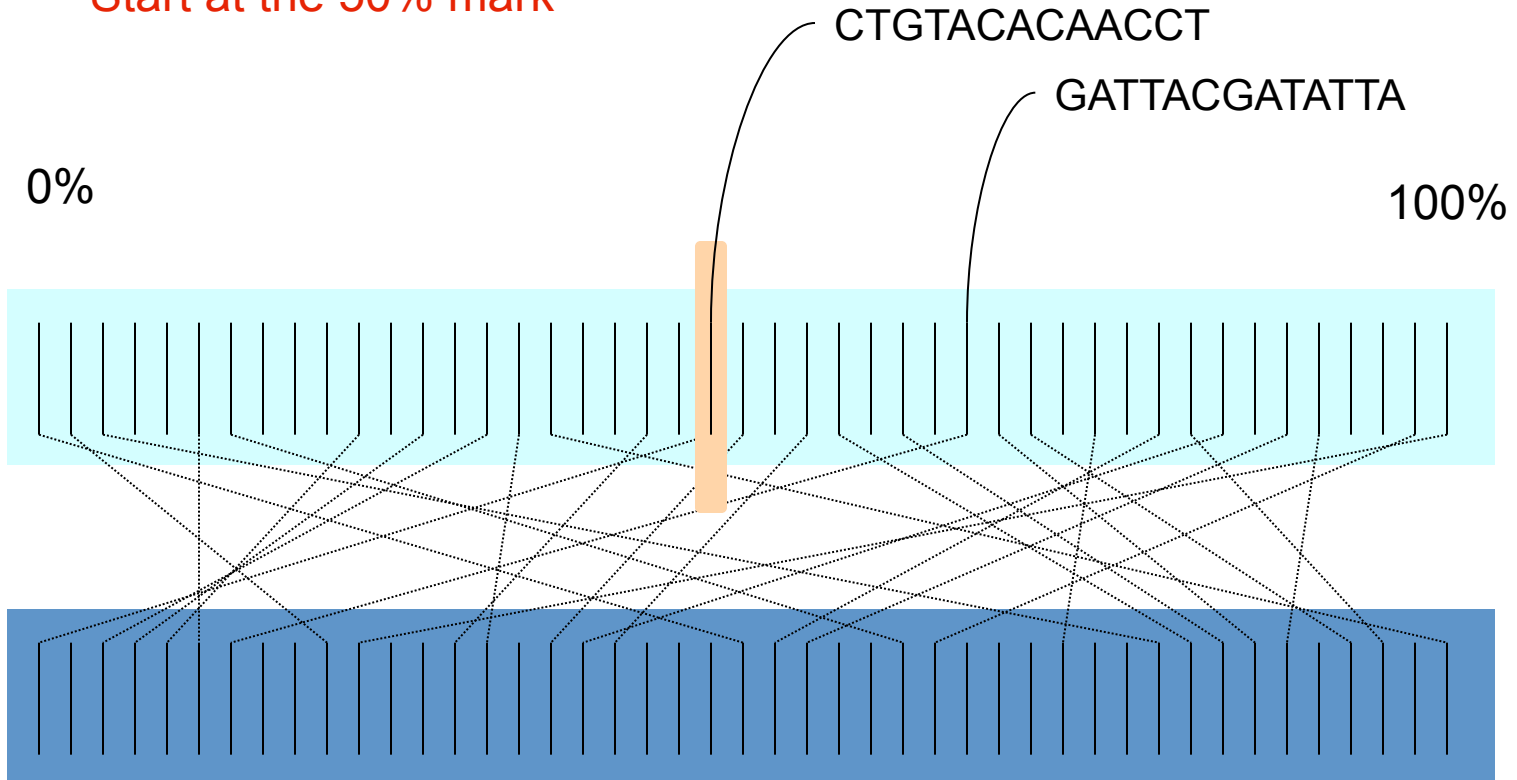
CTGTACACAACCT < GATTACGATATTA

time = 0

No match.

Skip to 75% mark

GATATTTTAAGC

GATTACGATATTA

0%

100%
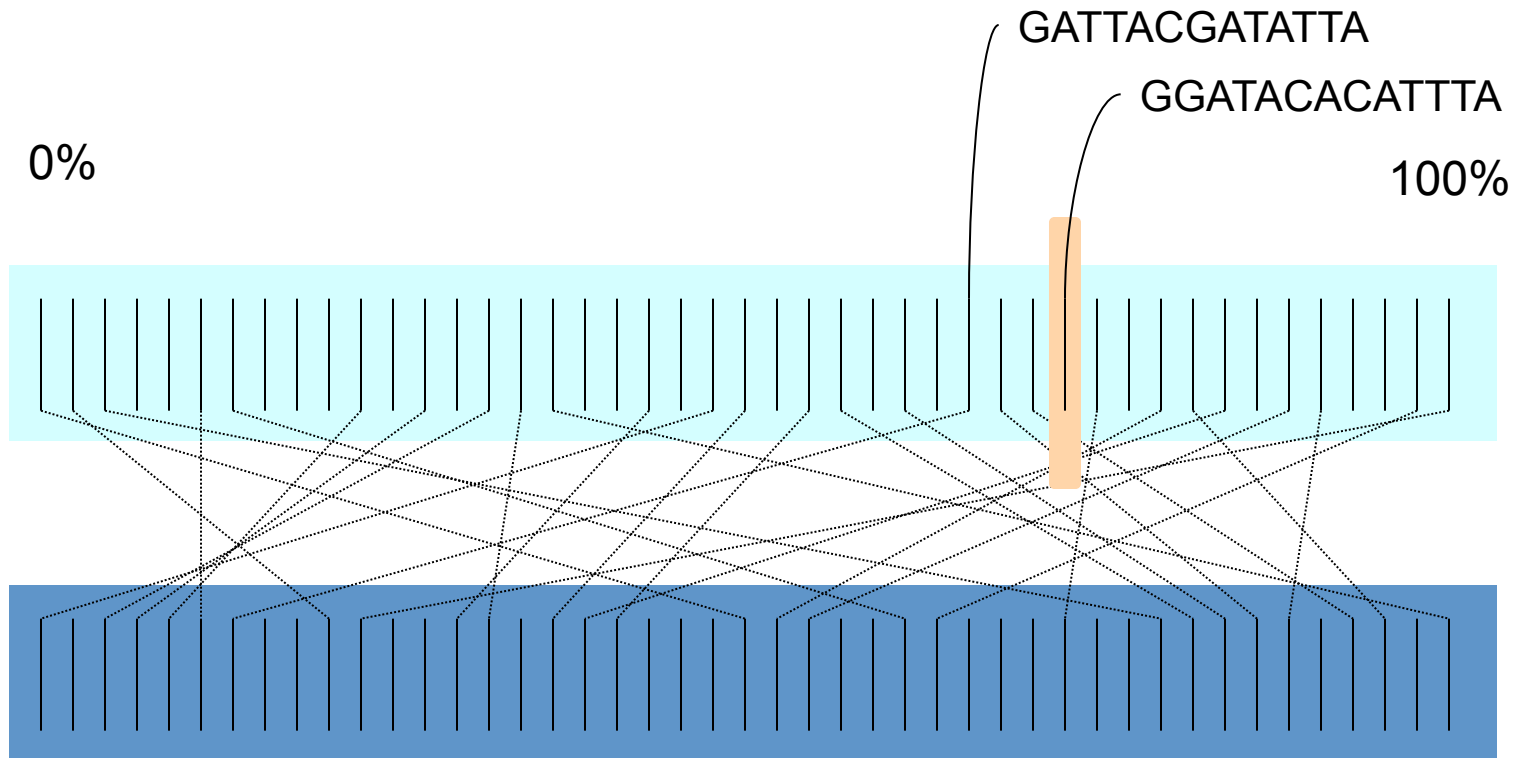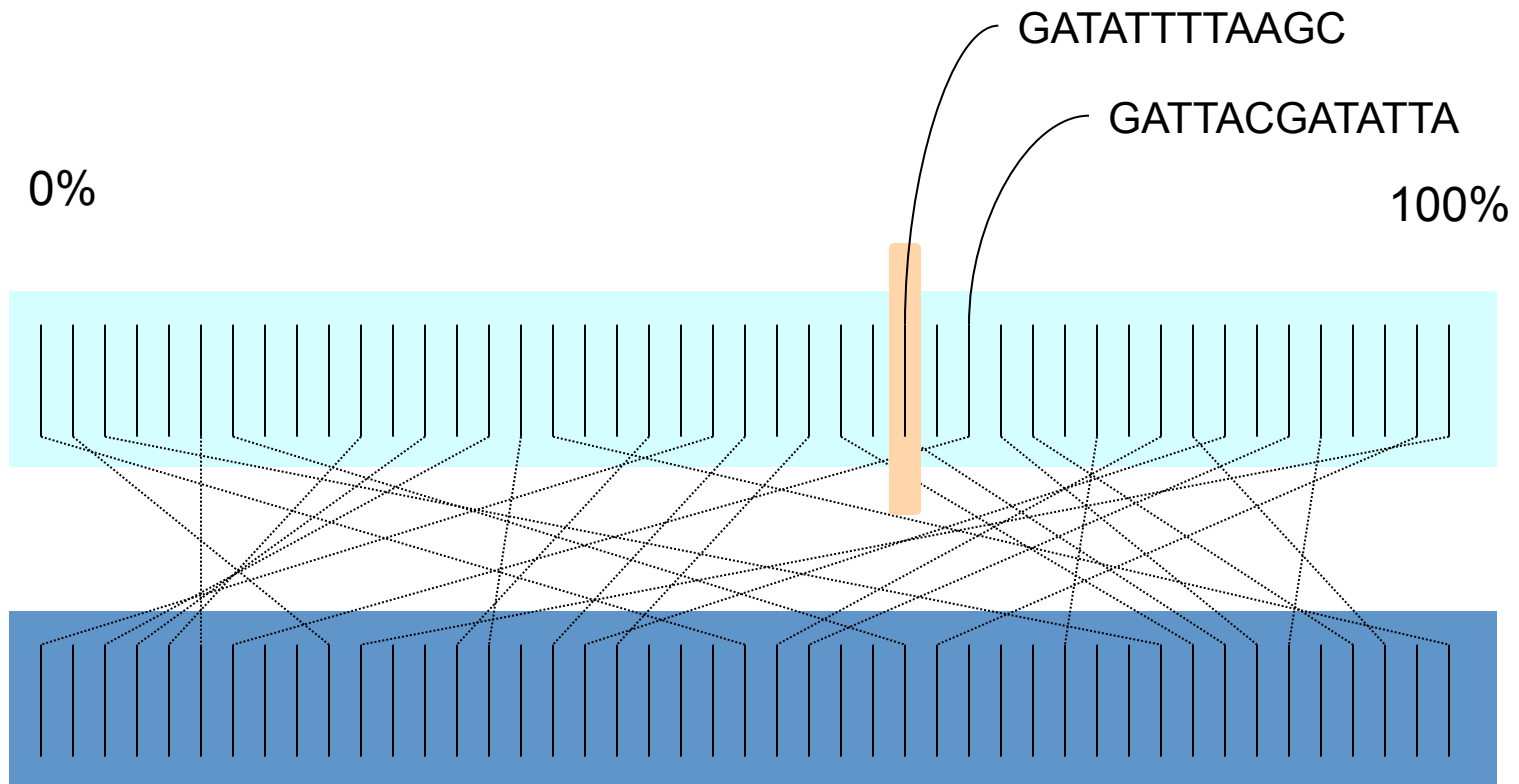
GATATTTTAAGC < GATTACGATATTA

No match.

Skip back to 68.75% mark

GATTACGATATTA

0%   100%

GATTACGATATTA = GATTACGATATTA

Match!

Walk through the records until we fail to match.

GATTACGATATTA

0%   100%
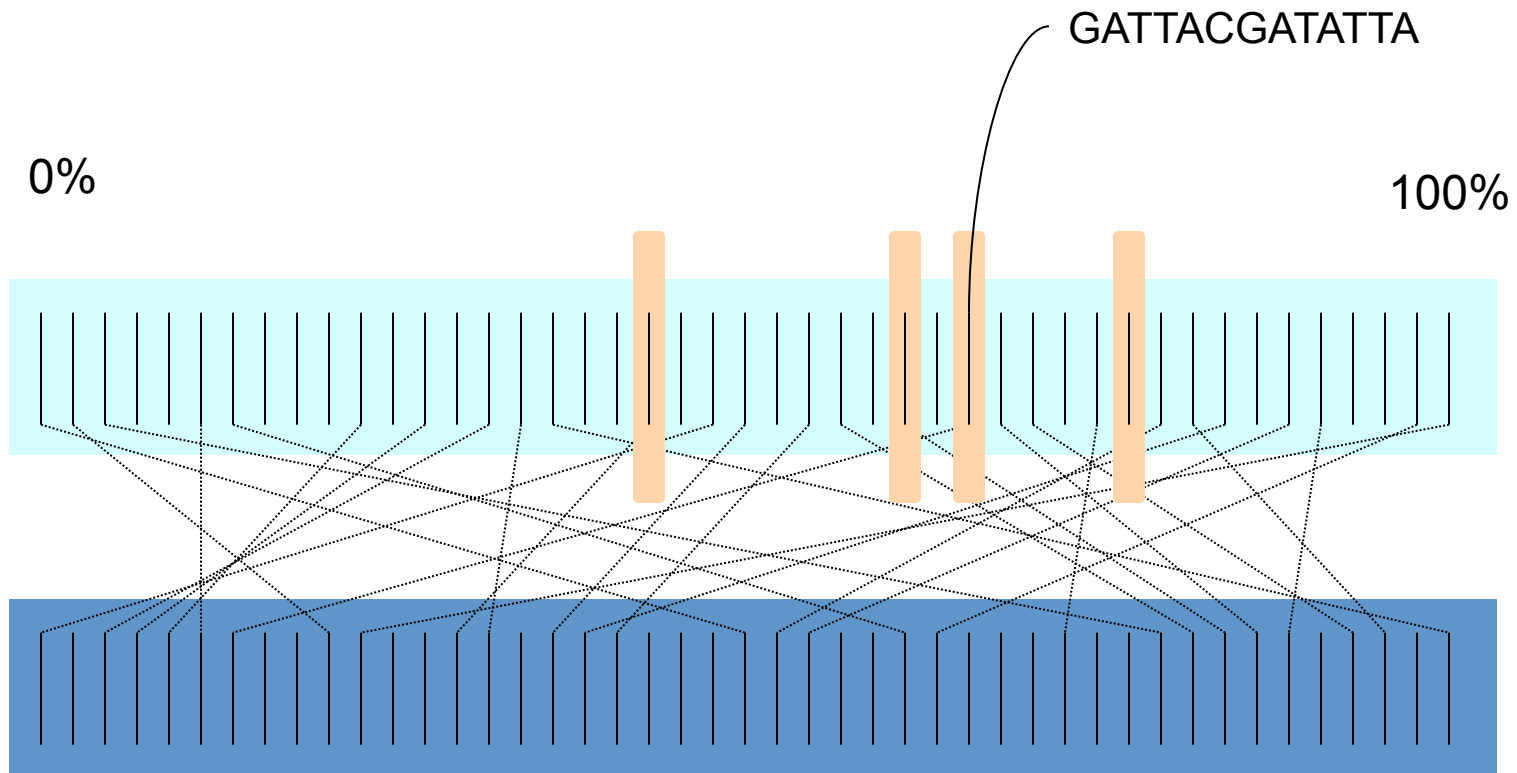
How many comparisons did we do?

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**
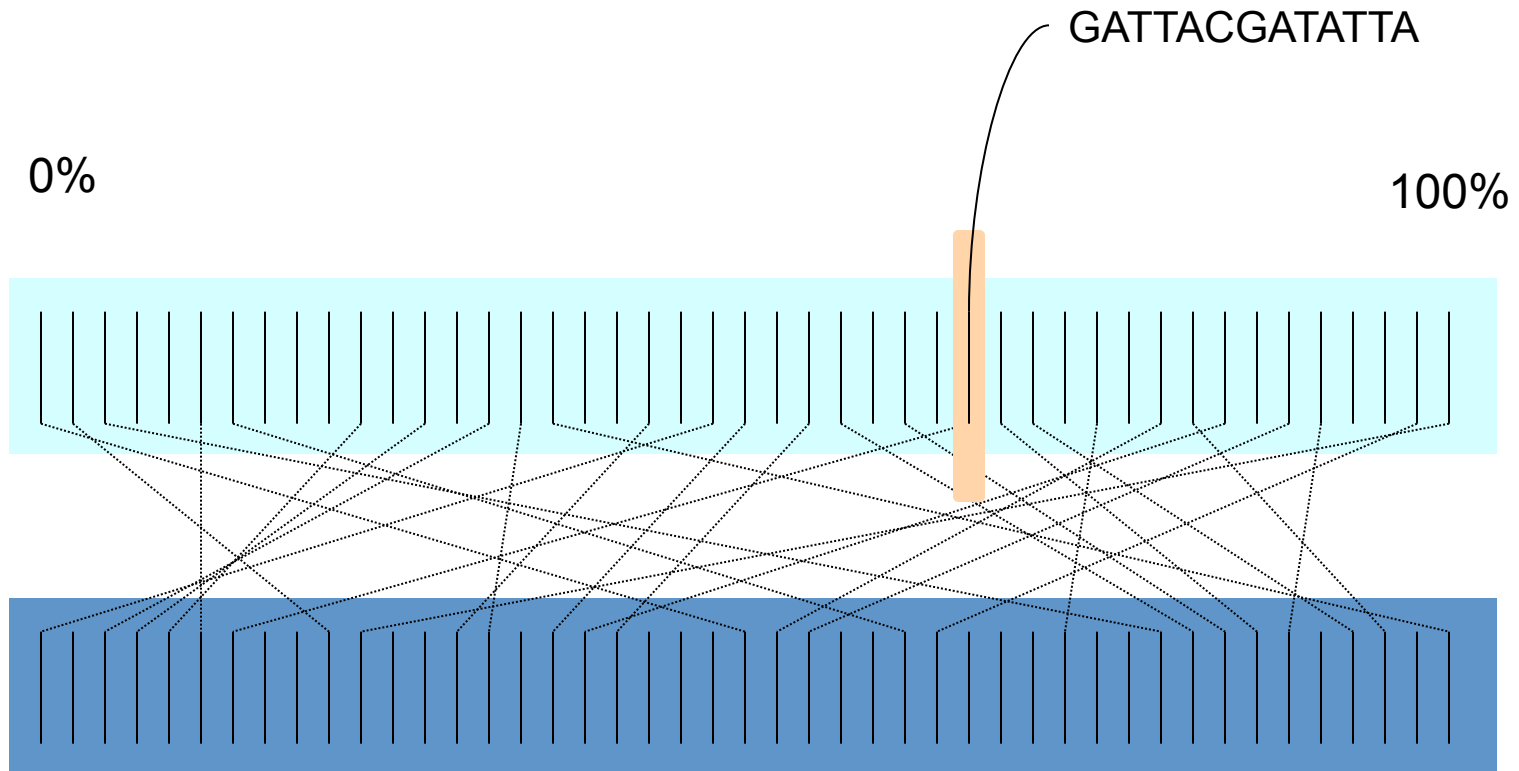 - **Your query will always* finish, regardless of dataset size.**

  - **Indexes are <u>easily built</u> and <u>automatically used</u> 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
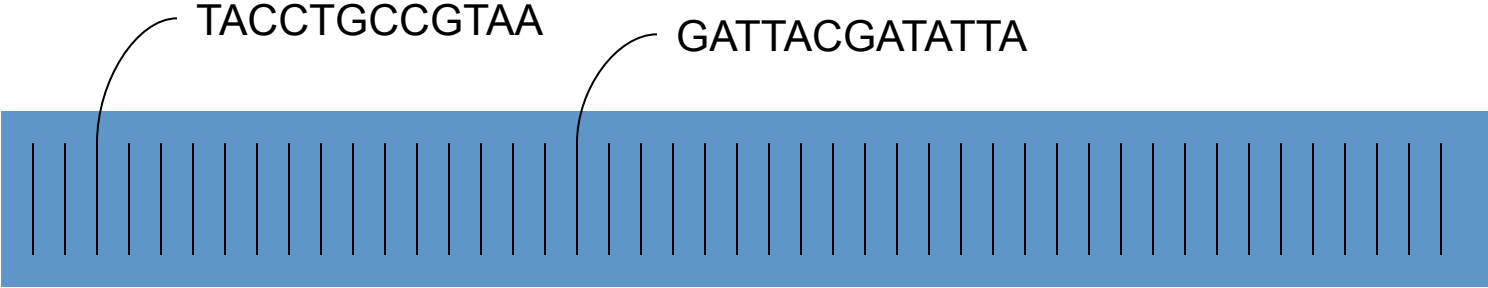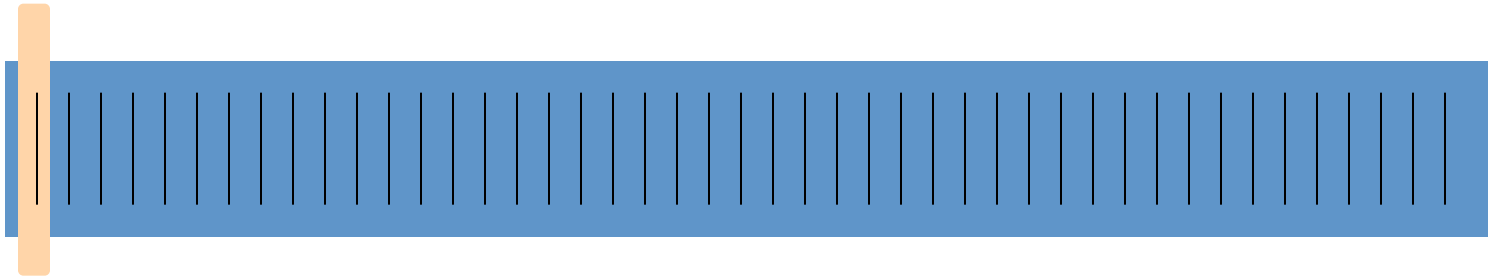
TACCTGCCGTAA becomes TACCT

time = 0
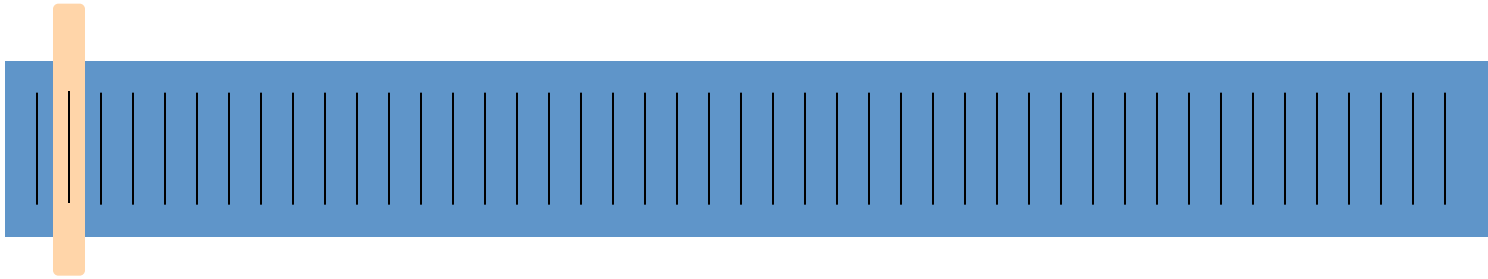
CCCCCAATGAC becomes CCCCC

time = 1
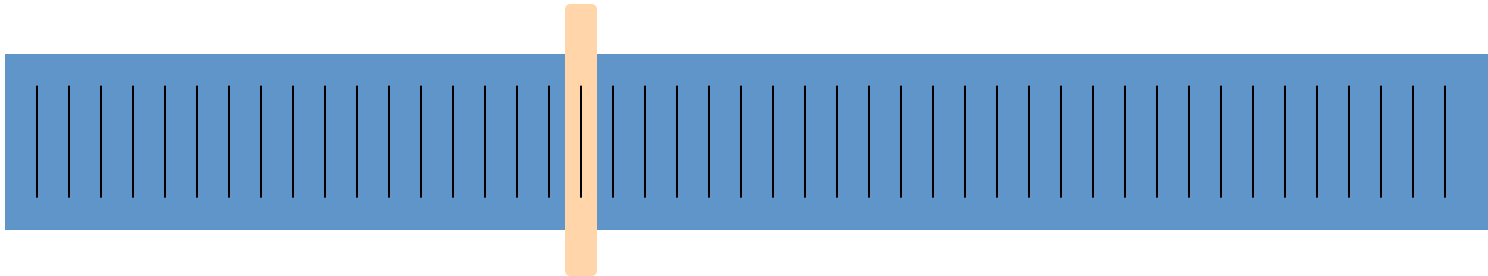
GATTACGATATTA becomes GATTA

time = 17
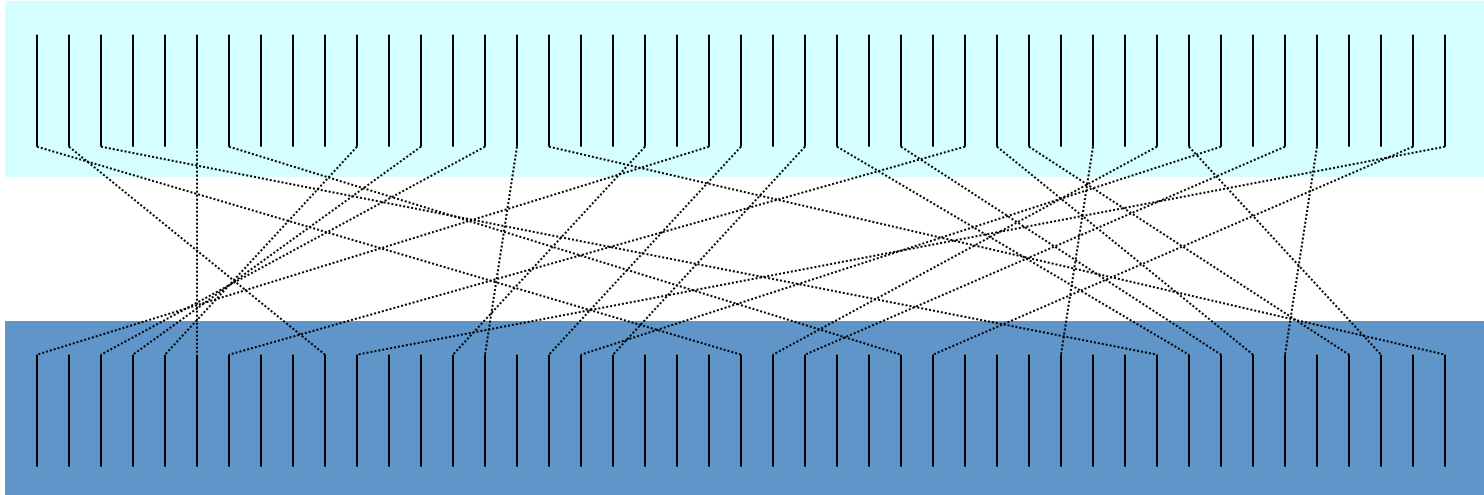
Can we use an index?

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

The task is fundamentally O(N)

Can we do any better?
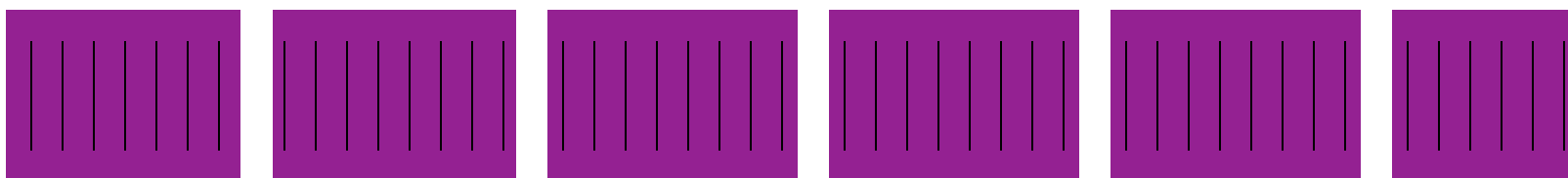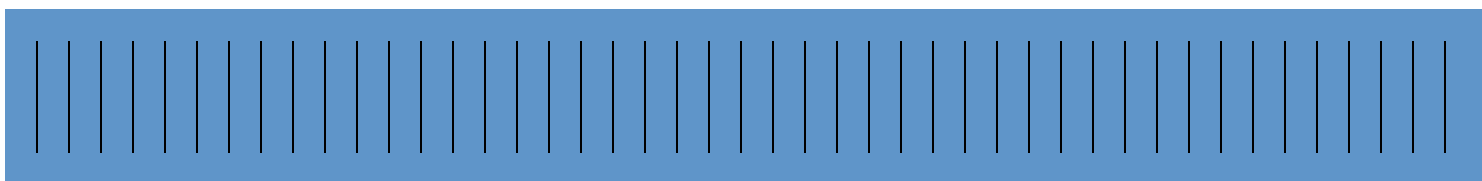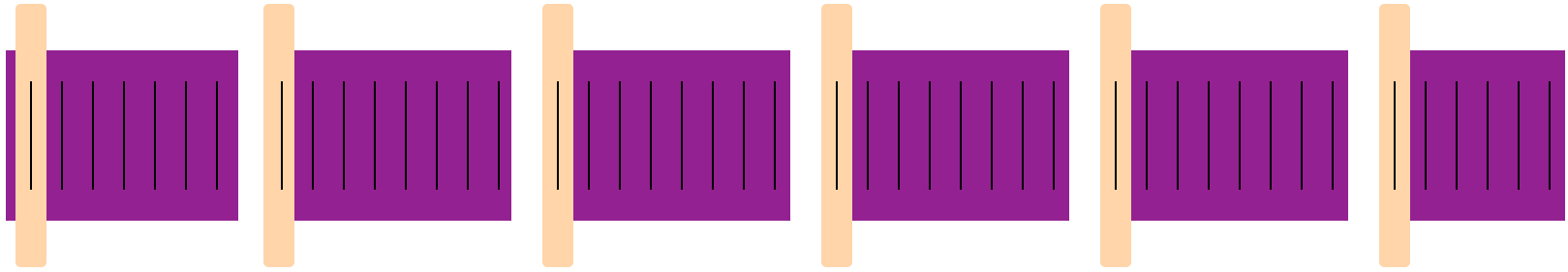
time = 0
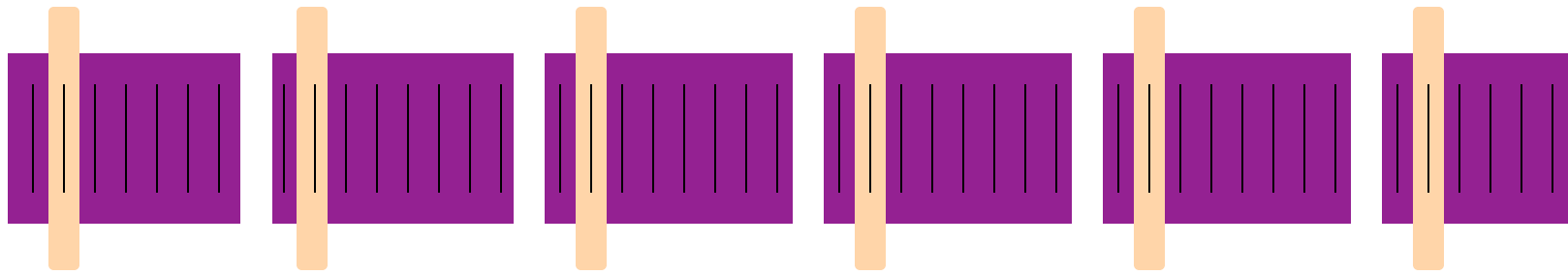
time = 1
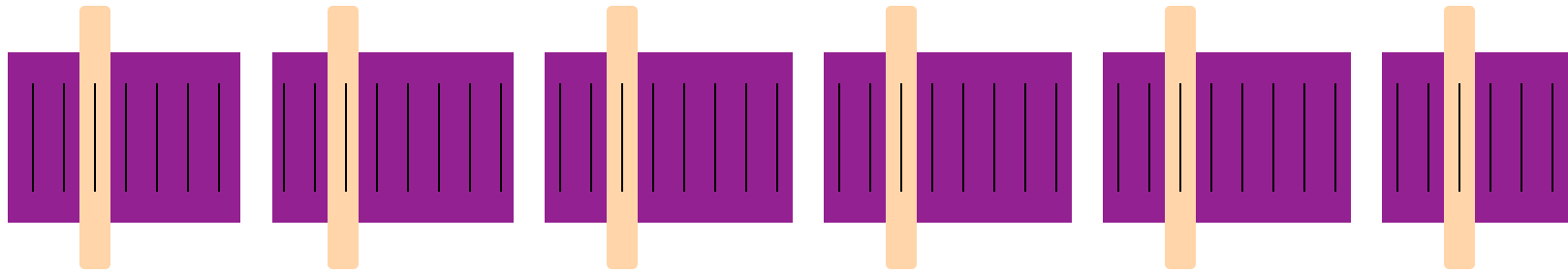
time = 2
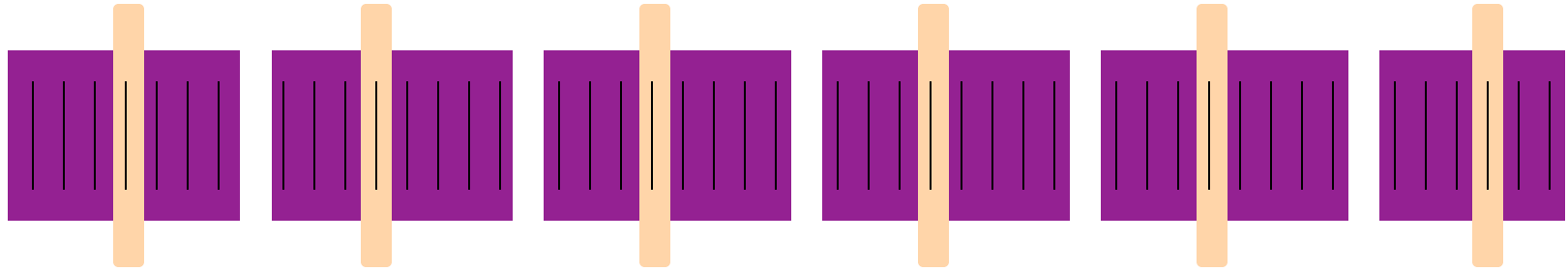
time = 3
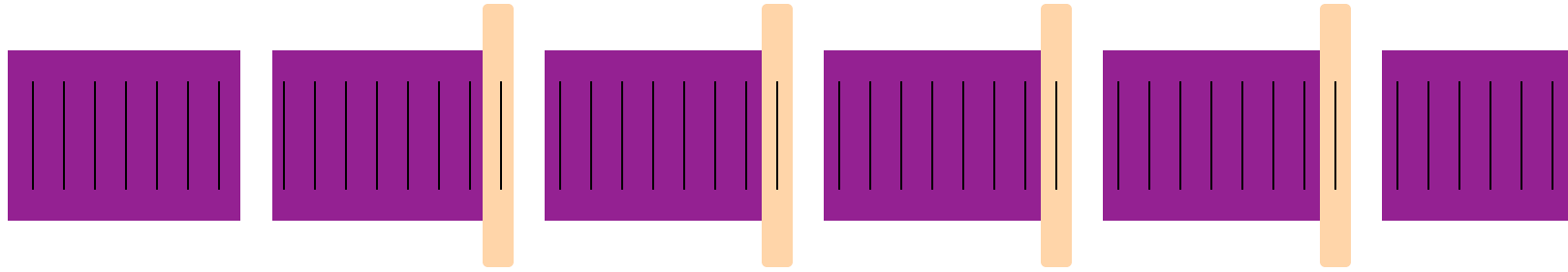
How much time did this take?

7 cycles

40 records, 6 workers

O(N/k)

time = 7

Taxonomy of Parallel Architectures

Scales to 1000s of computers
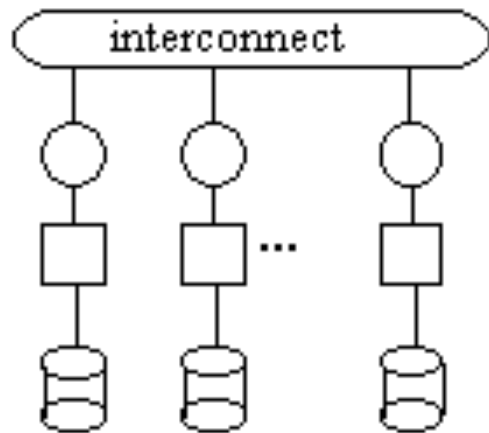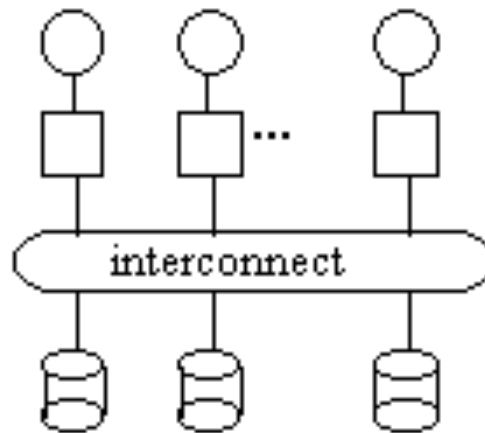
a) shared nothing   b) shared disc   c) shared memory
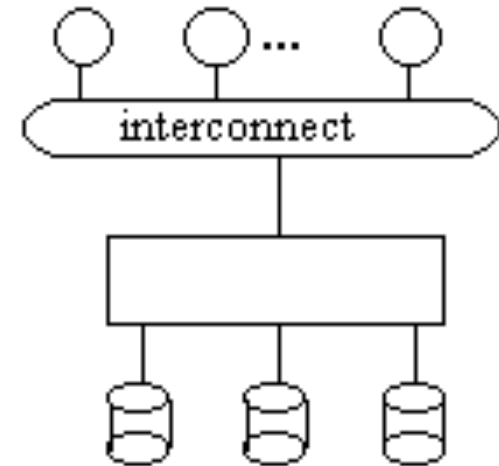
Easiest to program, but **$$$$**

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

Design Space

Internet

Grid

In a few weeks

The area we're discussing

Data-parallel

Search

DISC

Shared memory

Private data center

Transaction

HPC

Latency

Throughput
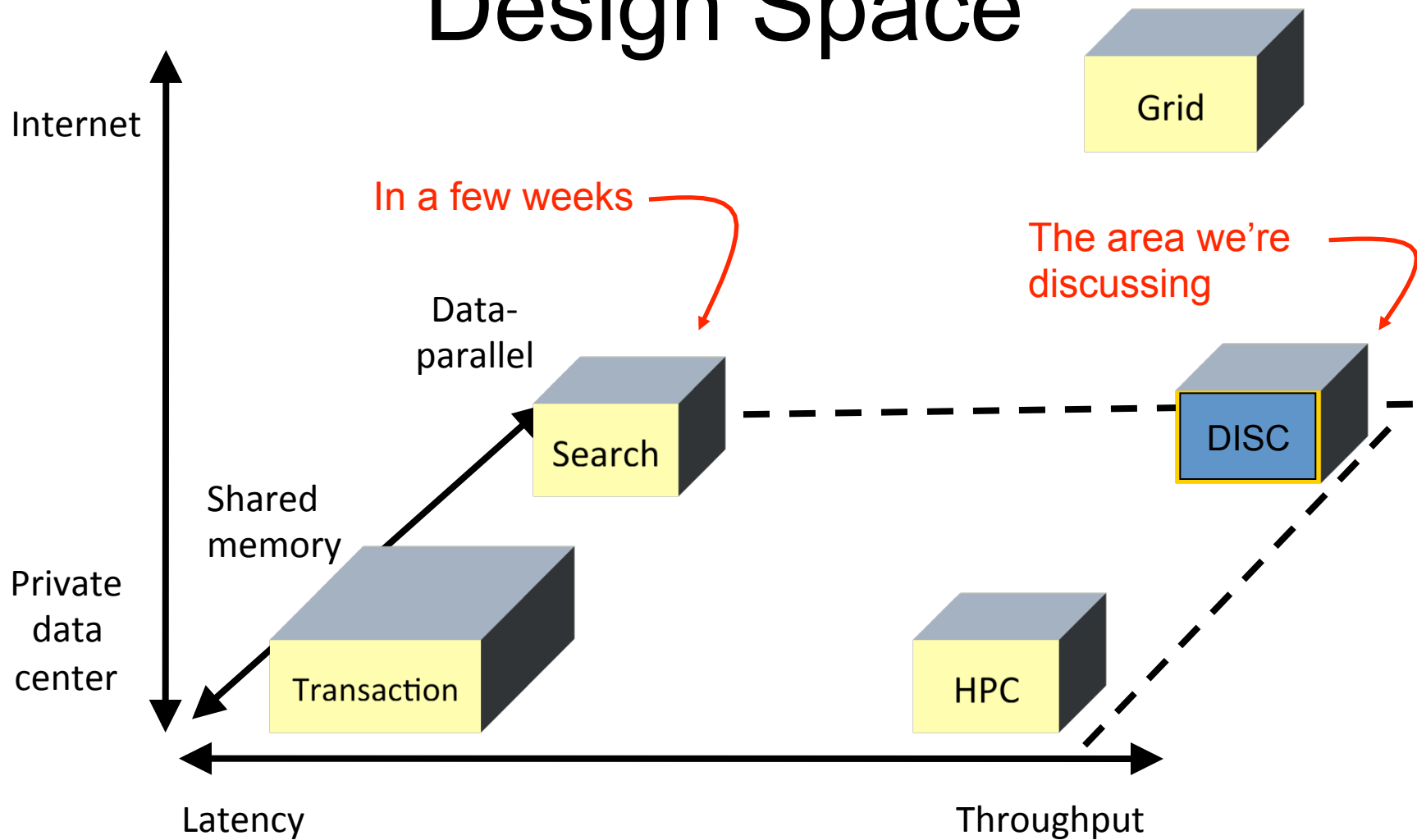
Some distributed algorithm…

Map

(Shuffle)
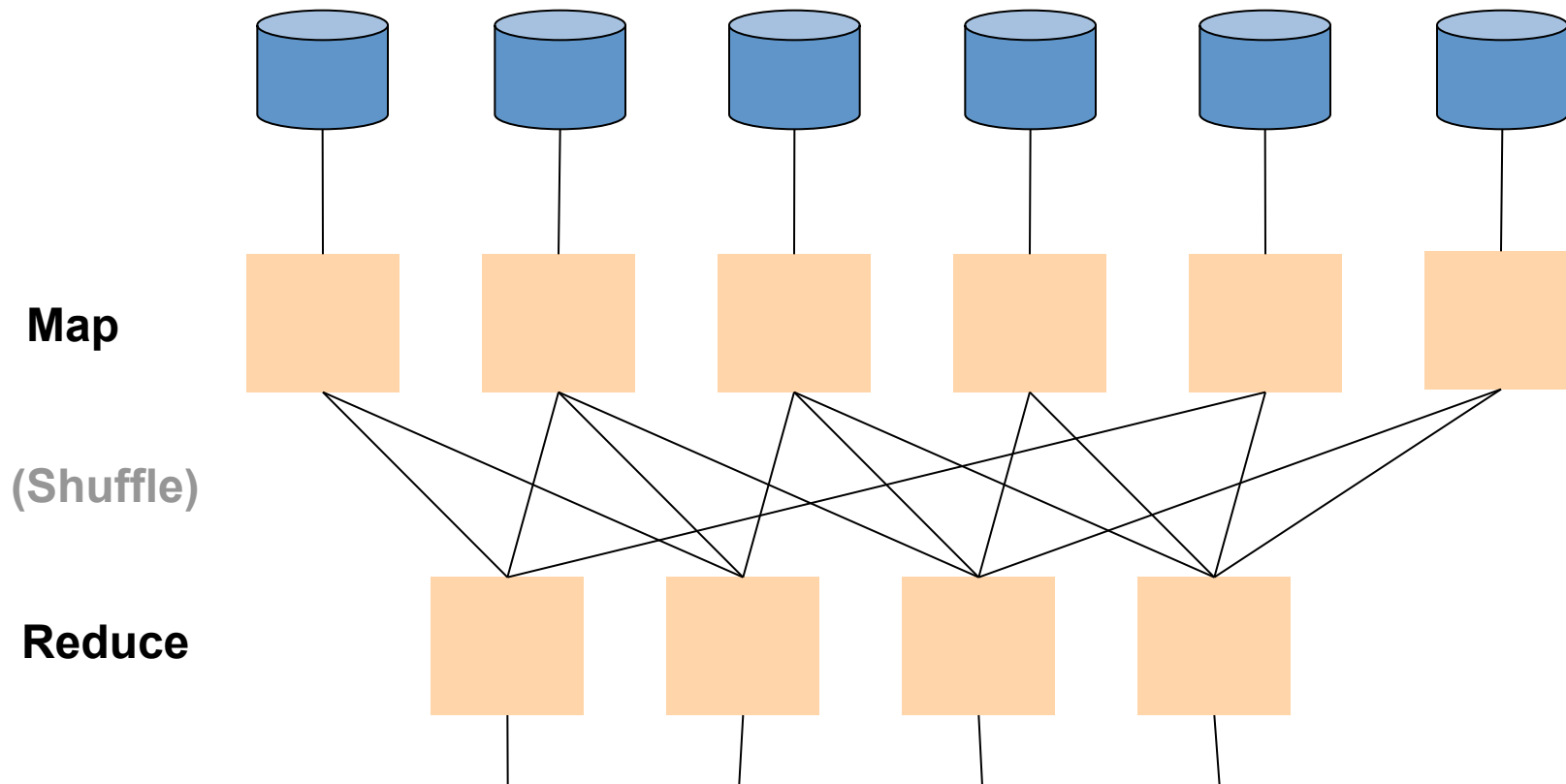
Reduce

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

slide source: Google, Inc.

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);
```

slide source: Google, Inc.

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 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

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 1

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.

Chunk 2

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 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

(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, 37)
(yellow, 20)

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

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

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

MR Phases

- Each Map and Reduce task has multiple phases:

Map Task

Reduce Task

{P 1} {P 2} {P 3} {P 4} {P 5}

| Split | Record Reader → Map → Combine | Copy | Sort | Reduce |

file

HDFS

Local storage

file

HDFS
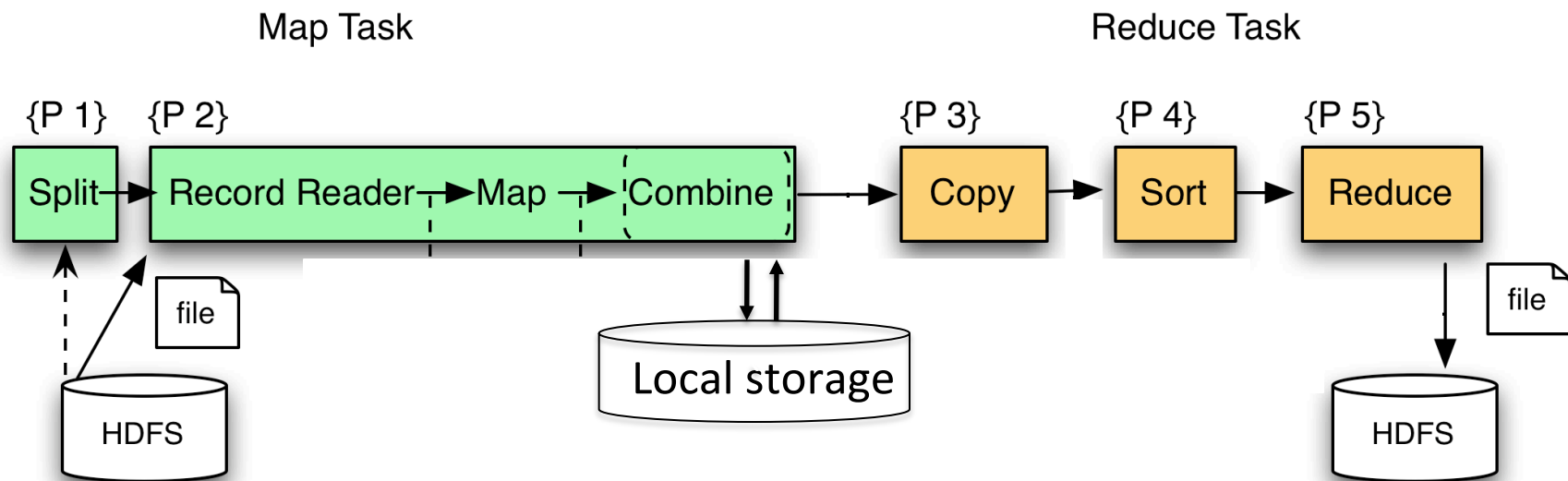
Hadoop in One Slide

*src: Huy Vo,
   NYU Poly*

MAP Tasks

REDUCE Tasks

(did1,v1)

(did2,v2)

(did3,v3)

. . . .

(w1,1)
(w2,1)
(w3,1)
...
(w1,1)
(w2,1)
...

(w1, (1,1,1,...,1))
(w2, (1,1,...))
(w3,(1...))
...
...
...

(w1, 25)
(w2, 77)
(w3, 12)
...
...
...
...
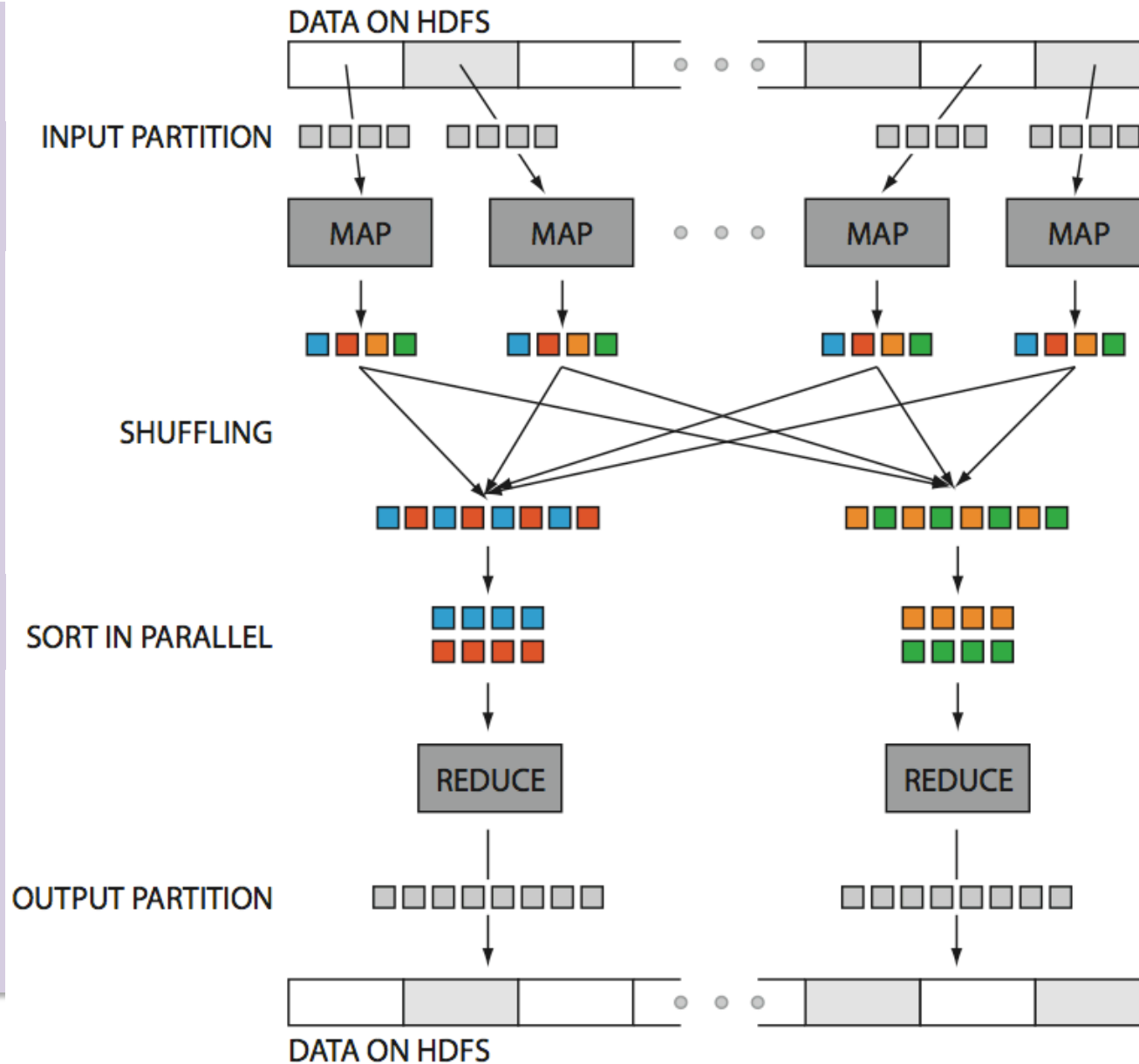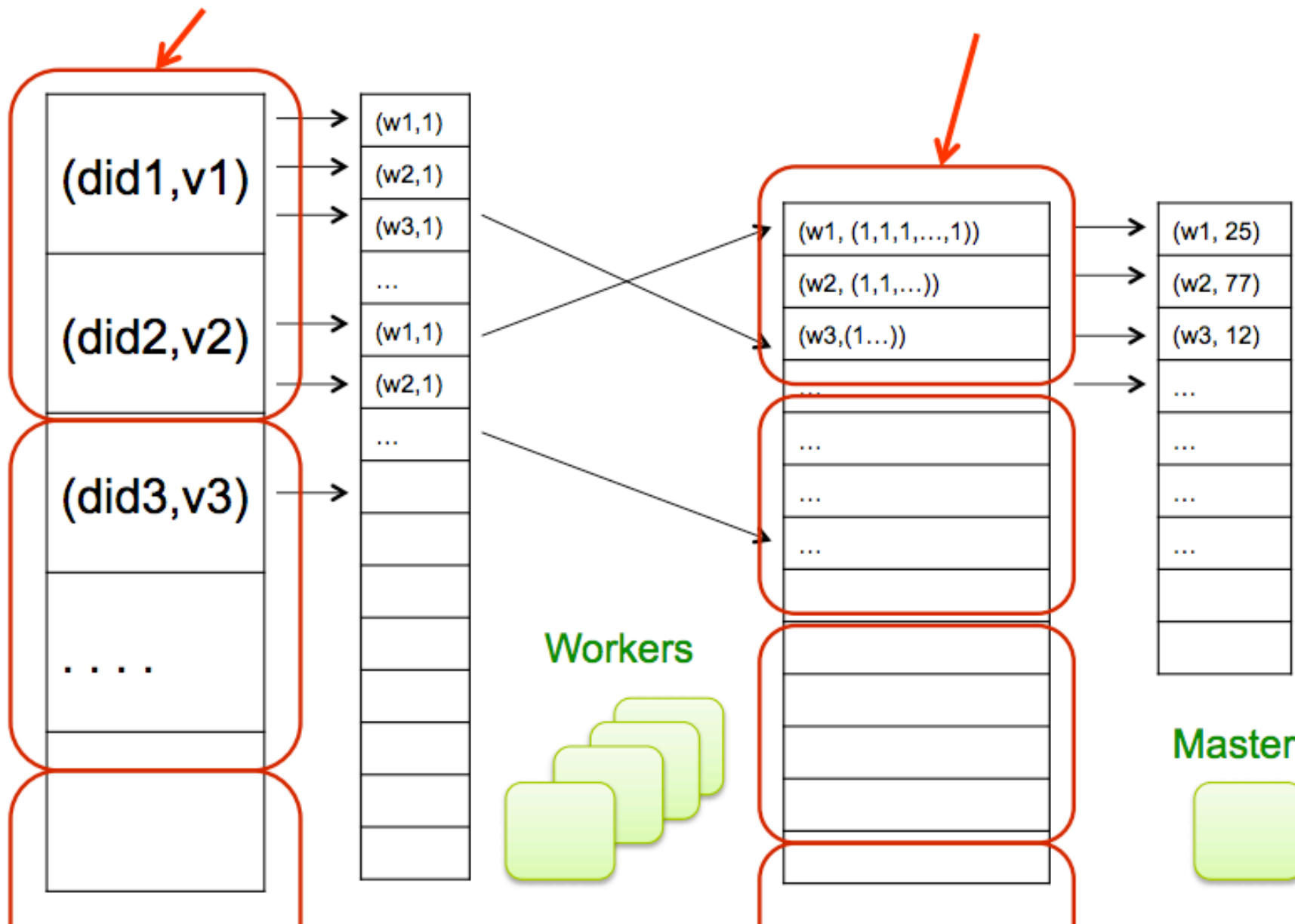
Workers

Master

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()
```
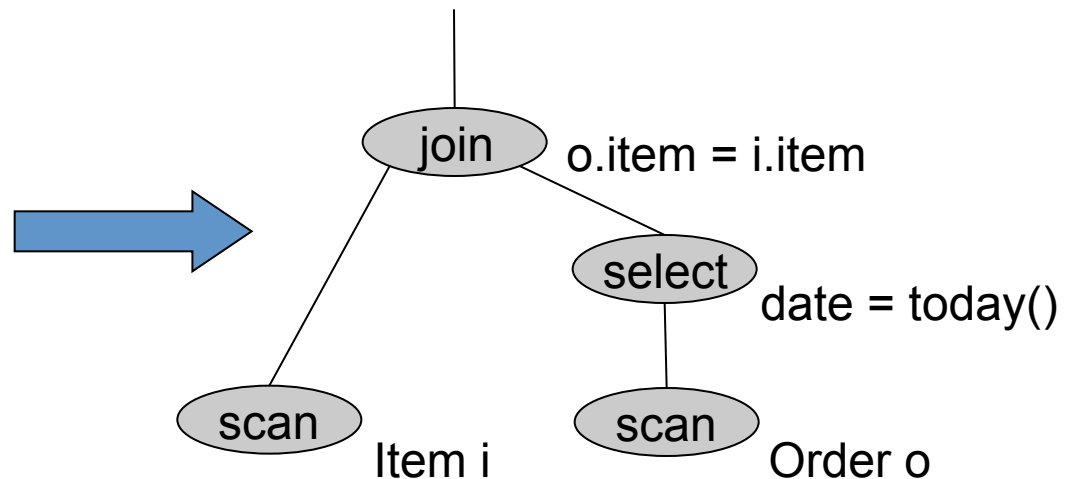
Two notions of parallel query processing

- "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
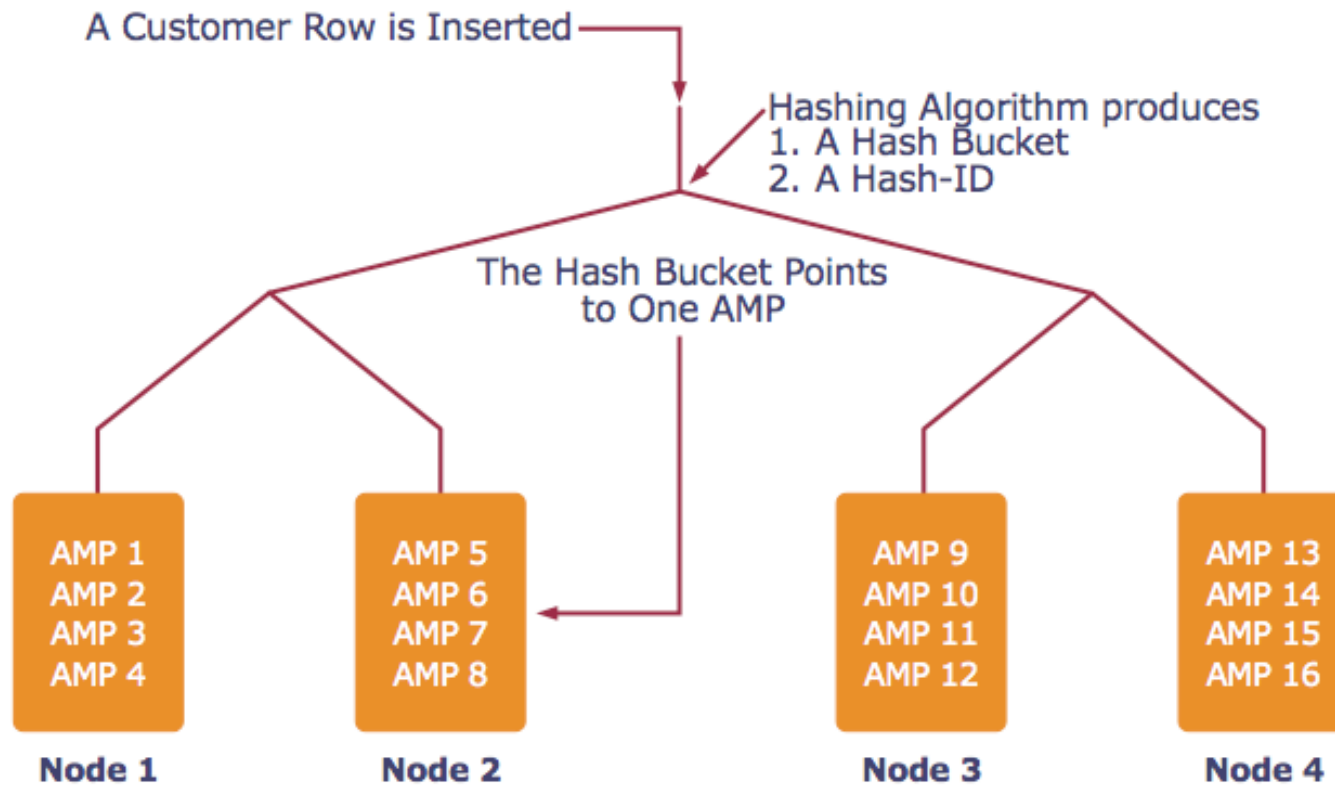
Distributed Query Example

```
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)
)
```

Parallel Query Example: Teradata

A Customer Row is Inserted

Hashing Algorithm produces
1. A Hash Bucket
2. A Hash-ID

The Hash Bucket Points
to One AMP

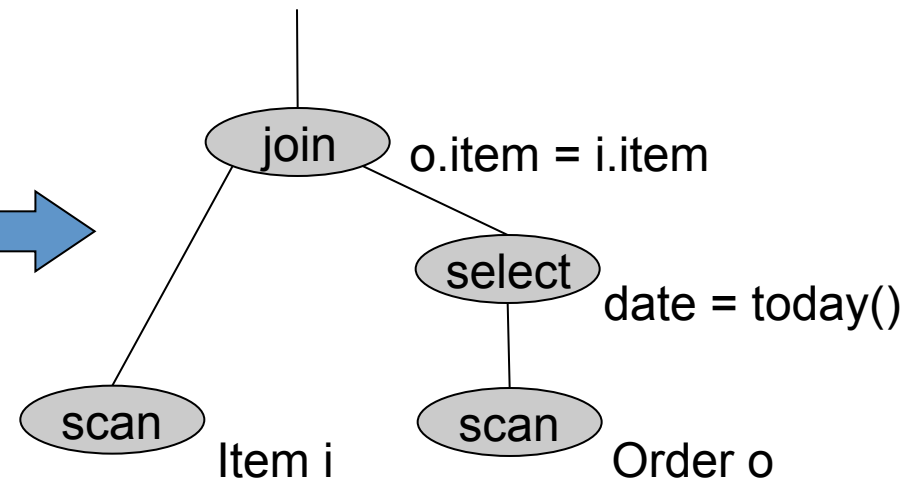| AMP 1 | AMP 5 | AMP 9 | AMP 13 |
AMP 2	AMP 6	AMP 10	AMP 14
AMP 3	AMP 7	AMP 11	AMP 15
AMP 4	AMP 8	AMP 12	AMP 16
Node 1	Node 2	Node 3	Node 4

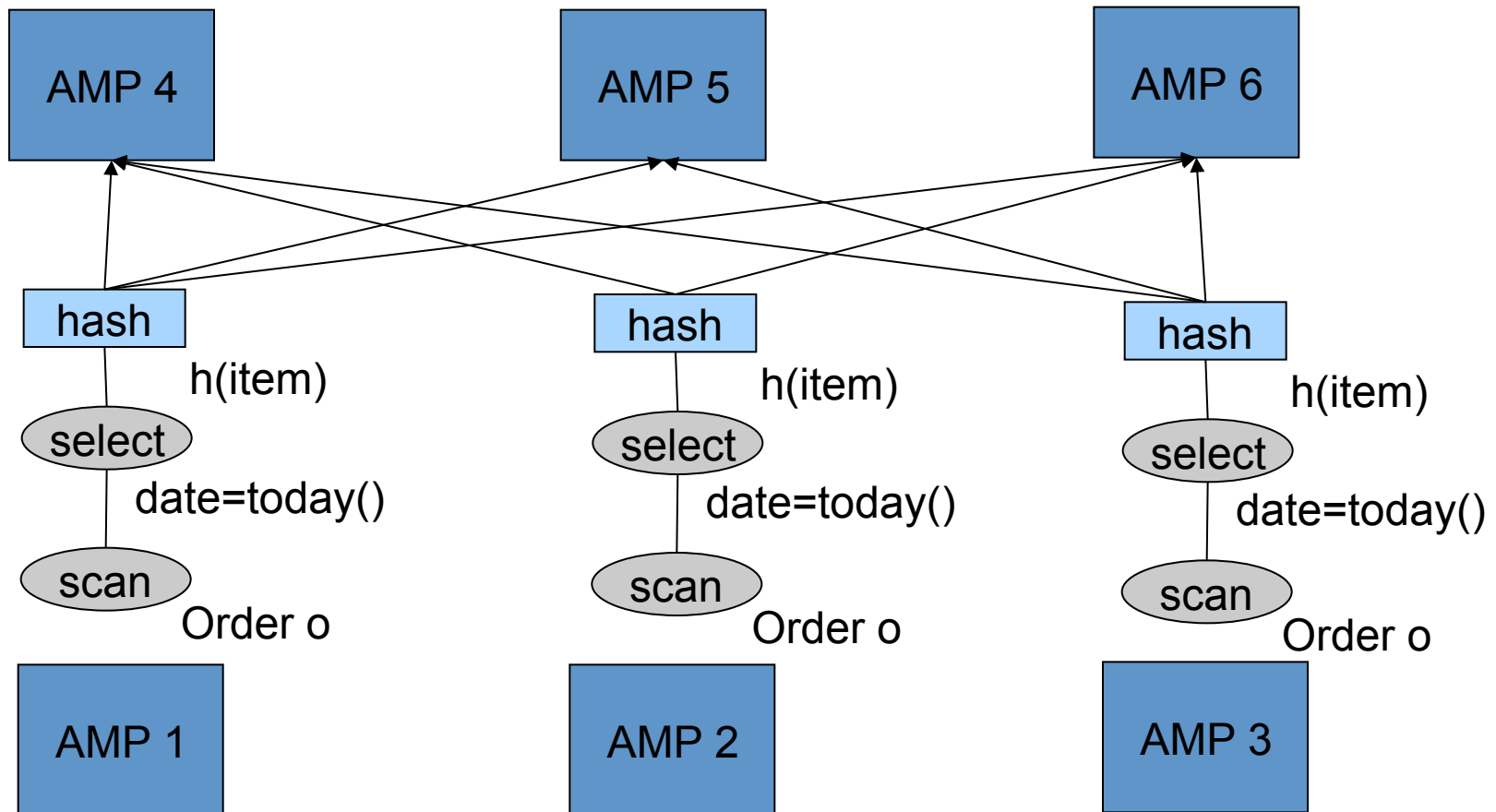*AMP = unit of parallelism*

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()
```

join   o.item = i.item

select   date = today()

scan   Item i

scan   Order o

Example System: Teradata

AMP 4   AMP 5   AMP 6

hash   h(item)   hash   h(item)   hash   h(item)

select   date=today()   select   date=today()   select   date=today()

scan   Order o   scan   Order o   scan   Order o

AMP 1   AMP 2   AMP 3

Example System: Teradata

Example System: Teradata

join

o.item = i.item

join

o.item = i.item

join

o.item = i.item

AMP 4

AMP 5

AMP 6

contains all orders and all lines where hash(item) = 3

contains all orders and all lines where hash(item) = 2

contains all orders and all lines where hash(item) = 1
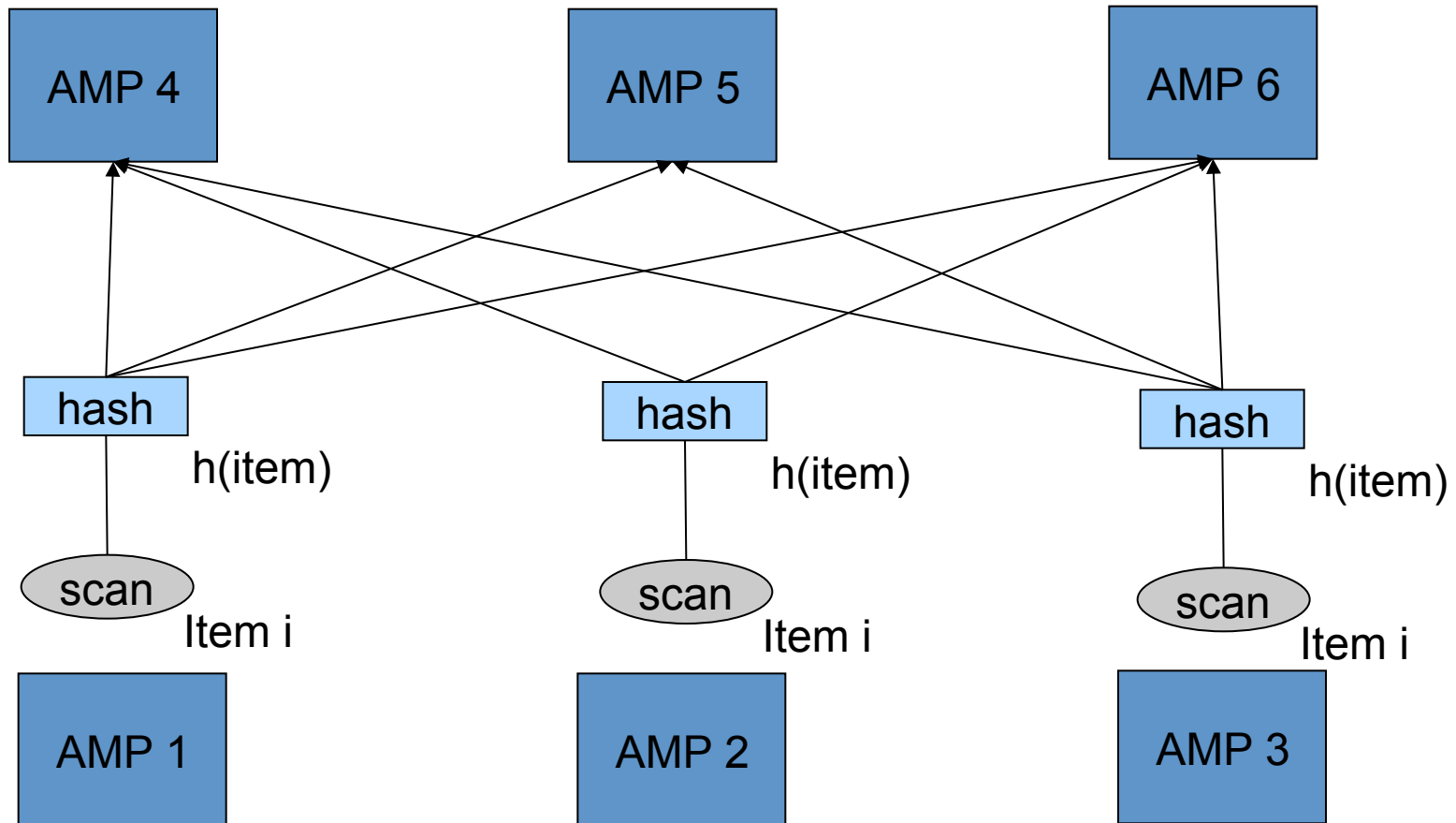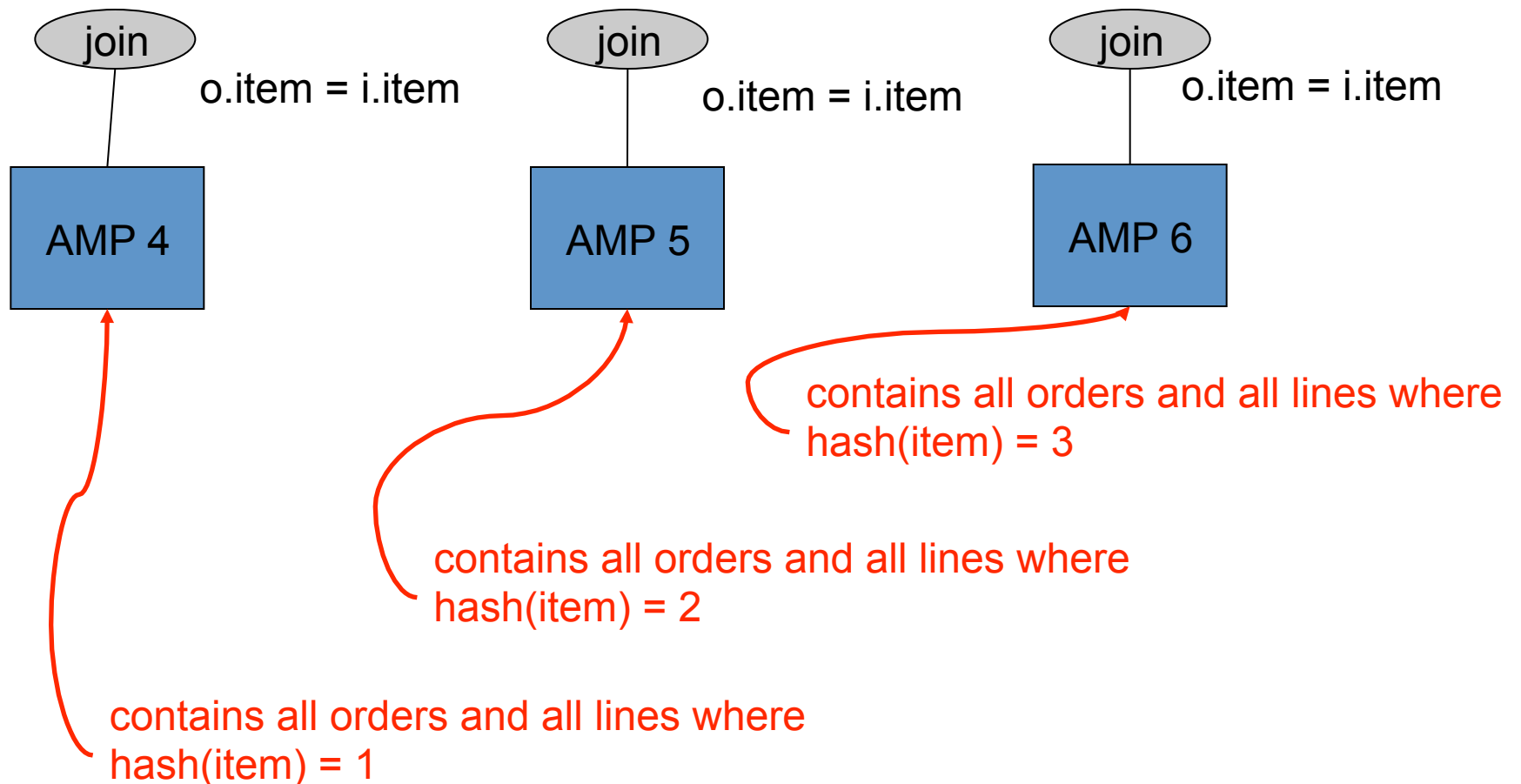
MapReduce Contemporaries

- 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   DryadLINQ, Pig, HIVE
  - Schemas   HIVE, Pig, DryadLINQ
  - Logical Data Independence
  - Indexing   Hbase
  - Algebraic Optimization   Pig, (Dryad, HIVE)
  - Caching/Materialized Views
  - *ACID/Transactions*

- MapReduce
  - High Scalability
  - Fault-tolerance
  - "One-person deployment"

	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

Today's Reading, Pavlo 2009

MR VS. DATABASES

Pavlo et al. 2009

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.
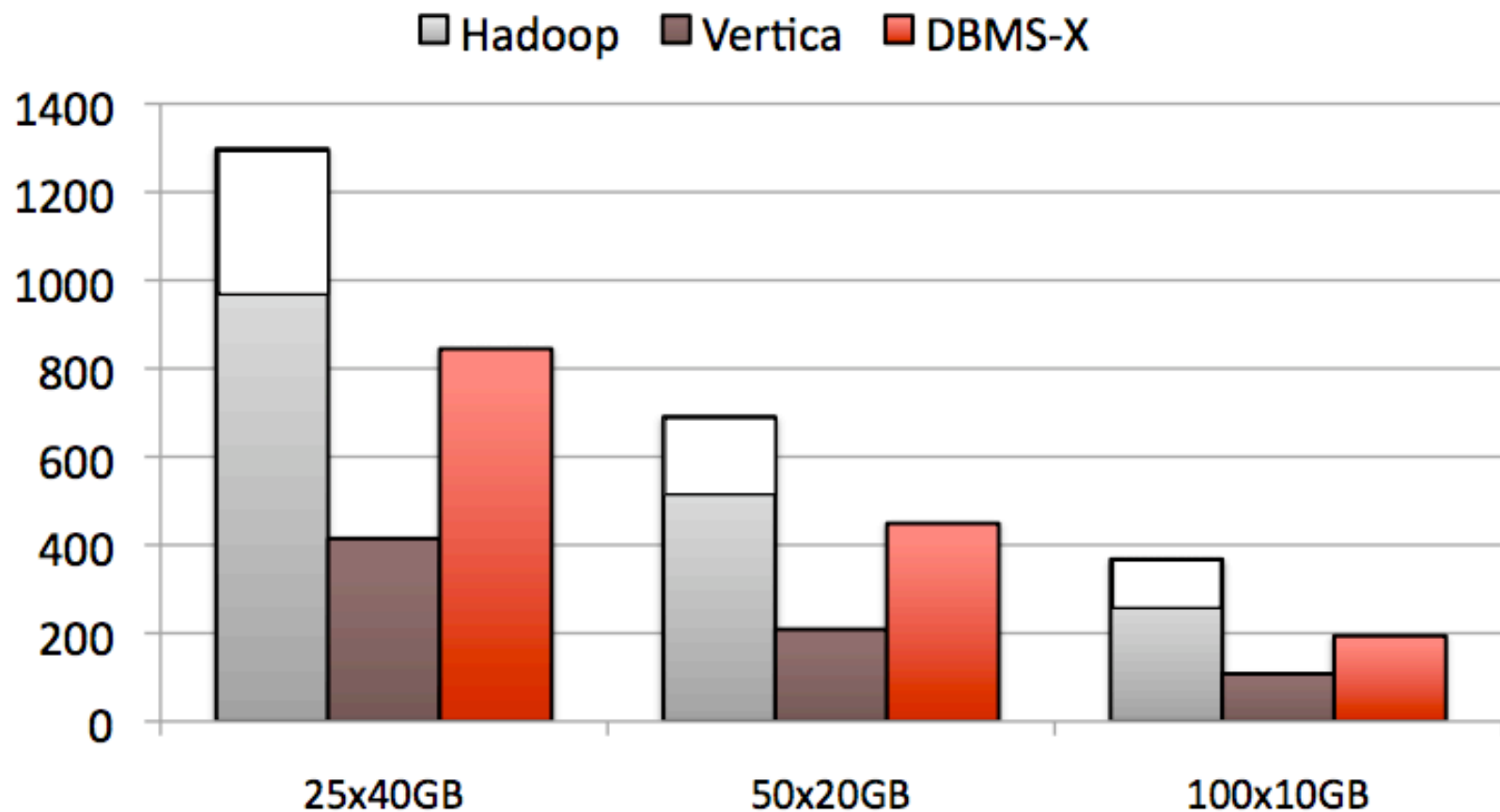- Quantitative
  - Data loading, different types of queries

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)**
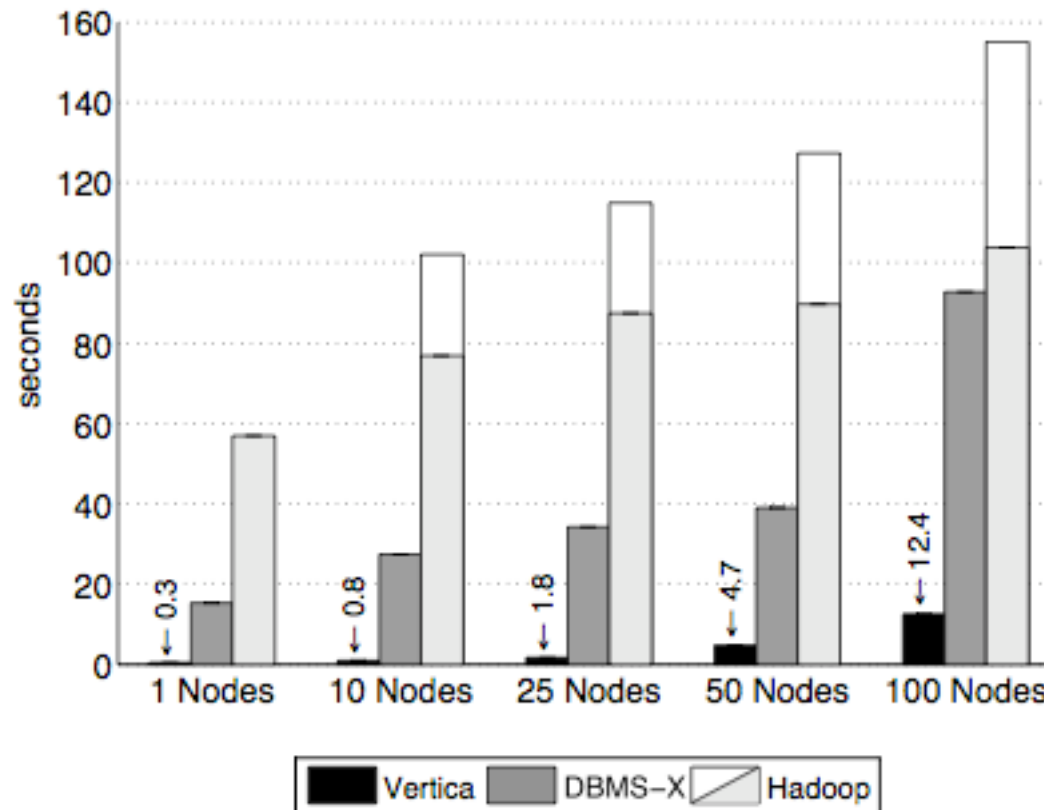
Grep Task Loading Results
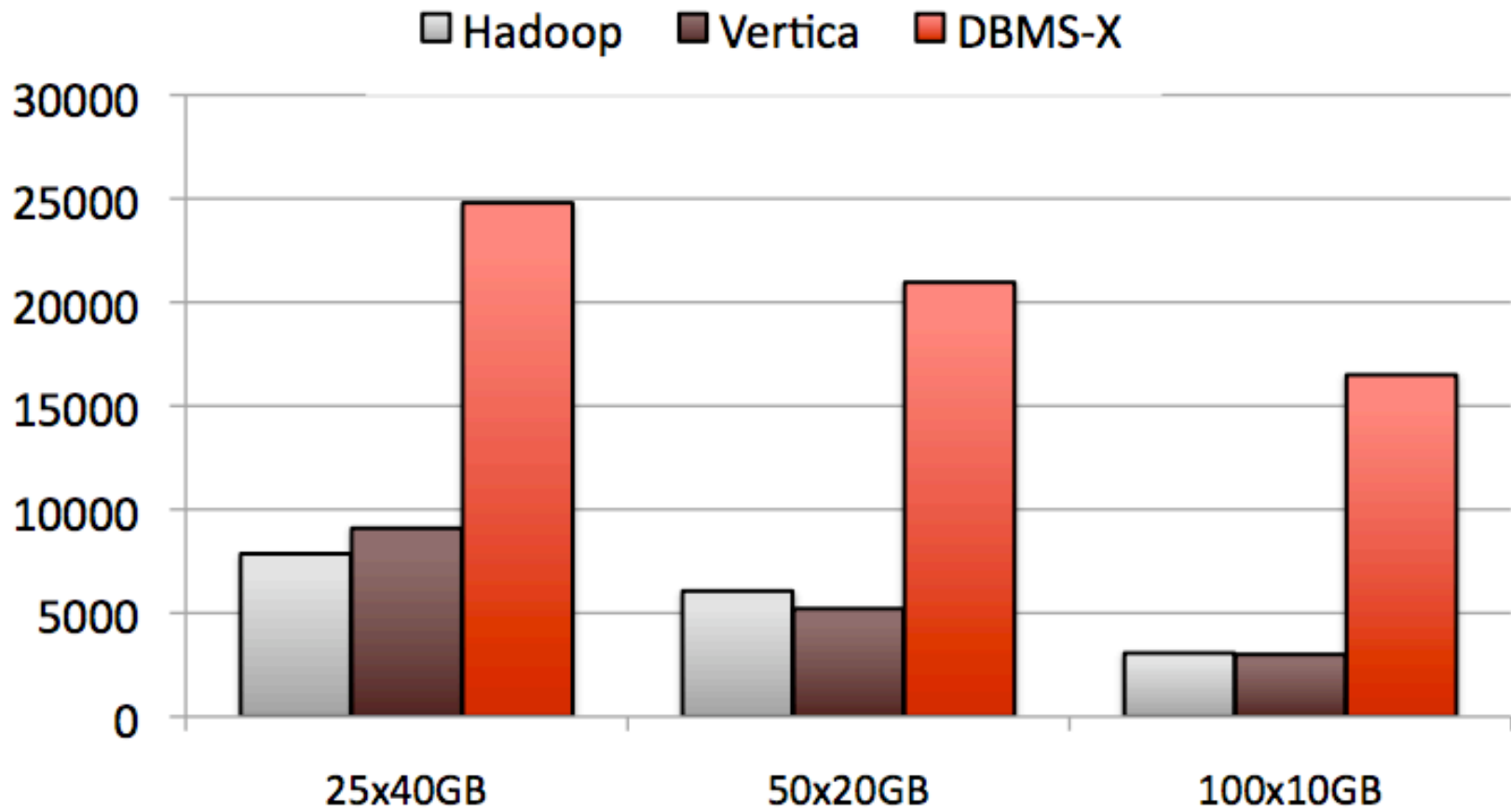
Grep Task Execution Results

Selection Task

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


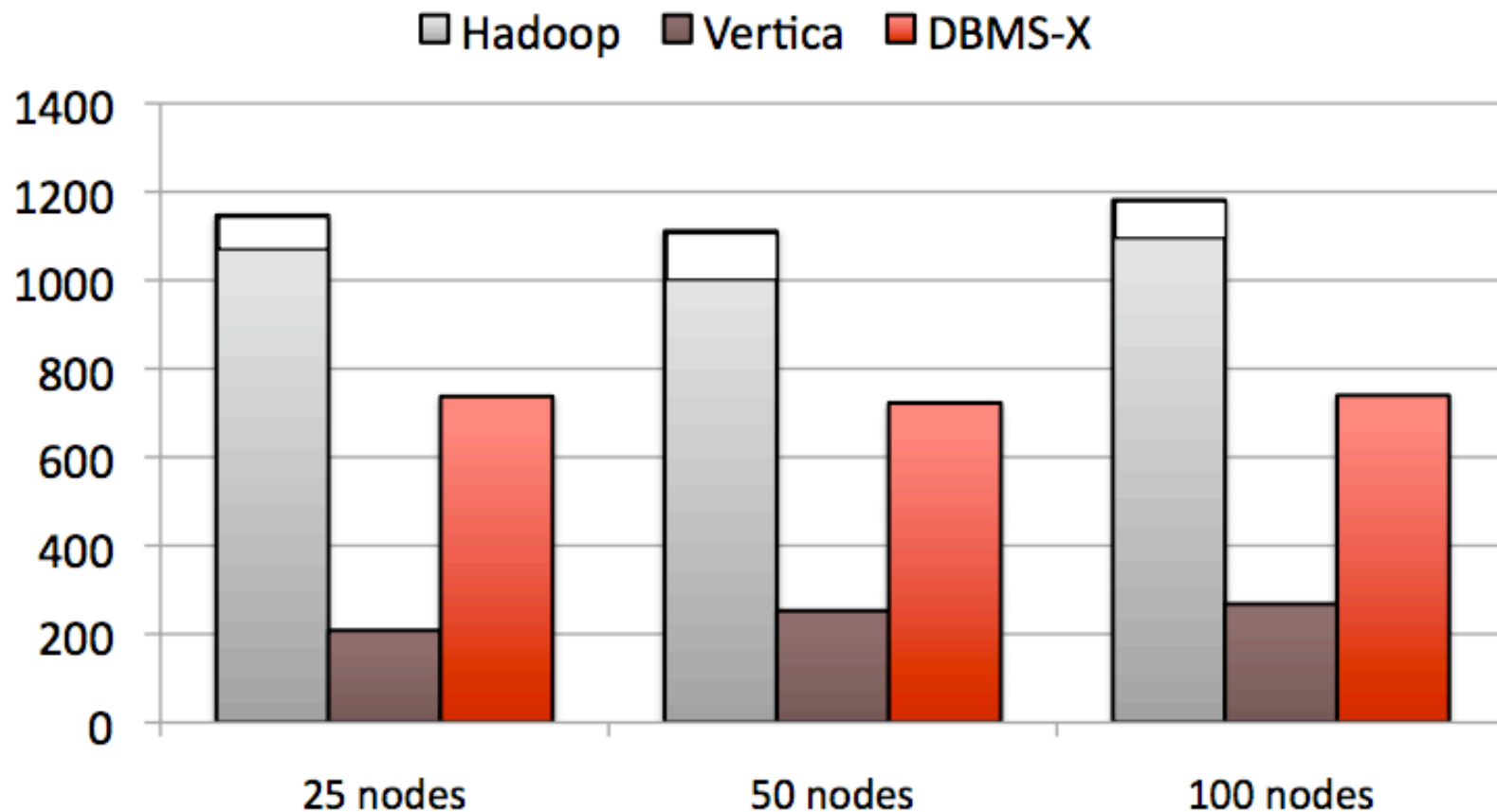1 GB / node

Analytical Tasks

- **Simple web processing 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 userVistits
 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:**

  - *DBMSs – Complex SQL using temporary table.*
  - *MapReduce – Three separate MR programs.*

Join Task

```sql
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
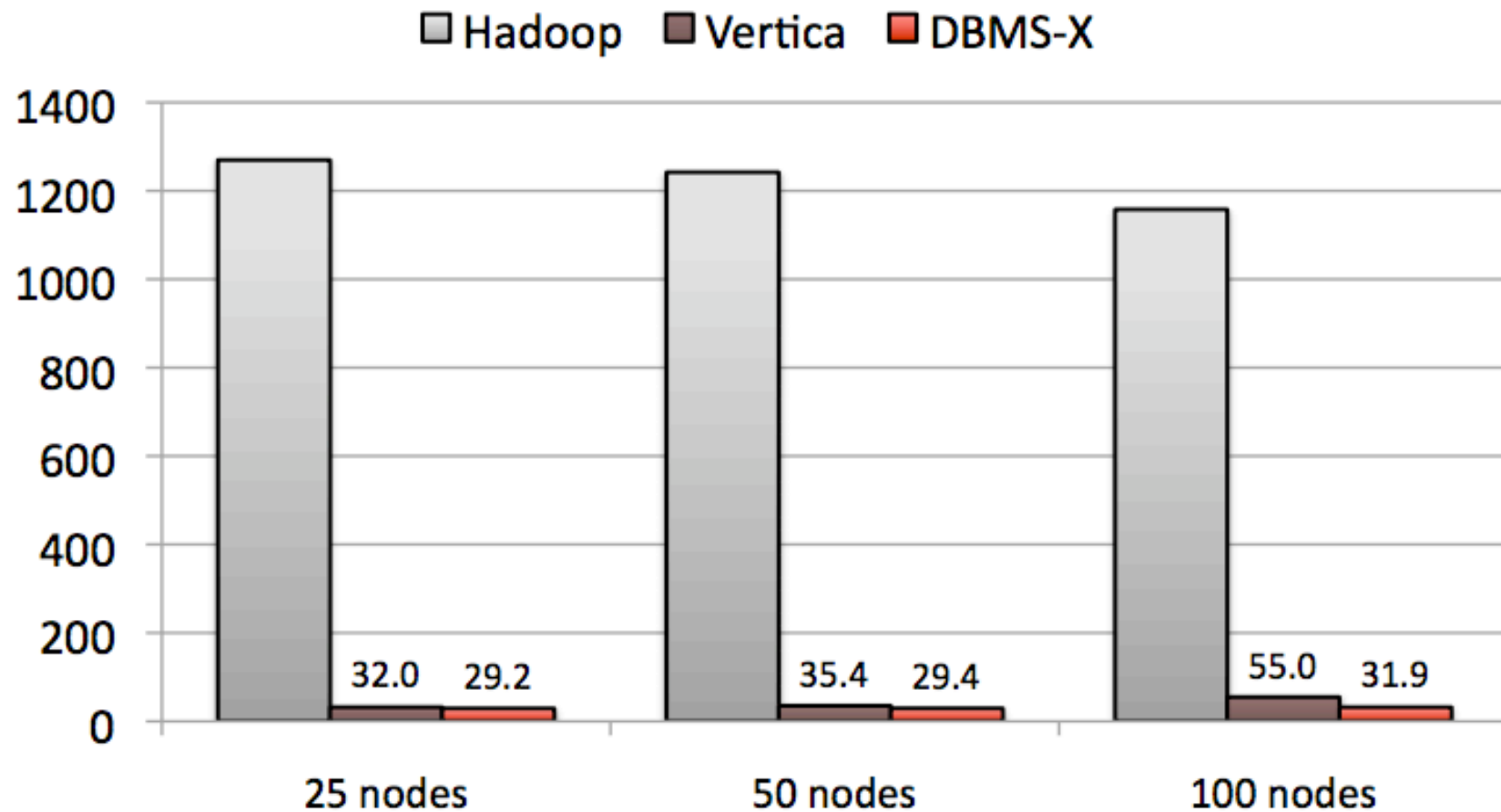
Problems with this analysis?

- 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
   http://developer.yahoo.net/blogs/hadoop/2009/05/hadoop_sorts_a_petabyte_in_162.html
[3] Hive - A Petabyte Scale Data Warehouse using Hadoop – June 10th, 2009
   http://www.facebook.com/note.php?note_id=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.*

Other Benchmarked Systems

- **HadoopDB (Abadi '09 - Yale)**
  - *Replaced Hadoop filesystem with Postgres.*
  - *Makes JDBC calls inside of MR functions.*

- **Hive (Thusoo '09 - Facebook)**
  - *Data warehouse interface on top of Hadoop.*
  - *Converts SQL-like language to MR programs.*