INTRO TO DATA SCIENCE MAP-REDUCE

I. BIG DATA
II. PROGRAMMING MODEL
III. IMPLEMENTATION DETAILS
IV. WORD COUNT EXAMPLE

EXERCISE:

V. MAP-REDUCE USING PYTHON

INTRO TO DATA SCIENCE

I. BIG DATA

Q: What does "big data" actually refer to?

Q: What does "big data" actually refer to?

A: Scalability; in particular, storing & processing web-scale (multi-terabyte) datasets...

Q: What does "big data" actually refer to?

A: Scalability; in particular, storing & processing web-scale (multi-terabyte) datasets...

But this is only half of the story...how would you do this?

One approach would be to get a huge supercomputer.

One approach would be to get a huge supercomputer.

But this has some obvious drawbacks:

- expensive
- difficult to maintain
- scalability is bounded

Instead of one huge machine, what if we got a bunch of regular (commodity) machines?

Instead of one huge machine, what if we got a bunch of regular (commodity) machines?

This has obvious benefits!

- cheaper
- easier to maintain
- scalability is unbounded (just add more nodes to the *cluster*)

Now we can give a complete answer to our earlier question.

Q: What does "big data" actually refer to?

Now we can give a complete answer to our earlier question.

Q: What does "big data" actually refer to?

A: Scalability; in particular, storing & processing web-scale (multi-terabyte) datasets using clusters of multiple computing nodes.

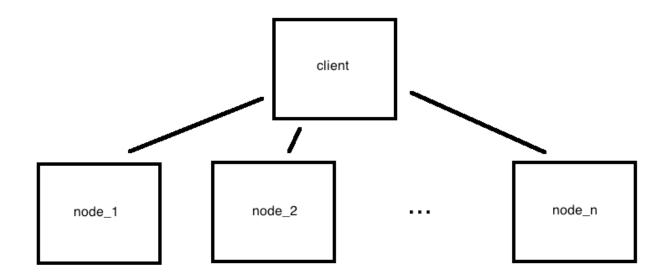
Now we can give a complete answer to our earlier question.

Q: What does "big data" actually refer to?

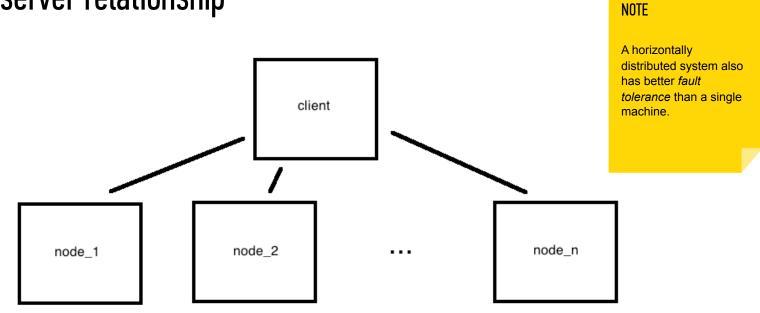
A: Scalability; in particular, storing & processing web-scale (multi-terabyte) datasets using clusters of multiple computing nodes.

"Scale out vs scale up!"

We can visualize this *horizontal* cluster architecture as a single client-multiple server relationship



We can visualize this *horizontal* cluster architecture as a single client-multiple server relationship



There are two ways to process data in a distributed architecture:

1) move data to cores (& code)

2) move code to data (& cores)

There are two ways to process data in a distributed architecture:

- 1) move data to cores (& code)
 - SETI@home

- 2) move code to data (& cores)
 - map-reduce → less overhead (network traffic, disk I/0)

"Computing nodes are the same as storage nodes."

Divide and conquer is a fundamental algorithmic technique for solving a given task, whose steps include:

Divide and conquer is a fundamental algorithmic technique for solving a given task, whose steps include:

- 1) split task into subtasks
- 2) solve these subtasks *independently*
- 3) recombine the subtask results into a final result

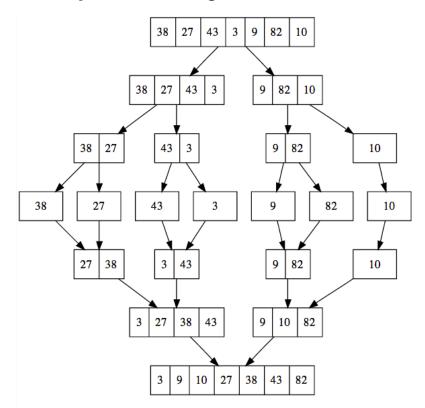
Divide and conquer is a fundamental algorithmic technique for solving a given task, whose steps include:

- 1) split task into subtasks
- 2) solve these subtasks *independently*
- 3) recombine the subtask results into a final result

This is how recursive algorithms work, for example.

One famous example of divide and conquer is *merge sort*.





Map-reduce leverages the divide and conquer approach by splitting a large dataset into several smaller datasets and performing a computation on each of these in parallel.

Map-reduce leverages the divide and conquer approach by splitting a large dataset into several smaller datasets and performing a computation on each of these in parallel.

In fact, running a map-reduce job with identity (eg, do-nothing) mappers and reducers is similar to merge sort!

Map-reduce leverages the divide and conquer approach by splitting a large dataset into several smaller datasets and performing a computation on each of these in parallel.

In fact, running a map-reduce job with identity (eg, do-nothing) mappers and reducers is similar to merge sort!

(The similarity is approximate, because results are output in multiple sets, and data is not broken down to single-element subsets.)

The defining characteristic of a problem that is suitable for the divide and conquer approach is that it can be broken down into *independent* subtasks.

The defining characteristic of a problem that is suitable for the divide and conquer approach is that it can be broken down into *independent* subtasks.

Tasks that can be *parallelized* in this way include:

- count, sum, average
- grep, sort, inverted index
- graph traversals, some ML algorithms

The defining characteristic of a problem that is suitable for the divide and conquer approach is that it can be broken down into *independent* subtasks.

Tasks that can be *parallelized* in this way include:

- count, sum, average
- grep, sort, inverted index
- graph traversals, some ML algorithms

NOTE

Parallelizing an ML algorithm can be a non-trivial exercise!

II. PROGRAMMING MODEL

As we've discussed, the map-reduce approach involves splitting a problem into subtasks and processing these subtasks in parallel.

As we've discussed, the map-reduce approach involves splitting a problem into subtasks and processing these subtasks in parallel.

This takes place in two phases:

As we've discussed, the map-reduce approach involves splitting a problem into subtasks and processing these subtasks in parallel.

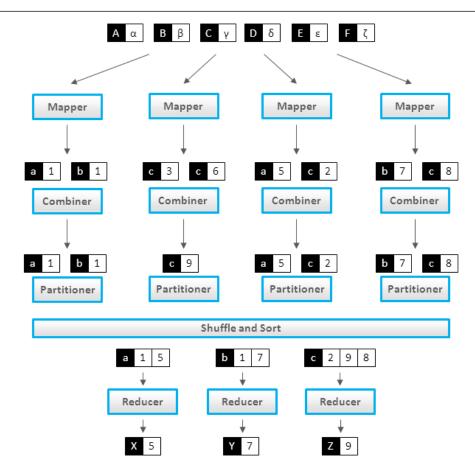
This takes place in two phases:

- 1) the **mapper** phase
- 2) the **reducer** phase

As we've discussed, the map-reduce approach involves splitting a problem into subtasks and processing these subtasks in parallel.

This takes place in (*approximately*) two phases:

- 1) the **mapper** phase
- 1.5) shuffle/sort
- 2) the **reducer** phase



Map-reduce uses a *functional programming* paradigm. The data processing *primitives* are mappers and reducers, as we've seen.

Map-reduce uses a *functional programming* paradigm. The data processing *primitives* are mappers and reducers, as we've seen.

mappers – filter & transform datareducers – aggregate results

MAP-REDUCE

Map-reduce uses a *functional programming* paradigm. The data processing *primitives* are mappers and reducers, as we've seen.

mappers – filter & transform datareducers – aggregate results

The functional paradigm is good at describing how to solve a problem, but not very good at describing data manipulations (eg, relational joins).

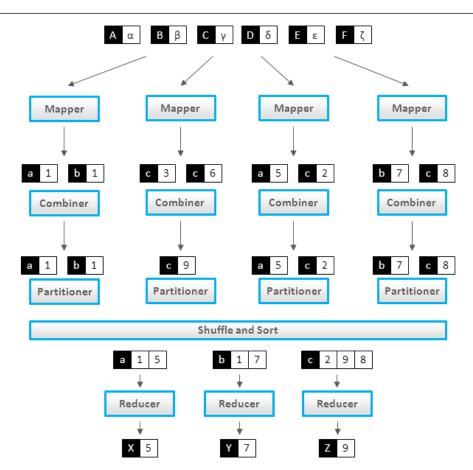
MAP-REDUCE

As our earlier diagram suggests, there are additional intermediate steps in a map-reduce workflow.

mappers – filter & transform datareducers – aggregate results

As our earlier diagram suggests, there are additional intermediate steps in a map-reduce workflow.

MAP-REDUCE



It's possible to overlay the map-reduce framework with an additional declarative syntax.

This makes operations like select & join easier to implement and less error prone.

Popular examples include Pig and Hive.

Why Pig?

Because I bet you can read the following script.

A Real Pig Script

```
000
                                                               top 5.pig
  users = load 'users.csv' as (username: chararray, age: int); -
  users 1825 = filter users by age >= 18 and age <= 25;
  pages = load 'pages.csv' as (username: chararray, url: chararray); -
  joined = join users 1825 by username, pages by username;
  grouped = group joined by url; -
  summed = foreach grouped generate group as url, COUNT(joined) AS views; -
  sorted = order summed by views desc;
  top 5 = limit sorted 5;
      e top_5 into 'top_5_sites.csv';
```

Now, just for fun... the same calculation in vanilla Hadoop MapReduce.

No, seriously.

```
Ja. werdowyna ferfilmer Fent. olaenii

Ja. werdowyna fent. olaenii

Ja. werdowyna fent. olaenii

Felt. olaenii
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    // Do the cross gradual and collect the values for disting al , first; {
    per collect ( min ) }
    per collect ( min ) ;
    per collection ( min ) ;
    per collection ( min ) ;
    per collection ( min ) ;
    reports; sendatural ( min ) ;
    reports; sendatural ( min ) ;

   port org.specke.tadoop.fs.Fsth.
port.org.specke.badoop.to.tonspr(table)
port org.specke.badoop.to.tasl
port org.specke.badoop.to.tasloop.
port org.specke.badoop.to.WittableComparable;
   sort org.openho.badong.magrod.fileingutformat.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  content bis " new debutef(TERRITARE) = 114 | healthch
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          public static class inadicinal saturds Maghaducedars implements Magnet-Test, Test, Test, LongStituble- i
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Jft.setHopper(lame:lame/lame/liter/bers.cleen)
FileInjer/brand.abd/sput/Arth/15s, now
lose/spite//defs*21)
FileInjer/brand.abd/spit/Arth/15s, now
lose/spite//defs*21)
FileDringer(brand)
lose Faith'/dess/spite/faith/literad_spite/*1);
Ift.setHopper(lame)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Test & Test to . The test of t
gort vra.apate.ha.hadong.omgared.dogumondfilminatus

port vra.apate.ha.hadong.omgared.dogumondfilminatus

port vra.apate.hadong.omgared.tettingstforest;

port vra.apate.hadong.omgared.jettingstforest;

port vra.apate.hadong.omgared.jebnostroi.phd.omgaret

port vra.apate.hadong.omgared.jebnostroi.phd.omgaret

port vra.apate.hadong.omgared.jebnostroi.yhdepager;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    Amounter reporters throws Discounting the Amounter of States and Control of States and C
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Nob loadfaces - new Nobelfuts
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       restricted to the community of memory (state);
point, restricted memory (s
                                                      poblic cold mappinespWritable k, Suss val.
                                                                                    gobilo static class Endorettle sutends HapReduceBase implements Reducer-Test, LongWoltsble, Writsble-Comparable, Writsble-4
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              FileInputFormat.mbHisputFath; buts, now
Fath; "/wow/gates/top/filtored_sect=");
FileOutputFormat.sectOutputFath; buts, new
Fath; "/www/gates/top/jollod";;
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       public veid reduce:

Yet a very concept trabler lier,

Organization and converted before a very trabler or,

Deportur reporter; blooms Differentiate (
// Add up all the velues or as
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               JobConf group = new JobConf(MRExample.01aem);
group.actJocRamm("Group URLs");
group.act(puthformat.fmy/mine/futhingutformat.dlamm);
group.act(http://dram.fmy/mine/futhingutformat.dlamm);
group.act(http://dram.fmy/mine/futhingutformat.dlamm);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        iong sum = 0;
while (iner.hashesn()) {
sum = n iner.hash() = sen();
reporter.estitation('US');
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               group, seriotypstvalourias (Longwirlashia (1888))
group, seriotypstropalarias (september) (seriotypolytomas, vlass))
group, seriotypstrolises (issahiniasi, siaas))
group, seriotypstrolises (issahiniasi, siaas))
group, seriotypstrolises (seriotypstrolise)
group, seriotypstrolises (seriotypstrolise)
group, seriotypstrolises (seriotypstrolise)
                                                      public wold maggiorgeritable b. Test val.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        or.collectibey, new LongWritable(sum));
                                                                                       tice would map (Longfor Lobble b. Took val.

(Augustian Service Service)

(7) Poil the bay one

(8) Poil the bay one

(8) Poil the bay one

(9) Poil the bay one

(1) Poil the bay one

(2) Poil the bay one

(3) Poil the bay one

(4) Poil the bay one

(5) Poil the bay one

(6) Poil the bay one

(7) Poil the bay one

(6) Poil the

(6) Poil the bay one

(7) Poil the bay one

(7) Poil the bay one

(8) Poil the bay one

                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Fath("/uses/gates/tep/joined");;
FilebatgutFormat, cetCotgutFathogroup, new
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          public static class LondClicks estando MagMediceMane
implements Mapper WritableComparable, Writable, toogWritable,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           Falk! '/weer/aptheast.astockgutPathigeroug
Falk! '/weer/apthea/tap/aptheastockgutPathigeroug
/who grouphes = new /obcgroup!
grouphes = see /obcgroup!
grouphes addreganding/bb(3010/061)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 public weld mage
WillableComparable Ray,
DisputibleComparable Ray,
DisputibleCommissionisticable, Vesti or,
Reporter reporter; throme Description (
or Tellonic(10)mpriciable(1), (Testi Ray))
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               Judionf Englis - now Johnstimmscapis slans;
englis .est.buckens; "Englis slans";;
englis .est.buckens; ("Englis englis sputformat.class);
englis .est.buckens; ("Englis englis slans);
englis .est.buckets; ("Englis englis .est.buckets);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       English one integral browned imagement is designated to intercept the tage of the control of the
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          public static class LimitClicks sounds HapdeduceBase implements RestoratelyLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLi
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    int such = 0;

public void orders;

its sour Test | its;

chaptical instance | its;

chaptical instance | instance | its |

Reported instance | instance | its |
                                                      public void reduce/Pest hay,

Iterator/Pest' iter,

Thepstolianter/Pest, Test' or,

Reported reported throws Domination (

// For each value, Figure out which file it's from and
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        // Only output the first 100 records while count = 100 km iter.teamset[]] ( oc-collect(key, iter.comt());
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           Authoritic je = new Authoritic find top 198 mitem for users

20 to 20 and Authoritic Temperary;

21 and Authoritic Temperary;

22 and Authoritic Temperary;

23 and Authoritic Temperary;

24 and Authoritic Temperary;
                                                                                                // accordingly.
Listostring: Siret = new ArreptietoString:[];
Listostring: second = new ArreptietoString:[];
                                                                                                while (ther hashest()) (
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             String value - t.tostring;;;
car, adding low, substituted by
```

II. IMPLEMENTATION DETAILS

IMPLEMENTATION DETAILS

The map-reduce framework handles a lot of messy details for you:

The map-reduce framework handles a lot of messy details for you:

- parallelization & distribution (eg, input splitting)
- partitioning (shuffle/sort/redirect)
- fault-tolerance (fact: tasks/nodes will fail!)
- I/O scheduling
- status and monitoring

IMPLEMENTATION DETAILS

The map-reduce framework handles a lot of messy details for you:

- parallelization & distribution (eg, input splitting)
- partitioning (shuffle/sort/redirect)
- fault-tolerance (fact: tasks/nodes will fail!)
- I/O scheduling
- status and monitoring

This (along with the functional semantics) allows you to focus on solving the problem instead of accounting & housekeeping details.

HADOOP

Hadoop is a popular open-source Java-based implementation of the map-reduce framework (including file storage for input/output).

Hadoop is a popular open-source Java-based implementation of the map-reduce framework (including file storage for input/output).

You can download Hadoop and configure a set of machines to operate as a map-reduce cluster, or you can run it as a *service* via Amazon's Elastic Map-Reduce.

Hadoop is a popular open-source Java-based implementation of the map-reduce framework (including file storage for input/output).

You can download Hadoop and configure a set of machines to operate as a map-reduce cluster, or you can run it as a *service* via Amazon's Elastic Map-Reduce.

Hadoop is written in Java, but the *Hadoop Streaming* utility allows client code to be supplied as executables (eg, written in any language).

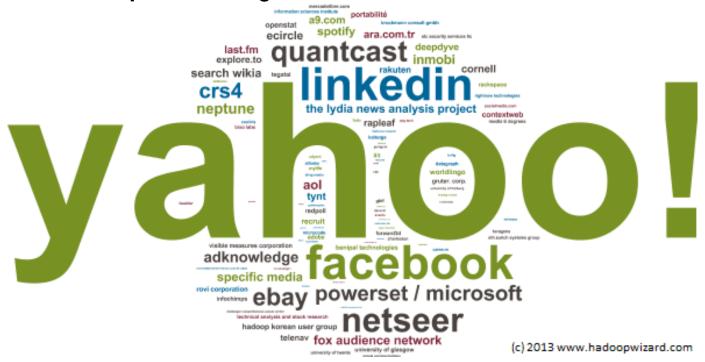
Frequently when people say "map-reduce" they're referring to Hadoop, but there are some exceptions:

HADOOP

Frequently when people say "map-reduce" they're referring to Hadoop, but there are some exceptions:

- many NoSQL databases support native map-reduce queries
- commercial distributions (Cloudera, MapR, etc)
- Google's internal implementation

That said, Hadoop has a large user base.



Data is replicated in the (distributed) file system across several nodes.

Data is replicated in the (distributed) file system across several nodes.

This permits locality optimization (and fault tolerance) by allowing the mapper tasks to run on the same nodes where the data resides.

Data is replicated in the (distributed) file system across several nodes.

This permits locality optimization (and fault tolerance) by allowing the mapper tasks to run on the same nodes where the data resides.

So we move code to data (instead of data to code), thus avoiding a lot of network traffic and disk I/O.

the same as storage

nodes."

Data is replicated in the (distributed) file system across se

This permits locality optimization (and fault tolerance) by a mapper tasks to run on the same nodes where the data re

So we move code to data (instead of data to code), thus avoiding a lot of network traffic and disk I/O.

The *Google File System* (GFS) was developed alongside map-reduce to serve as the native file system for this type of processing.

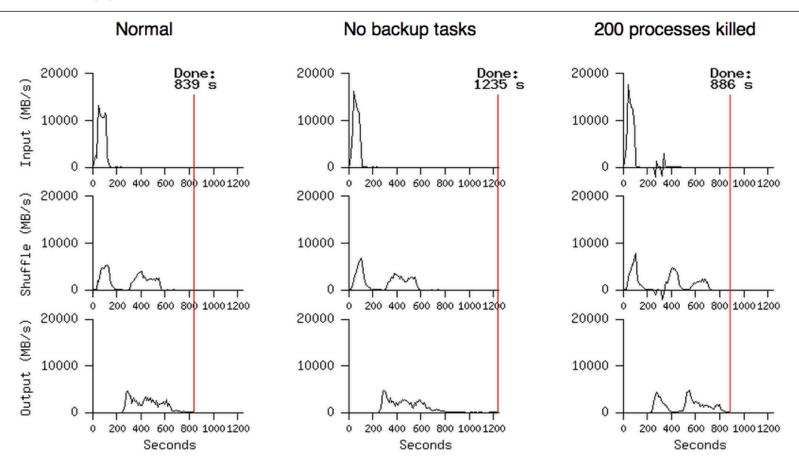
The *Google File System* (GFS) was developed alongside map-reduce to serve as the native file system for this type of processing.

The Hadoop platform is bundled with an open-source implementation of this file system called *HDFS*.

The *Google File System* (GFS) was developed alongside map-reduce to serve as the native file system for this type of processing.

The Hadoop platform is bundled with an open-source implementation of this file system called *HDFS*.

If you use Amazon EMR, you can use Amazon S3.



III. WORD COUNT EXAMPLE

Map-reduce processes data in terms of *key-value pairs*:

```
input <k1, v1> mapper <k1, v1> \rightarrow <k2, v2> (partitioner) <k2, v2> \rightarrow <k2, [all k2 values]> reducer <k2, [all k2 values]> \rightarrow <k3, v3>
```

MAP-REDUCE EXAMPLE

Using the following input, we can implement the "Hello World" of map-reduce: a word count.

MAP-REDUCE EXAMPLE: MAPPER INPUT

Using the following input, we can implement the "Hello World" of map-reduce: a word count.

```
where in where in the where in the world where in the world is where in the world is carmen where in the world is carmen sandiego
```

MAP-REDUCE EXAMPLE: MAPPER

The first processing primitive is the mapper, which filters & transforms the input data, and *emits* transformed key-value pairs.

The first processing primitive is the mapper, which filters & transforms the input data, and *emits* transformed key-value pairs.

```
mapper(k1, v1):
// k1 = line number
// v1 = line contents (eg, space-delimited string)

words = tokenize(v1) // split string into words
for word in words:
    emit (word, 1)
```

MAP-REDUCE EXAMPLE: MAPPER OUTPUT

The mapper emits key-value pairs for each word encountered in the input data.

MAP-REDUCE EXAMPLE: MAPPER OUTPUT

The mapper emits key-value pairs for each word encountered in the input data.

```
where 1 where 1 where 1 in 1 the 1
```

MAP-REDUCE EXAMPLE: PARTITIONER

The partitioner is internal to the map-reduce framework, so we don't have to write this ourselves. It shuffles & sorts the mapper output, and redirects all intermediate results for a given key to a *single* reducer.

MAP-REDUCE EXAMPLE: PARTITIONER OUTPUT

The partitioner is internal to the map-reduce framework, so we don't have to write this ourselves. It shuffles & sorts the mapper output, and redirects all intermediate results for a given key to a *single* reducer.

```
where
    [1, 1, 1, 1, 1, 1, 1]
in
    [1, 1, 1, 1, 1, 1]
the
    [1, 1, 1, 1, 1]
world
is
    [1, 1, 1, 1]
carmen
    [1, 1]
sandiego
[1]
```

MAP-REDUCE EXAMPLE: REDUCER

Finally, the reducer receives all values for a given key and aggregates (in this case, sums) the results.

MAP-REDUCE EXAMPLE: REDUCER

Finally, the reducer receives all values for a given key and aggregates (in this case, sums) the results.

```
reducer(k2, k2_vals):
// k2 = word
// k2_vals = word counts
   emit k2, sum(k2_vals)
```

MAP-REDUCE EXAMPLE: REDUCER OUTPUT

Reducer output is aggregated...

```
where
in 6
the 5
world 4
is 3
carmen 2
sandiego 1
```

MAP-REDUCE EXAMPLE: REDUCER OUTPUT

Reducer output is aggregated & sorted by key.

```
carmen 2
is 3
in 6
the 5
sandiego 1
where 7
world 4
```

INTRO TO DATA SCIENCE

IV. EXERCISE