UNIVERSITY OF WATERLOO

Data-Intensive Distributed Computing

CS 431/631 451/651 (Winter 2019)

Part 1: MapReduce Algorithm Design (2/4)

January 10, 2019

Adam Roegiest
Kira Systems

These slides are available at http://roegiest.com/bigdata-2019w/


RTFM you must

Announcements

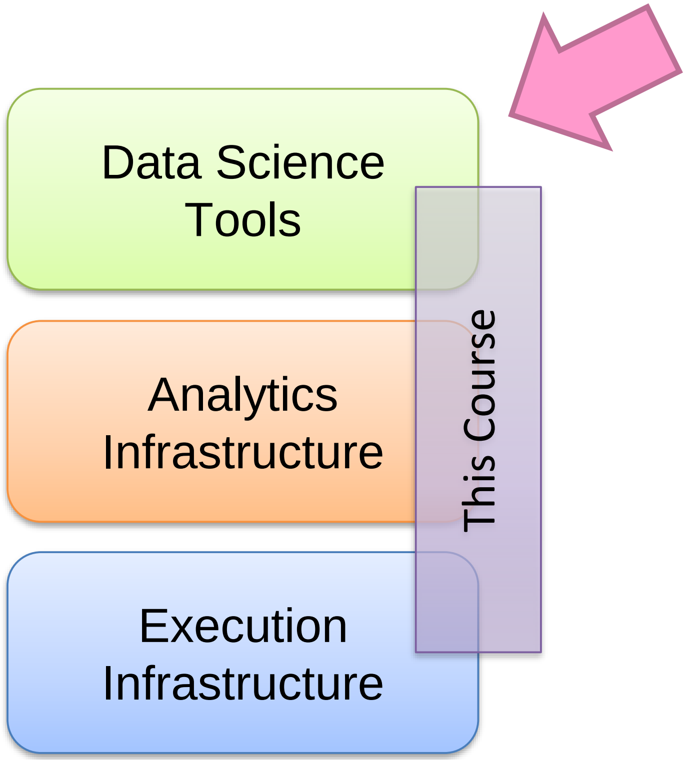
A0 finalized for all sections

CS 431/631 Only: If you want a challenge, you may elect to do **all of** the CS 451 assignments instead of the six CS 431 assignments.
This is a one-way road. No switching back.

Agenda for Today

Why big data?
Hadoop walkthrough

Why big data?



“big data stack”

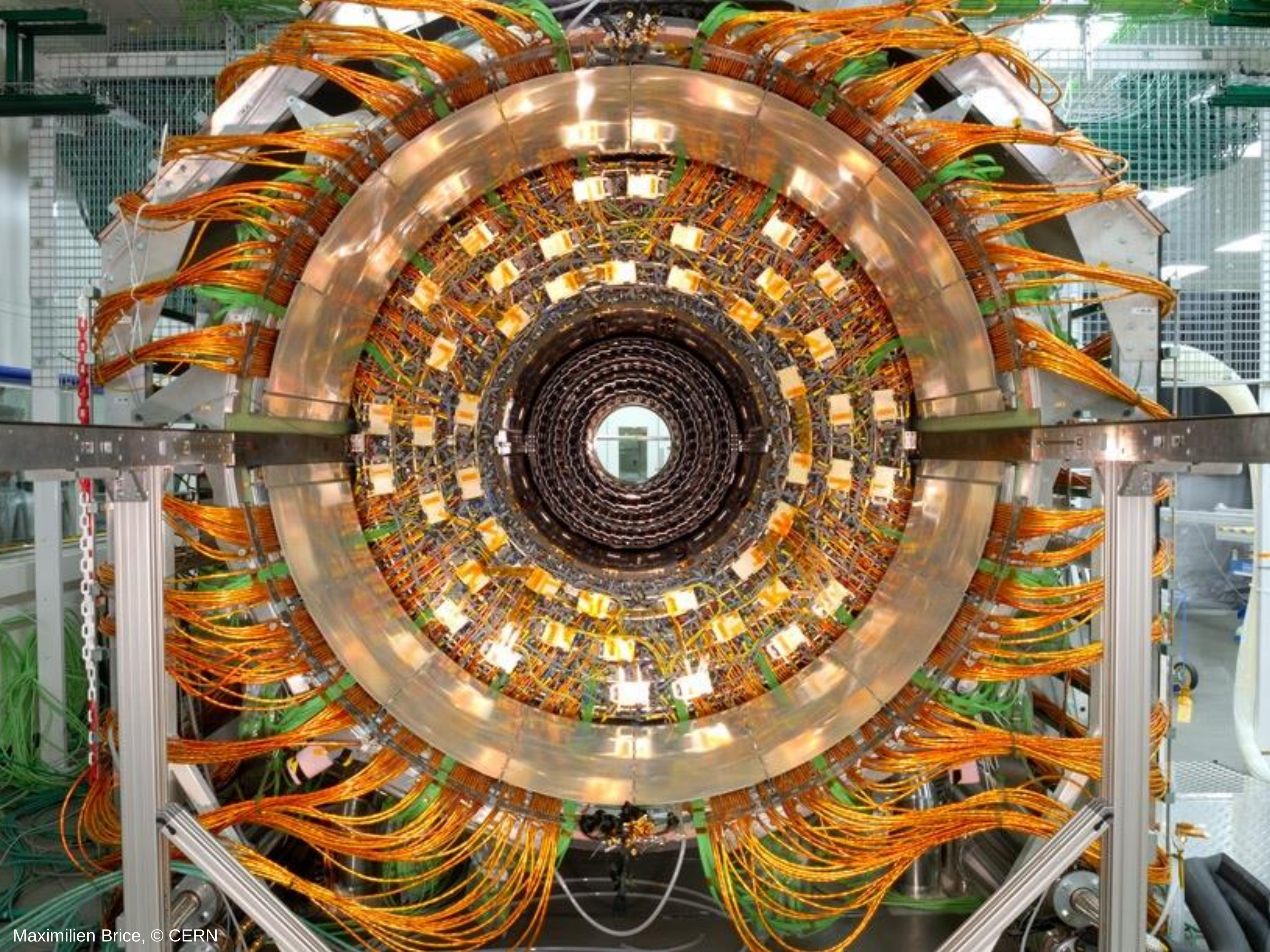

Why big data?
Science
Business
Society
Source: Wikipedia (Everest)

Science

Emergence of the 4th Paradigm

Data-intensive e-Science

Maximilien Brice, © CERN



Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC

The ATLAS Collaboration

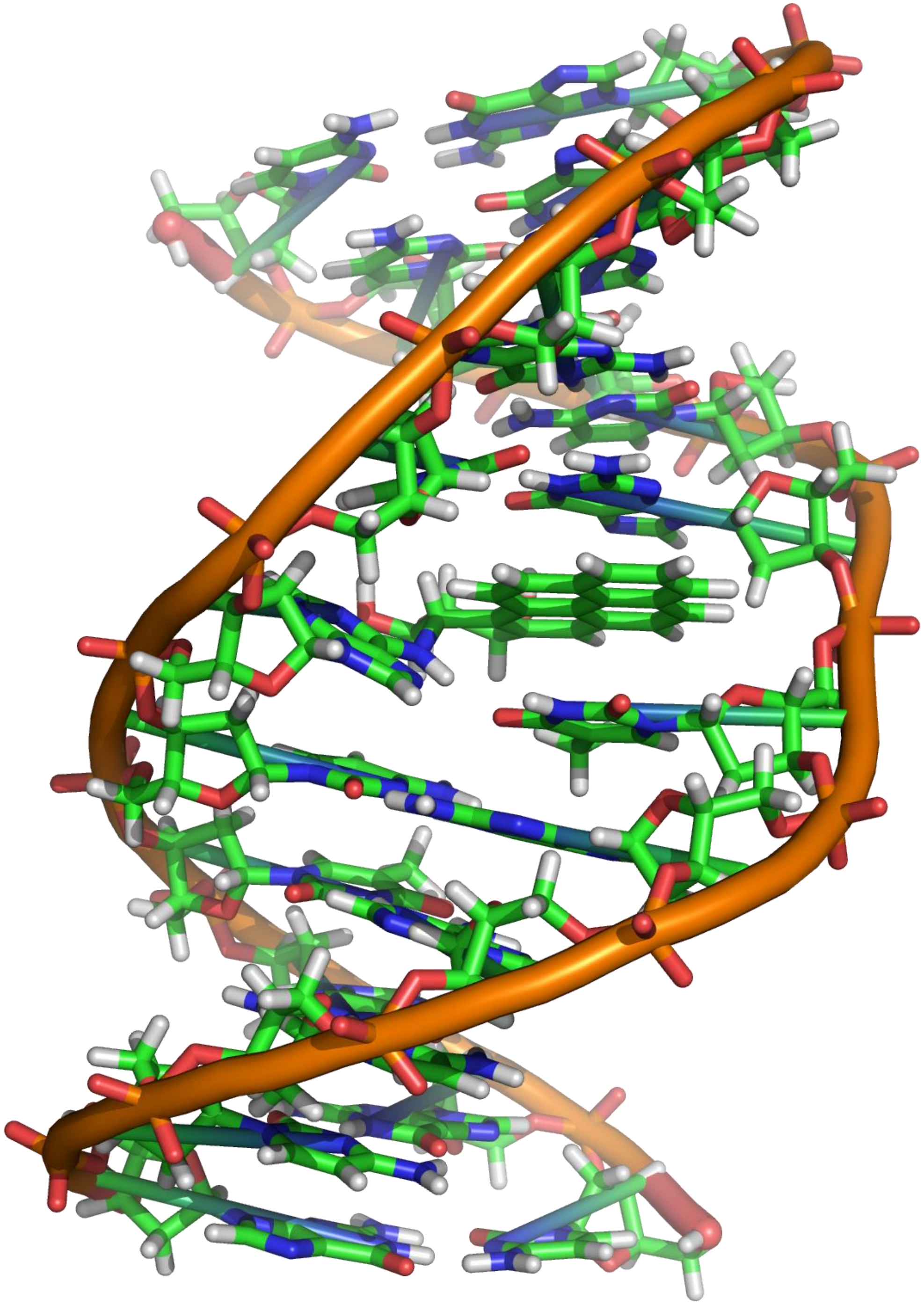
(Submitted on 31 Jul 2012 (v1), last revised 31 Aug 2012 (this version, v2))

A search for the Standard Model Higgs boson in proton-proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^-1 collected at sqrt(s) = 7 TeV in 2011 and 5.8 fb^-1 at sqrt(s) = 8 TeV in 2012. Individual searches in the channels H->ZZ^(*)->llll, H->gamma gamma and H->WW->e nu mu nu in the 8 TeV data are combined with previously published results of searches for H->ZZ^(*), WW^(*), bbbar and tau^+tau^- in the 7 TeV data and results from improved analyses of the H->ZZ^(*)->llll and H->gamma gamma channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of 126.0 +/- 0.4(stat) +/- 0.4(sys) GeV is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7x10^-9, is compatible with the production and decay of the Standard Model Higgs boson.

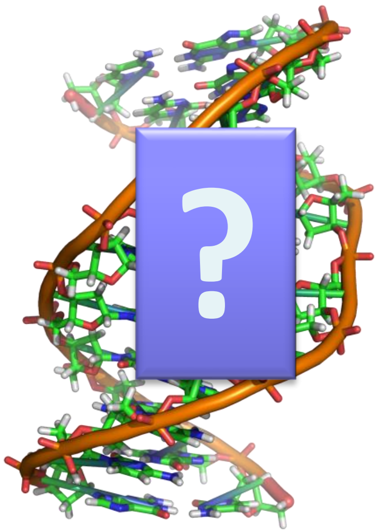
Comments:	24 pages plus author list (38 pages total), 12 figures, 7 tables, revised author list, matches version to appear in Physics Letters B
Subjects:	**High Energy Physics - Experiment (hep-ex)**
Journal reference:	Phys.Lett. B716 (2012) 1-29
DOI:	10.1016/j.physletb.2012.08.020
Report number:	CERN-PH-EP-2012-218
Cite as:	**arXiv:1207.7214 [hep-ex]**
	(or **arXiv:1207.7214v2 [hep-ex]** for this version)

Submission history

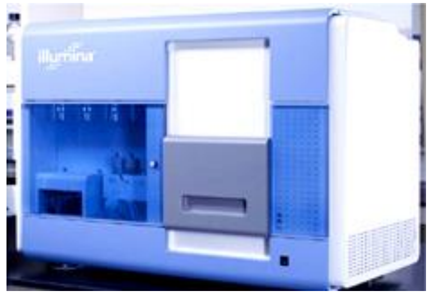
From: Atlas Publications [view email]
[v1] Tue, 31 Jul 2012 11:59:59 GMT (334kb)
[v2] Fri, 31 Aug 2012 19:29:54 GMT (334kb)



Source: Wikipedia (DNA)



Subject genome



Sequencer

GATGCTTACTATGCGGGCCCC
CGGTCTAATGCTTACTATGC
GCTTACTATGCGGGCCCCTT
AATGCTTACTATGCGGGCCCCTT
TAATGCTTACTATGC
AATGCTTAGCTATGCGGGC
AATGCTTACTATGCGGGCCCCTT
AATGCTTACTATGCGGGCCCCTT
CGGTCTAGATGCTTACTATGC
AATGCTTACTATGCGGGCCCCTT
CGGTCTAATGCTTAGCTATGC
ATGCTTACTATGCGGGCCCCTT

Reads

Human genome: 3 gbp
A few billion short reads
(~100 GB compressed data)

Business

Data-driven decisions

Data-driven products

Source: Wikiedia (Shinjuku, Tokyo)

Business Intelligence

An organization should retain data that result from carrying out its mission and exploit those data to generate insights that benefit the organization, for example, market analysis, strategic planning, decision making, etc.

Duh!?

This is not a new idea!

In the 1990s, Wal-Mart found that customers tended to buy diapers and beer together. So they put them next to each other and increased sales of both.*

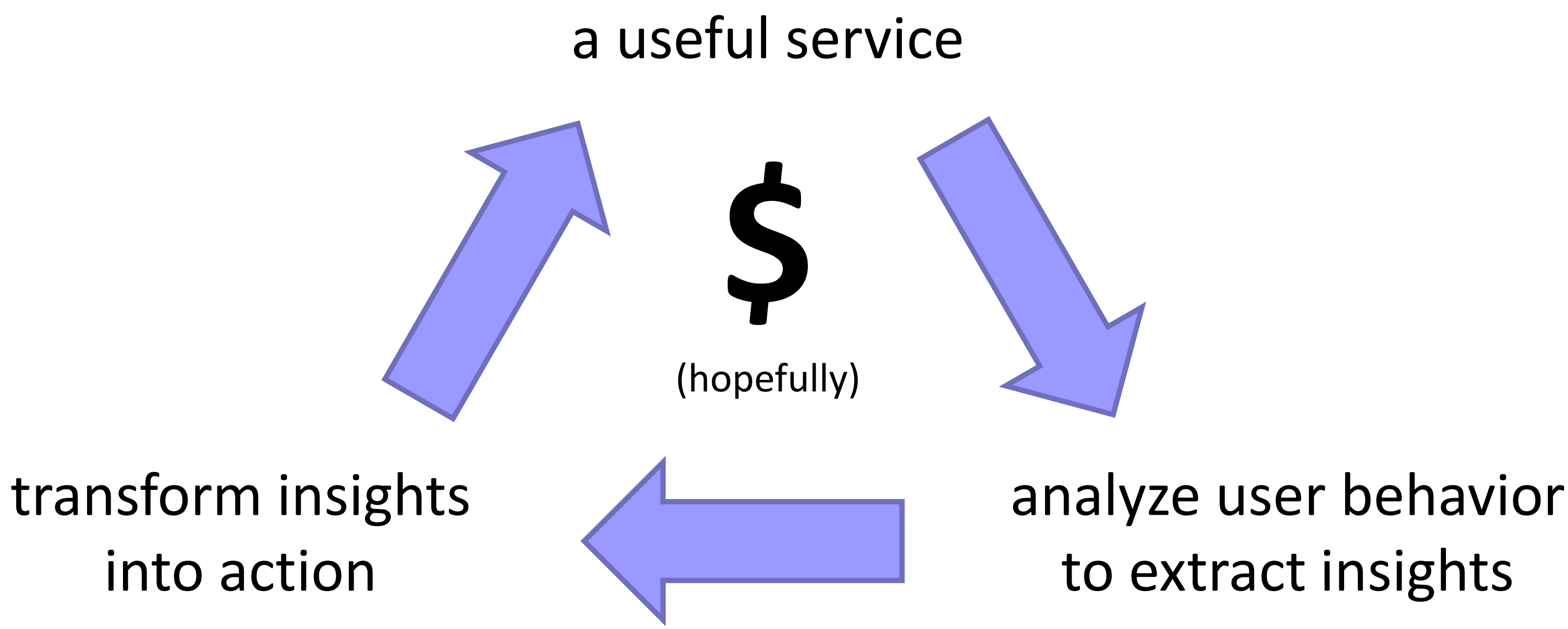
So what's changed?

More compute and storage

Ability to gather behavioral data

* BTW, this is completely apocryphal. (But it makes a nice story.)

Virtuous Product Cycle

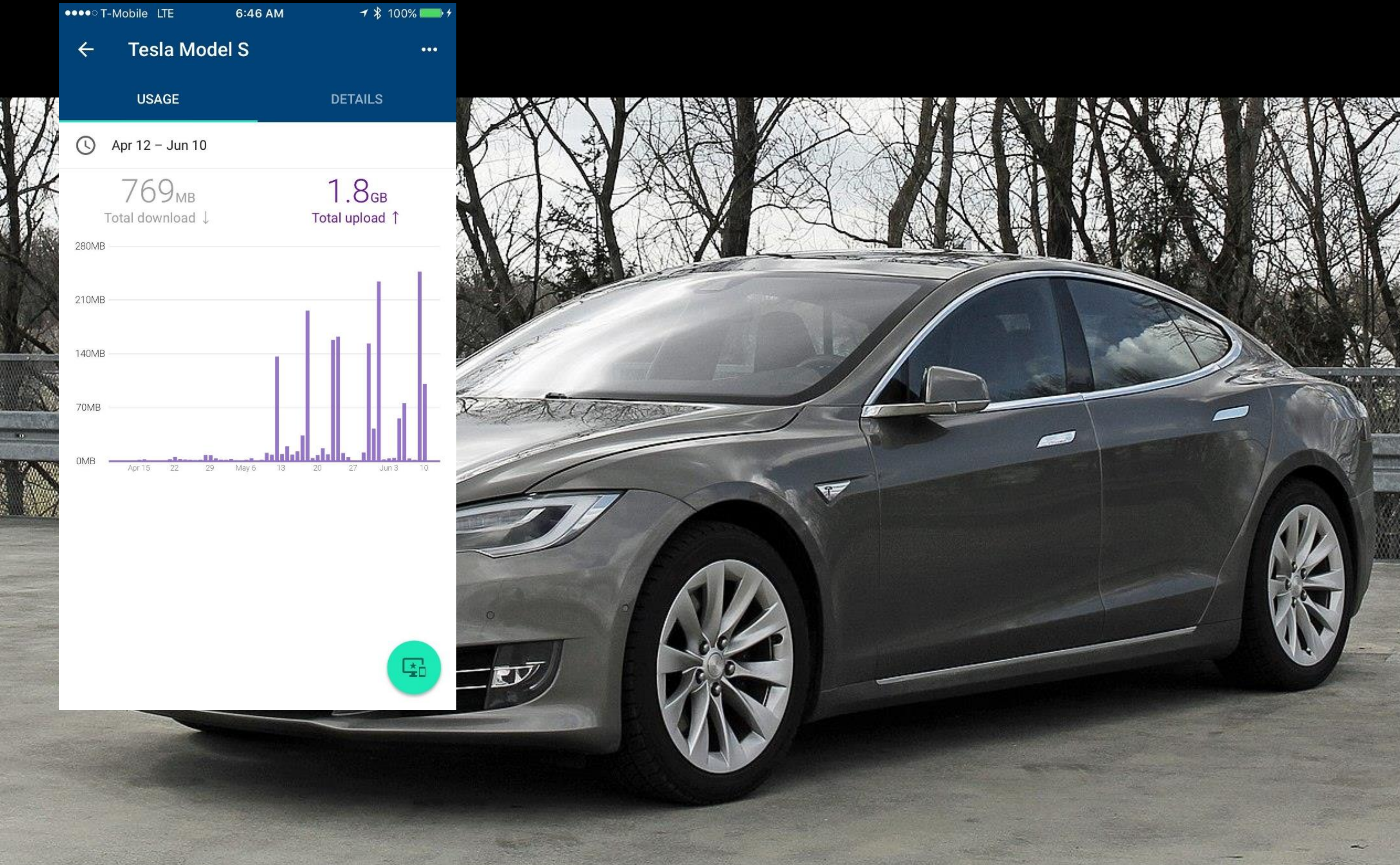


Google. Facebook. Twitter. Amazon. Uber.

data products

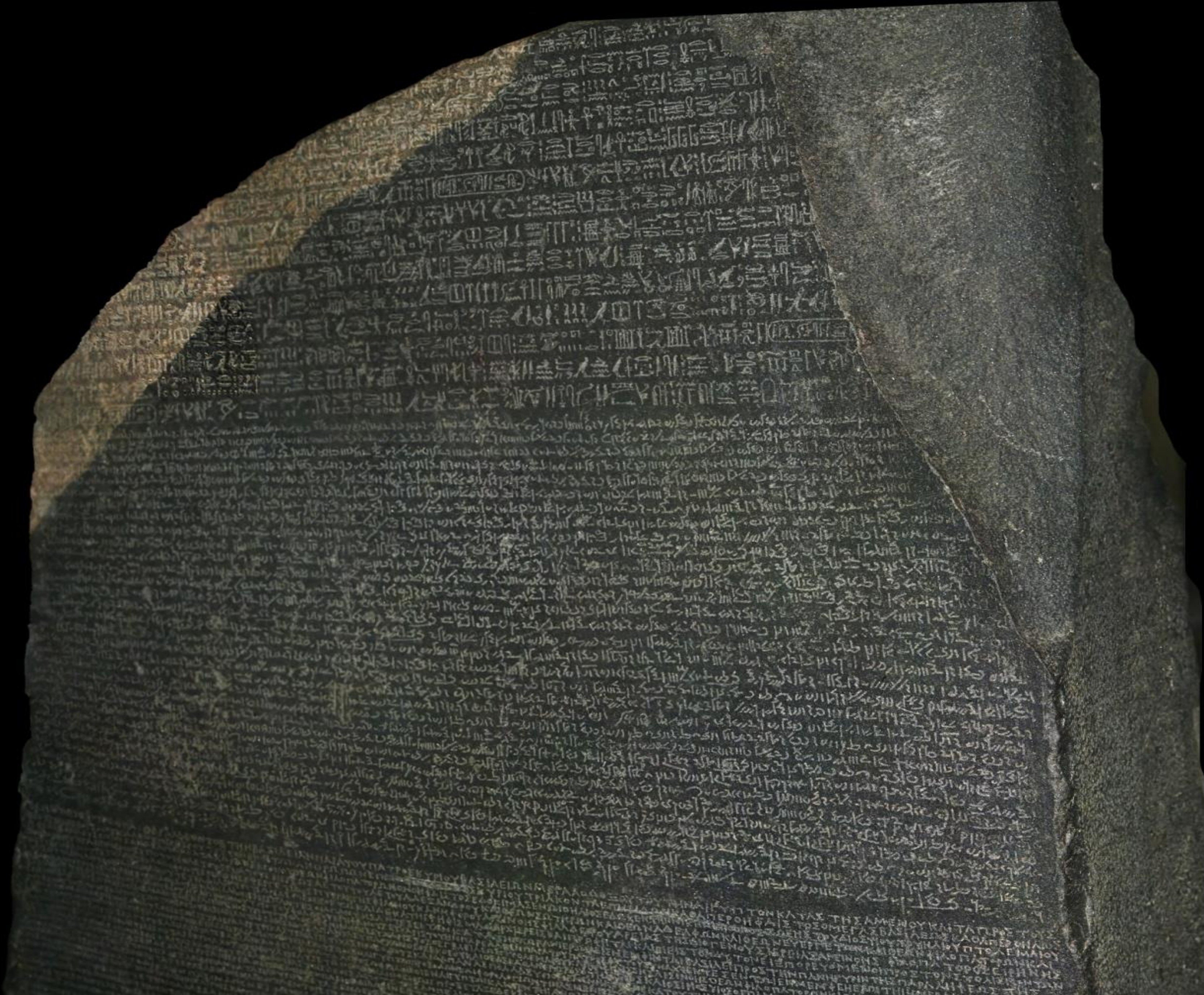
data science

Source: https://images.lookhuman.com/render/standard/8002245806006052/pillow14in-whi-z1-t-netflixing.png



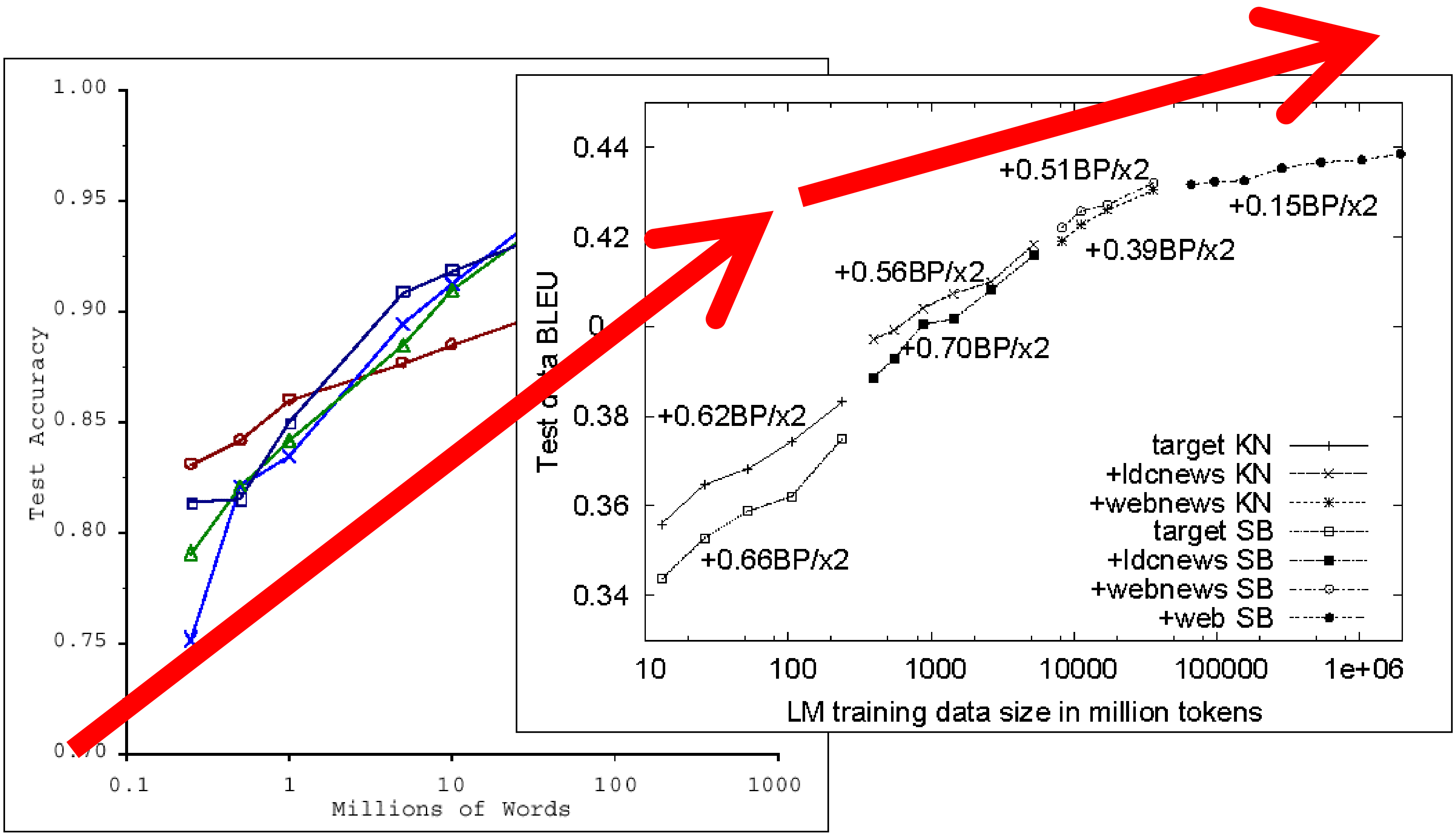
Source: https://www.reddit.com/r/teslamotors/comments/6gsc6v/i_think_the_neural_net_mining_is_just_starting/ (June 2017)


Translate
Turn off instant translation
Chinese
English
Spanish
Detect language
English
Chinese (Traditional)
Chinese (Simplified)
Translate
How does Google translate English into Chinese?
47/5000
Google如何将英语翻译成中文?
Suggest an edit
Google rúhé jiāng yīngyǔ fānyì chéng zhōngwén?
Google Translate for Business:
Translator Toolkit
Website Translator



Source: Wikipedia (Rosetta Stone)

No data like more data!



(Banko and Brill, ACL 2001)
(Brants et al., EMNLP 2007)

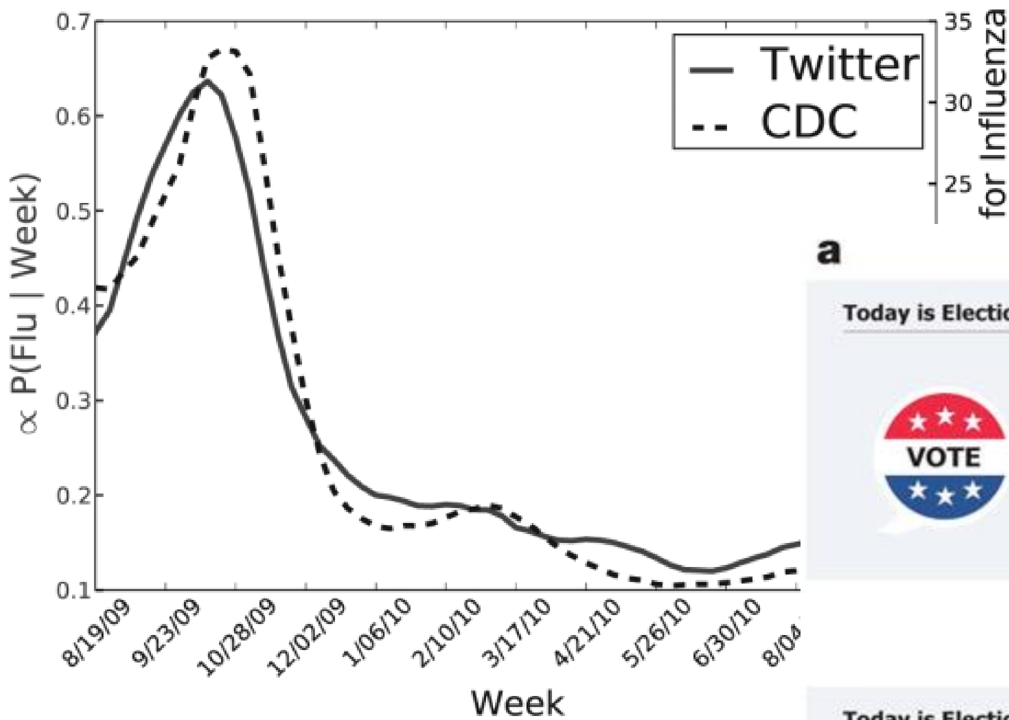
Society

Humans as social sensors

Computational social science

Source: Guardian

Predicting *X* with Twitter



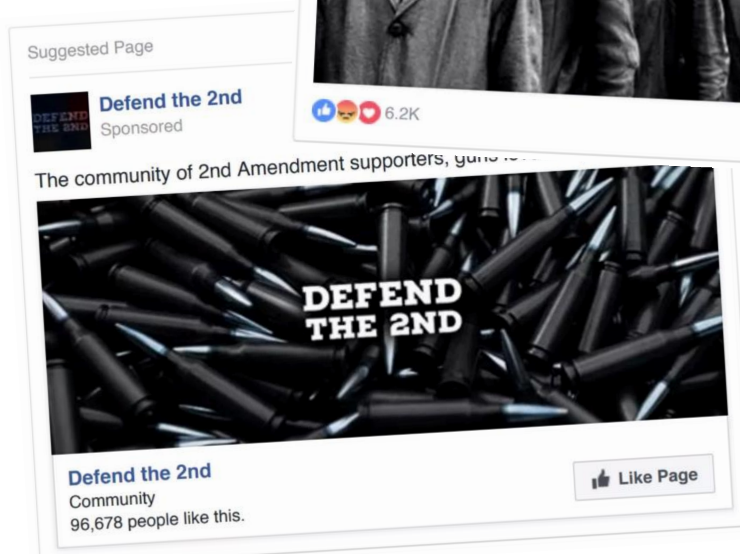
2010 US Midterm Elections:
60m users shown "I Voted" Messages

Summary: increased turnout by
60k directly and 280k indirectly

Woah! You should feel unsettled about this!

Political Mobilization on Facebook

(Paul and Dredze, ICWSM 2011; Bond et al., Nature 2011)



Source: https://www.theverge.com/2017/11/1/16593346/house-russia-facebook-ads

MACHINE BIAS

Facebook Enabled Advertisers to Reach 'Jew Haters'

After being contacted by ProPublica, Facebook removed several anti-Semitic ad categories and promised to improve monitoring.

by **Julia Angwin**, **Madeleine Varner** and **Ariana Tobin**, Sept. 14, 2017, 4 p.m. EDT

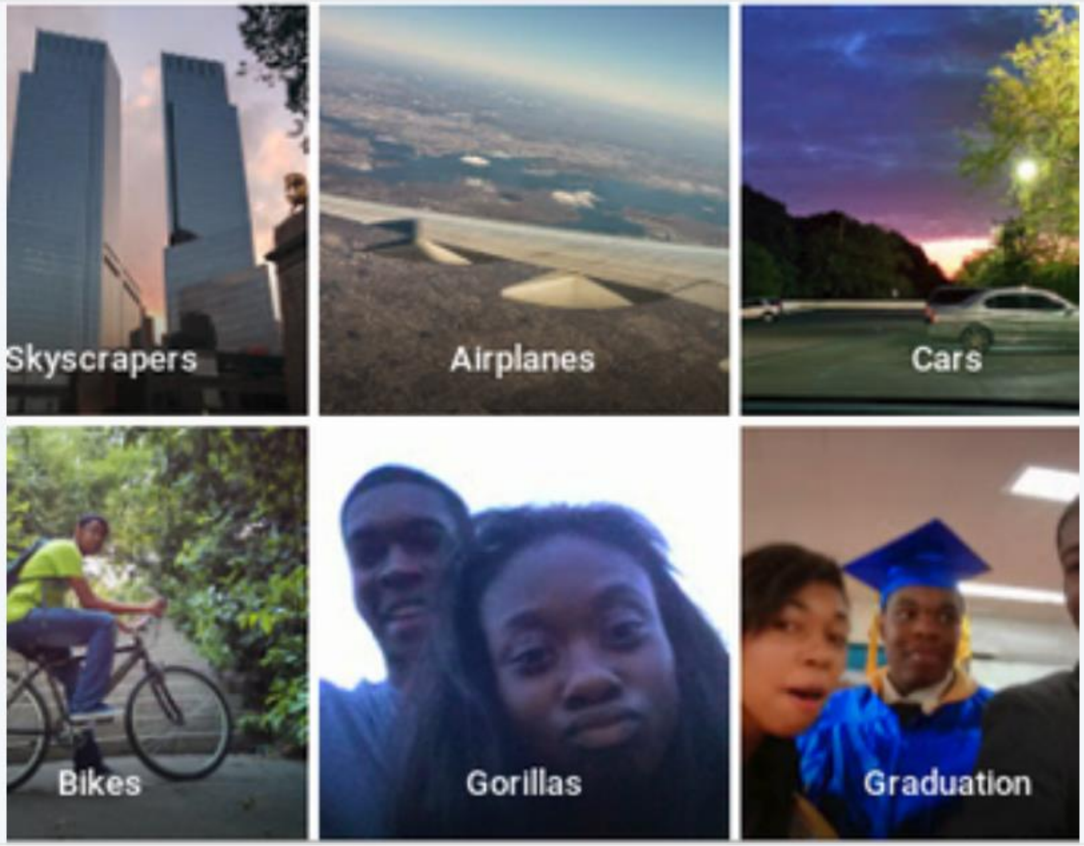
MACHINE BIAS
Investigating Algorithmic Injustice

Want to market Nazi memorabilia, or recruit marchers for a far-right rally? Facebook's self-service ad-buying platform had the right audience for you.

Until this week, when we asked Facebook about it, the world's largest social network enabled advertisers to direct their pitches to the news feeds of almost 2,300 people who expressed interest in the topics of "Jew hater," "How to burn jews," or, "History of 'why jews ruin the world.'"

To test if these ad categories were real, we paid $30 to target those groups with three "promoted posts" — in which a ProPublica article or post was displayed in their news feeds. Facebook approved all three ads within 15 minutes.

Source: https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters



Jacky
@jackyalcine

Google Photos, y'all fucked up. My friend's not a gorilla.

8:22 PM - Jun 28, 2015

225 3,209 2,059

× The photo you want to upload does not meet our criteria because:

- Subject eyes are closed

Please refer to the technical requirements.
You have 9 attempts left.

Check the photo requirements.

Read more about common photo problems and how to resolve them.

After your tenth attempt you will need to start again and re-enter the CAPTCHA security check.

Reference number: 20161206-81

Filename: Untitled.jpg

If you wish to contact us about the photo, you must provide us with the reference number given above.

INSTAGRAM.COM/RICHYFANCY

Source: http://www.cnn.com/2016/12/07/asia/new-zealand-passport-robot-asian-trnd/index.html

The Perils of Big Data

The end of privacy

Who owns your data and can the government access it?

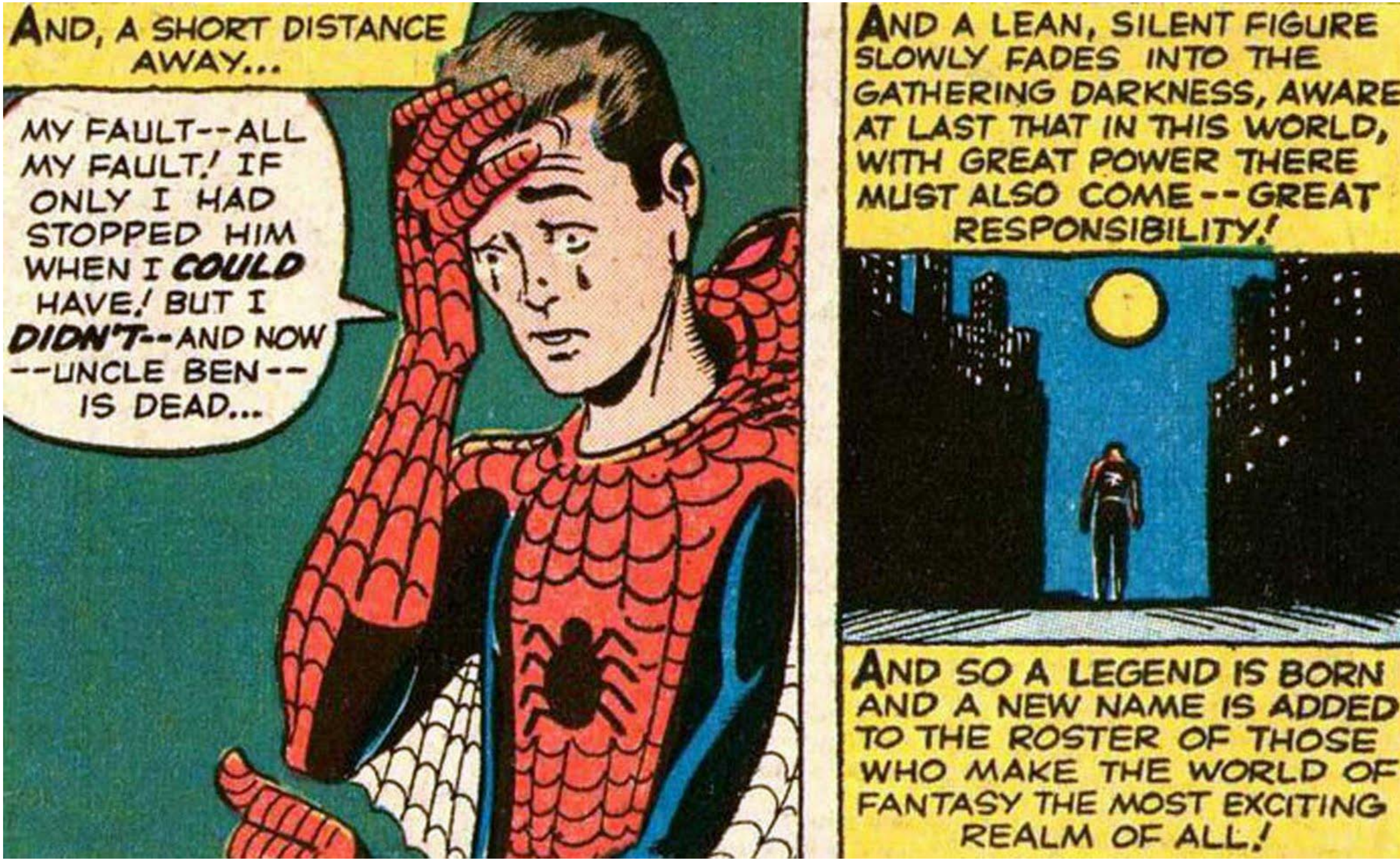
The echo chamber

Are you seeing only what you want to see?

The racist algorithm

Algorithms aren't racist, people are?

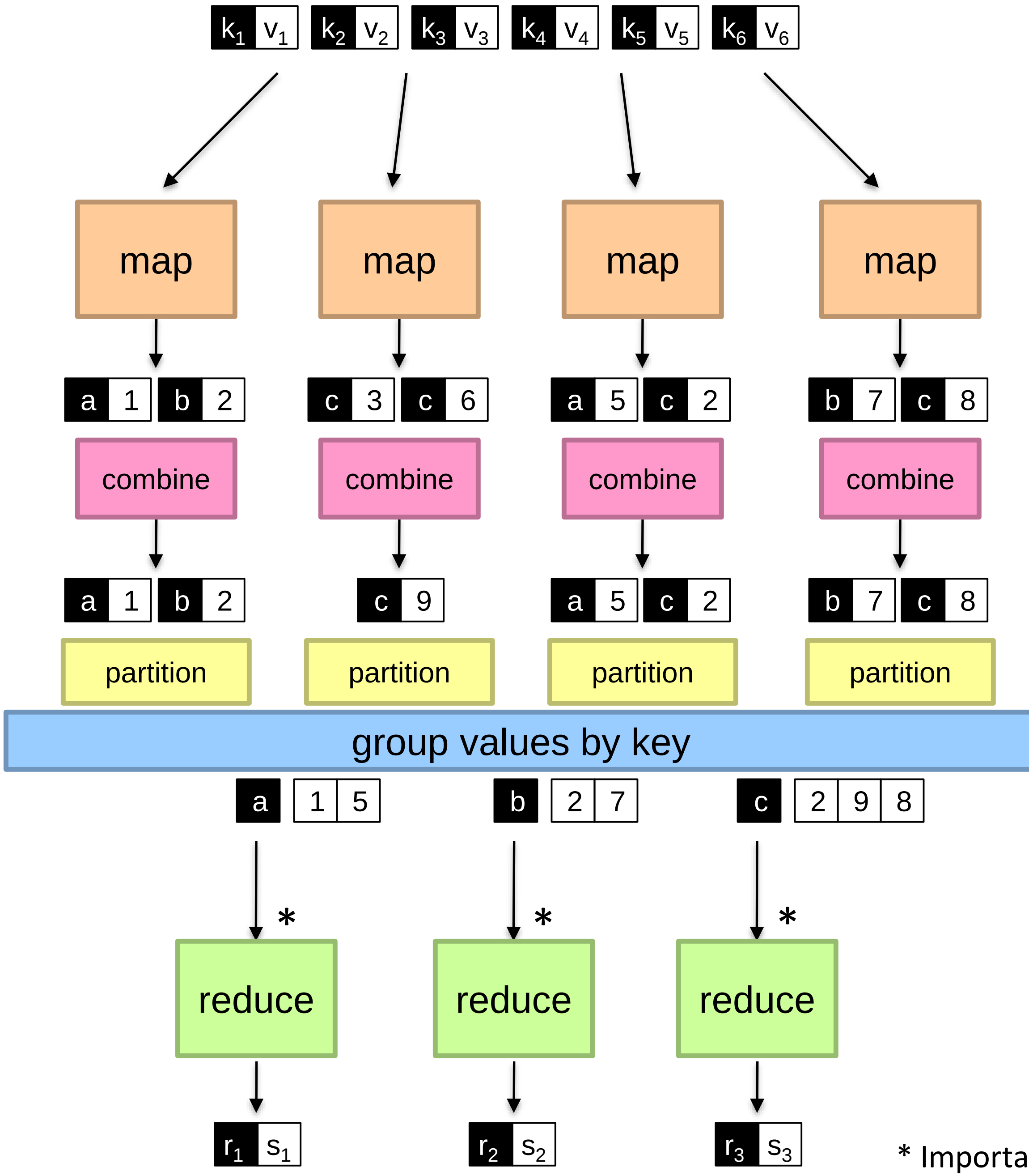
We *desperately* need "data ethics" to go with big data!


AND, A SHORT DISTANCE AWAY...
MY FAULT--ALL MY FAULT! IF ONLY I HAD STOPPED HIM WHEN I COULD HAVE! BUT I DIDN'T--AND NOW --UNCLE BEN-- IS DEAD...
AND A LEAN, SILENT FIGURE SLOWLY FADES INTO THE GATHERING DARKNESS, AWARE AT LAST THAT IN THIS WORLD, WITH GREAT POWER THERE MUST ALSO COME--GREAT RESPONSIBILITY!
AND SO A LEGEND IS BORN AND A NEW NAME IS ADDED TO THE ROSTER OF THOSE WHO MAKE THE WORLD OF FANTASY THE MOST EXCITING REALM OF ALL!

Source: Popular Internet Meme

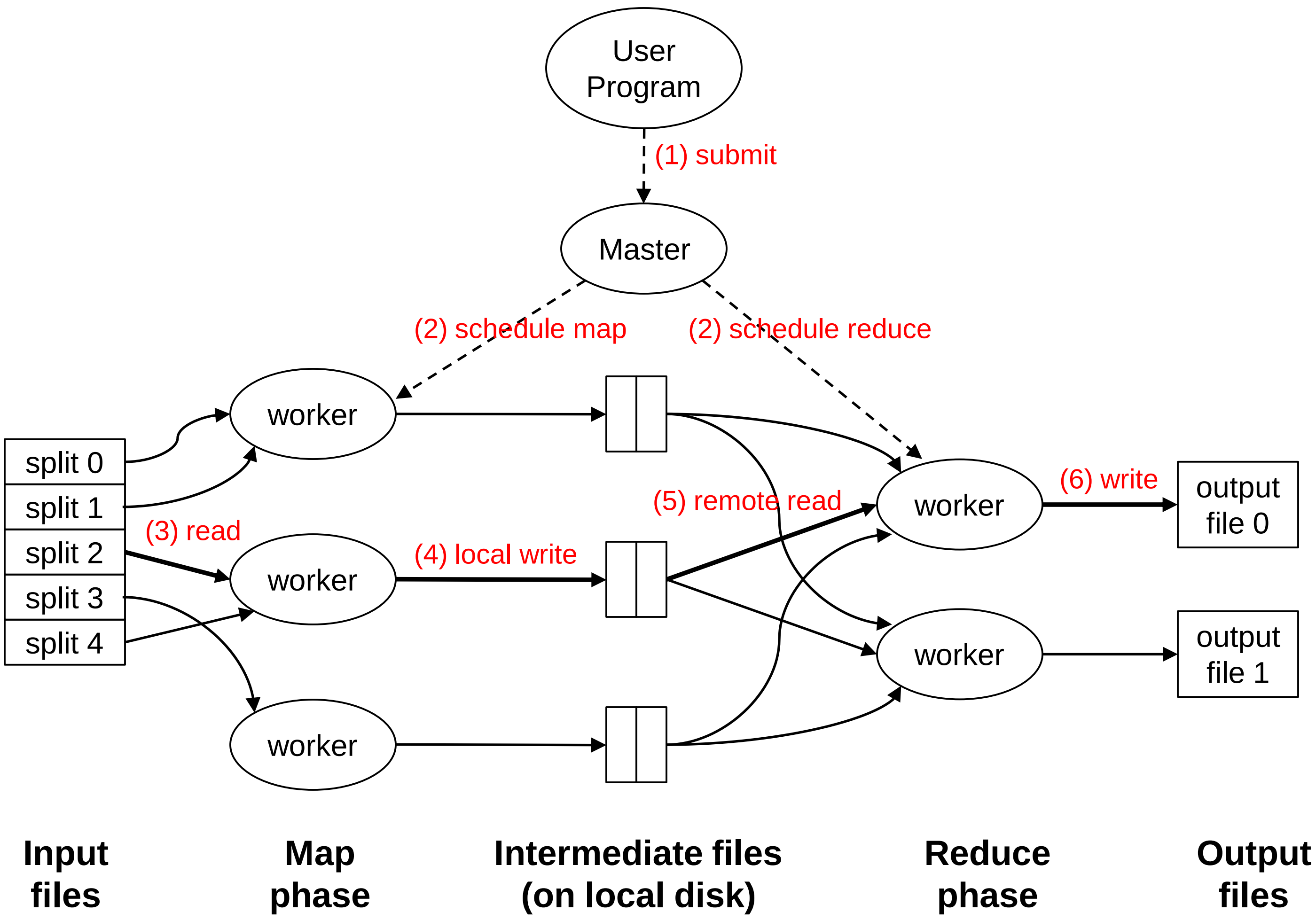

Tackling Big Data
Source: Google

Logical View



* Important detail: reducers process keys in sorted order

Physical View



Adapted from (Dean and Ghemawat, OSDI 2004)


The datacenter *is* the computer!
Source: Google

The datacenter *is* the computer!

It's all about the right level of abstraction

Moving beyond the von Neumann architecture

What's the "instruction set" of the datacenter computer?

Hide system-level details from the developers

No more race conditions, lock contention, etc.

No need to explicitly worry about reliability, fault tolerance, etc.

Separating the *what* from the *how*

Developer specifies the computation that needs to be performed

Execution framework ("runtime") handles actual execution

The datacenter *is* the computer!
"Big ideas"

Scale "out", not "up" *

Limits of SMP and large shared-memory machines

Assume that components will break

Engineer software around hardware failures

Move processing to the data *

Cluster have limited bandwidth, code is a lot smaller

Process data sequentially, avoid random access

Seeks are expensive, disk throughput is good

Seek vs. Scans

Consider a 1 TB database with 100 byte records

We want to update 1 percent of the records

Scenario 1: Mutate each record

Each update takes ~30 ms (seek, read, write)

10^8 updates = ~35 days

Scenario 2: Rewrite all records

Assume 100 MB/s throughput

Time = 5.6 hours(!)

Lesson? Random access is expensive!

Source: Ted Dunning, on Hadoop mailing list

Source: Wikipedia (Mahout)

So you want to drive the elephant!

(Aside, what about Spark?)

A tale of two packages…

org.apache.hadoop.mapreduce
org.apache.hadoop.mapred

Source: Wikipedia (Budapest)

MapReduce API*

Mapper<K_{in},V_{in},K_{out},V_{out}>

void setup(Mapper.Context context)

Called once at the start of the task

void map(K_{in} key, V_{in} value, Mapper.Context context)

Called once for each key/value pair in the input split

void cleanup(Mapper.Context context)

Called once at the end of the task

Reducer<K_{in},V_{in},K_{out},V_{out}>/Combiner<K_{in},V_{in},K_{out},V_{out}>

void setup(Reducer.Context context)

Called once at the start of the task

void reduce(K_{in} key, Iterable<V_{in}> values, Reducer.Context context)

Called once for each key

void cleanup(Reducer.Context context)

Called once at the end of the task

*Note that there are two versions of the API!

MapReduce API*

Partitioner<K, V>

int getPartition(K key, V value, int numPartitions)

Returns the partition number given total number of partitions

Job

Represents a packaged Hadoop job for submission to cluster

Need to specify input and output paths

Need to specify input and output formats

Need to specify mapper, reducer, combiner, partitioner classes

Need to specify intermediate/final key/value classes

Need to specify number of reducers (but not mappers, why?)

Don't depend of defaults!

*Note that there are two versions of the API!

Data Types in Hadoop: Keys and Values

Writable

Defines a de/serialization protocol.
Every data type in Hadoop is a Writable.

WritableComprable

Defines a sort order.
All keys must be of this type (but not values).

IntWritable
LongWritable
Text
…

Concrete classes for different data types.
Note that these are container objects.

SequenceFile

Binary-encoded sequence of key/value pairs.

“Hello World” MapReduce: Word Count

```
def map(key: Long, value: String) = {
  for (word <- tokenize(value)) {
    emit(word, 1)
  }
}


def reduce(key: String, values: Iterable[Int]) = {
  for (value <- values) {
    sum += value
  }
  emit(key, sum)
}
```

Word Count Mapper

```
private static final class MyMapper
    extends Mapper<LongWritable, Text, Text, IntWritable> {

   private final static IntWritable ONE = new IntWritable(1);
   private final static Text WORD = new Text();

   @Override
   public void map(LongWritable key, Text value, Context context)
       throws IOException, InterruptedException {
      for (String word : Tokenizer.tokenize(value.toString())) {
        WORD.set(word);
        context.write(WORD, ONE);
      }
   }
 }
```

Word Count Reducer

```
private static final class MyReducer
   extends Reducer<Text, IntWritable, Text, IntWritable> {

  private final static IntWritable SUM = new IntWritable();

  @Override
  public void reduce(Text key, Iterable<IntWritable> values,
     Context context) throws IOException, InterruptedException {
    Iterator<IntWritable> iter = values.iterator();
    int sum = 0;
    while (iter.hasNext()) {
      sum += iter.next().get();
    }
    SUM.set(sum);
    context.write(key, SUM);
  }
}
```

Three Gotchas

Avoid object creation

Execution framework reuses value object in reducer

Passing parameters via class statics doesn't work!

Getting Data to Mappers and Reducers

Configuration parameters

Pass in via Job configuration object

"Side data"

DistributedCache

Mappers/Reducers can read from HDFS in setup method

Complex Data Types in Hadoop

How do you implement complex data types?

The easiest way:

Encode it as Text, e.g., (a, b) = "a:b"

Use regular expressions to parse and extract data

Works, but janky

The hard way:

Define a custom implementation of Writable(Comprable)

Must implement: readFields, write, (compareTo)

Computationally efficient, but slow for rapid prototyping

Implement WritableComparator hook for performance

Somewhere in the middle:

Bespin (via lin.tl) offers various building blocks

Anatomy of a Job

Hadoop MapReduce program = Hadoop job

Jobs are divided into map and reduce tasks

An instance of a running task is called a task attempt

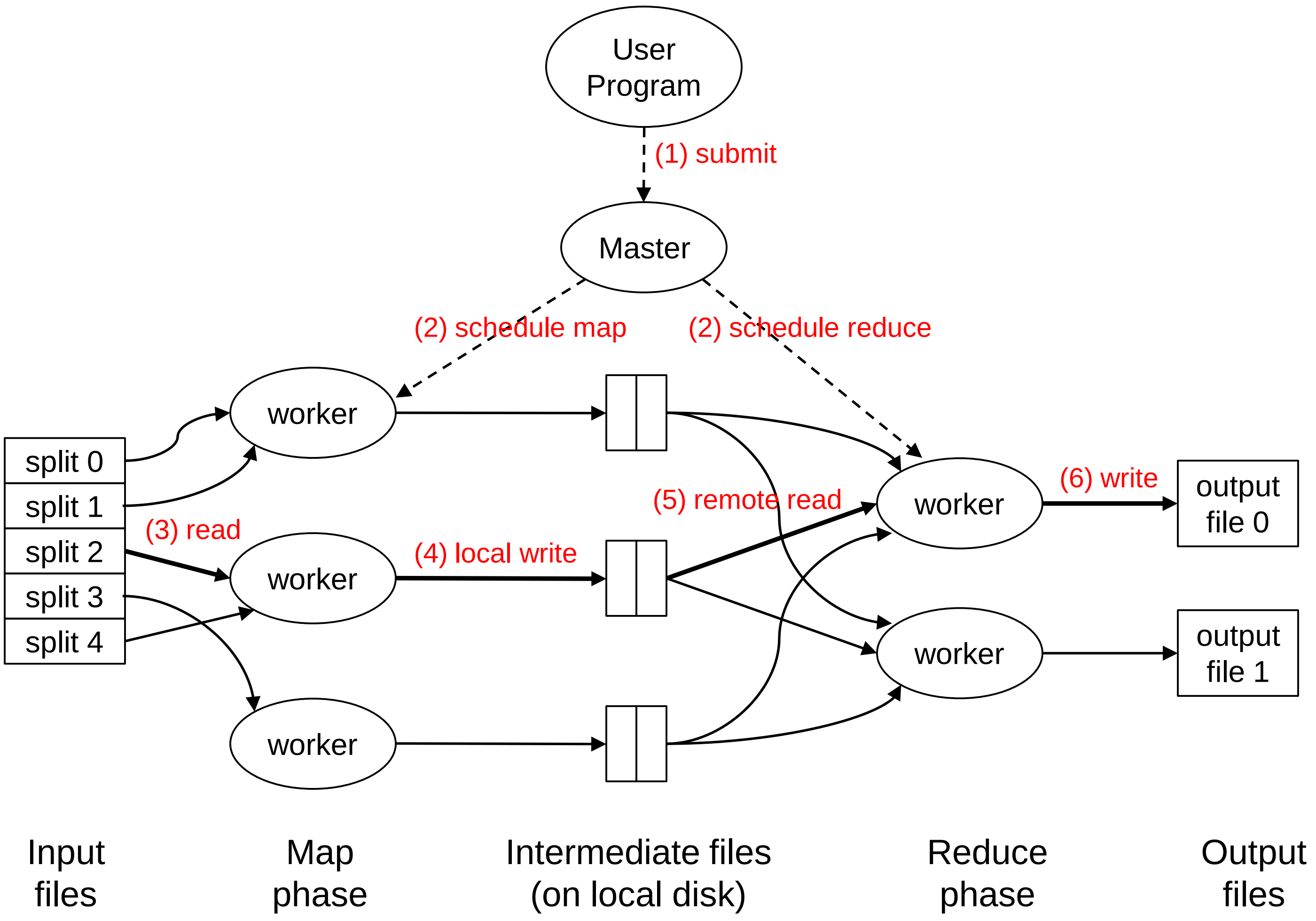
Each task occupies a slot on the tasktracker

Multiple jobs can be composed into a workflow

Job submission:

Client (i.e., driver program) creates a job,
configures it,
and submits it to jobtracker

That's it! The Hadoop cluster takes over…

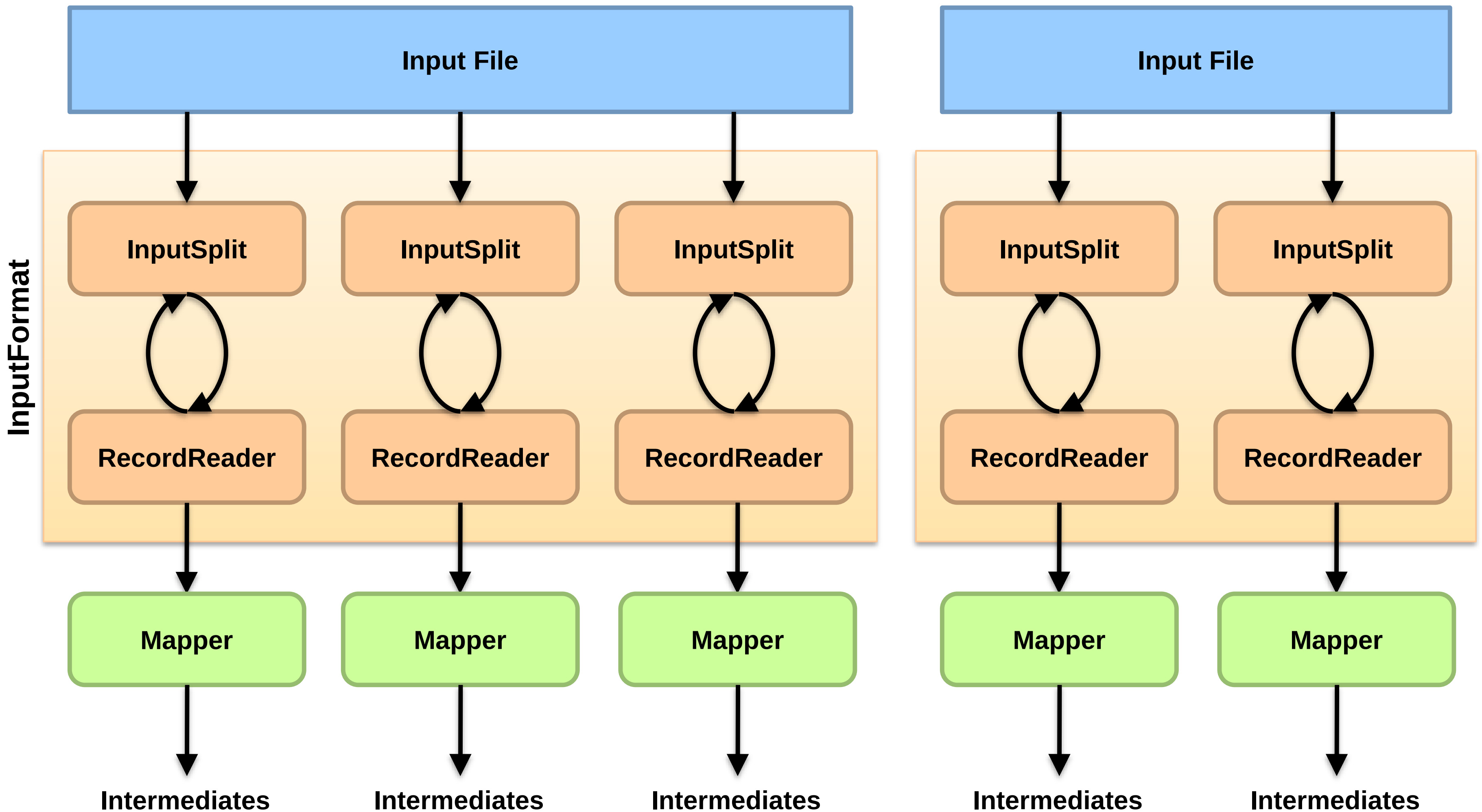


Adapted from (Dean and Ghemawat, OSDI 2004)

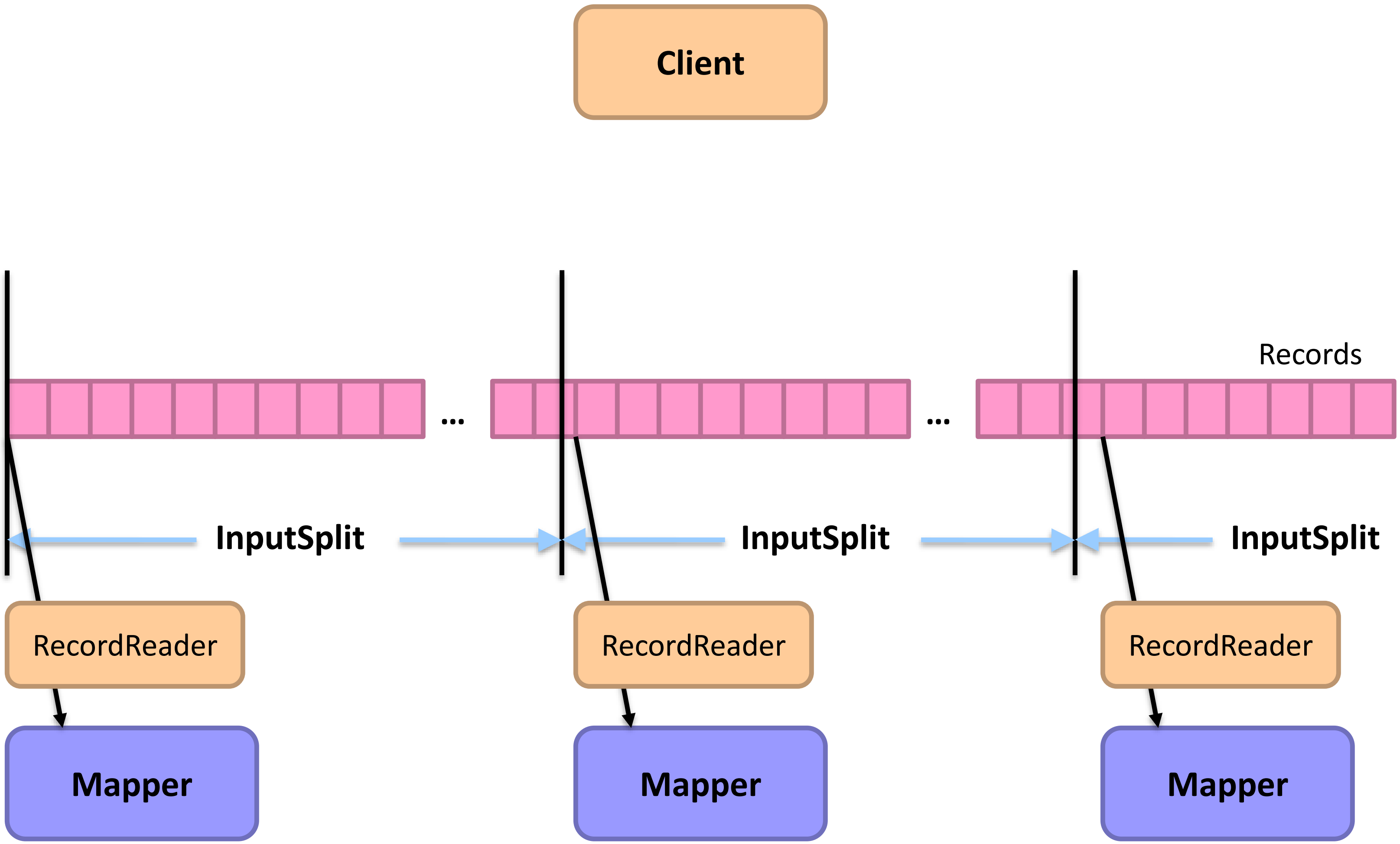
Anatomy of a Job

Behind the scenes:

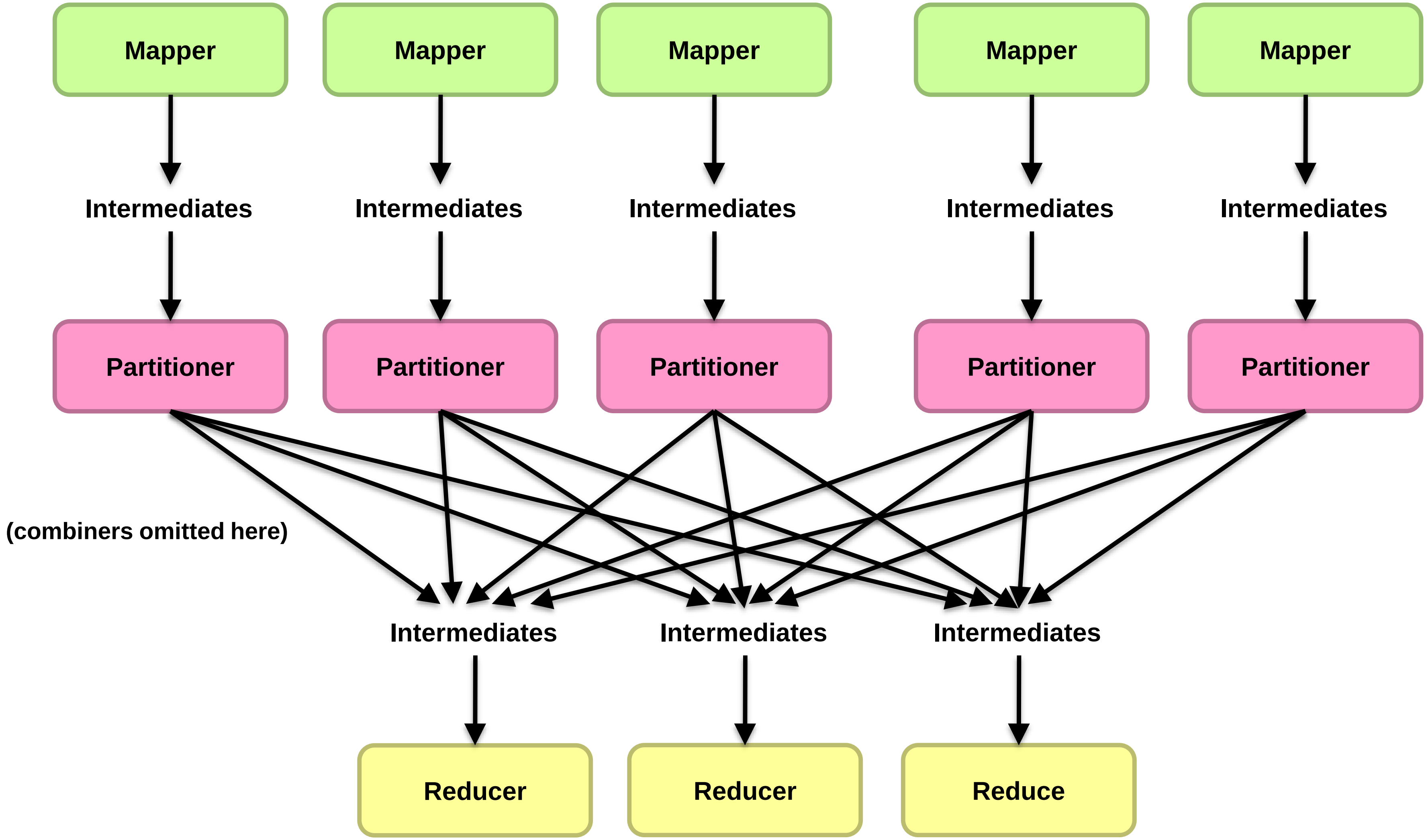
Input splits are computed (on client end)
Job data (jar, configuration XML) are sent to jobtracker
Jobtracker puts job data in shared location, enqueues tasks
Tasktrackers poll for tasks
Off to the races…



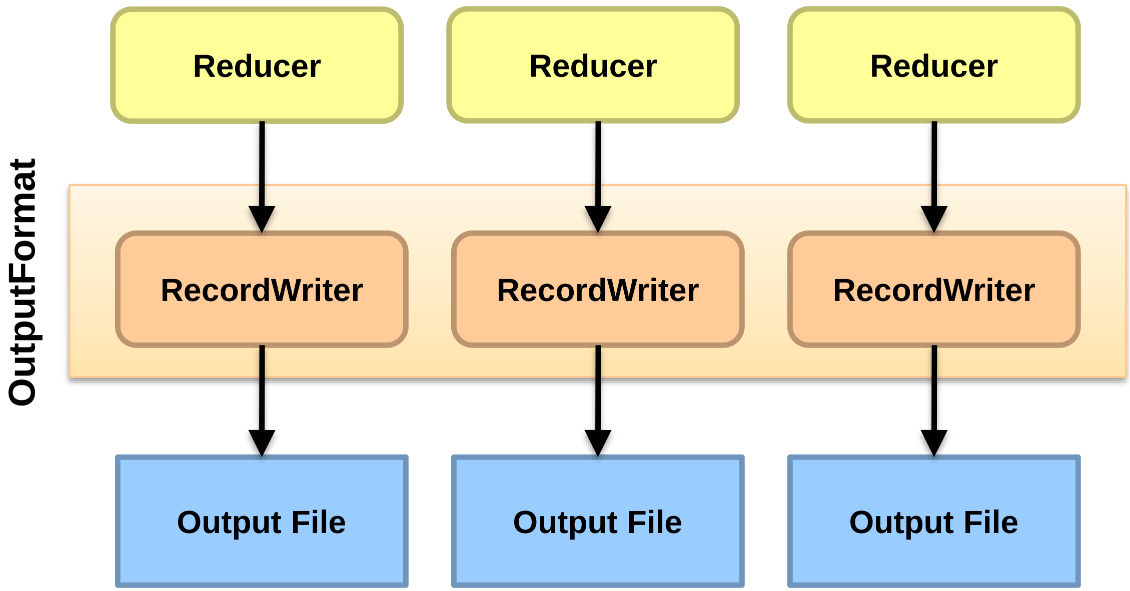
Source: redrawn from a slide by Cloduera, cc-licensed



Where’s the data actually coming from?



Source: redrawn from a slide by Cloduera, cc-licensed



Source: redrawn from a slide by Cloduera, cc-licensed

Input and Output

InputFormat

TextInputFormat

KeyValueTextInputFormat

SequenceFileInputFormat

…

OutputFormat

TextOutputFormat

SequenceFileOutputFormat

…

Spark also uses these abstractions for reading and writing data!

Hadoop Workflow

You

Submit node
(datasci)

Hadoop Cluster

Getting data in?
Writing code?
Getting data out?

Where's the actual
data stored?

Debugging Hadoop

First, take a deep breath
Start small, start locally
Build incrementally

Source: Wikipedia (The Scream)

Code Execution Environments

Different ways to run code:

Local (standalone) mode

Pseudo-distributed mode

Fully-distributed mode

Learn what's good for what

Hadoop Debugging Strategies

Good ol' System.out.println

Learn to use the webapp to access logs

Logging preferred over System.out.println

Be careful how much you log!

Fail on success

Throw RuntimeExceptions and capture state

Use Hadoop as the "glue"

Implement core functionality outside mappers and reducers

Independently test (e.g., unit testing)

Compose (tested) components in mappers and reducers

Source: Wikipedia (Japanese rock garden)