INTRO TO DATA SCIENCE

LECTURE 18: HADOOP

- I. BIG DATA
- II. WHAT IS HADOOP?
- III. HDFS
- IV. MAP-REDUCE
- V. MACHINE LEARNING
- VI. MODERN HADOOP + THE FUTURE

INTRO TO DATA SCIENCE

I. BIG DATA

As you have probably heard, big data is a hot topic these days.

As you have probably heard, big data is a hot topic these days.

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

As you have probably heard, big data is a hot topic these days.

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

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

As you have probably heard, big data is a hot topic these days.

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

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.

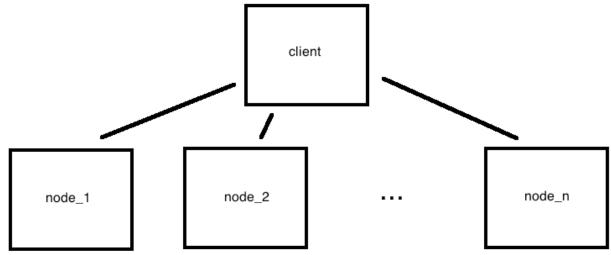
Now we can give a complete answer to our earlier question.

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.

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 se NOTE relationshin A horizontally distributed system also has better fault tolerance than a single client machine. node 2 node n node 1

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

1) move data to code (& processing power)

2) move code to data

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

- 1) move data to code (& processing power)
 - SETI

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

ALGORITHMS

Divide and conquer is a fundamental algorithmic

technique for solving a given task, whose steps include:

ALGORITHMS

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

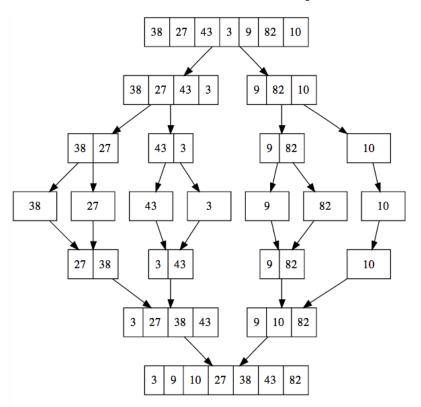
ALGORITHMS

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

Che famous example of divide and conquer is





ALGORITHMS

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!

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.

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

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. NOTE

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

- count, sum, average
- grep, sort, inverted index

Parallelizing an ML algorithm can be a nontrivial exercise!

IL PROGRAMMING MODEI

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.

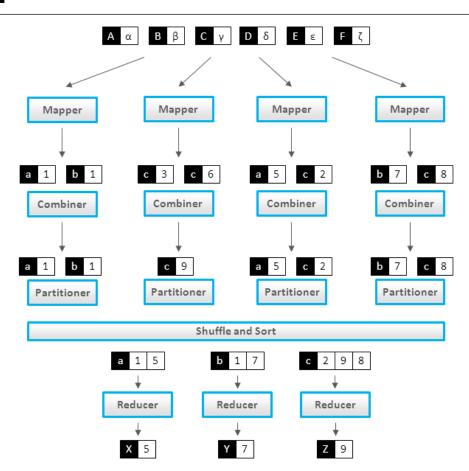
This takes place in two phases:

- 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 data reducers – aggregate results

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

mappers – filter & transform data reducers – aggregate results

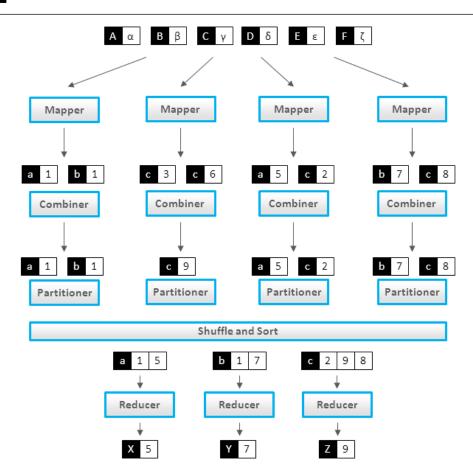
The functional paradigm is good at describing how to solve a problem, but not very good at

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

mappers – filter & transform data reducers – aggregate results

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

mappers – filter & transform data combiners – perform reducer operations on the mapper node (optional step, to reduce network traffic and disk I/O).



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.

```
// Do the cross gradual and collect the values for disting al , first; {
    per collect ( m) / ( m)
                              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.mogrod.fileingutformat.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     debtief his " new debtief(TEEmangle-stann);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               public static class inadicinal saturds Maghaducedars implements Magnet-Test, Test, Test, LongStituble- i
                      myerk or a spanik. Lading, magrad. Maghadoselana;
yari ora apanik. Lading, angrad. Maghadoselana;
yari ora apanik. Lading, angrad. Vanashila lading.
yari ora apanik. Lading, angrad. Maghadosel;
yari ora apanik. Lading, angrad.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  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 & Te
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       Amounter reporters throws Discognized a grant of the property of the control of t
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     Nob loadfaces - new Nobelfuts
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          restricted to the community of memory (state);
point, and holdens of just Charles of Fagure (state);
point, and holdens of just Charles of Fagure (state);
point, and holdens of memory of the state of leading (state);
point, and holdens of memory of the state of leading (state);
point, and holdens of memory of the state of leading (state);
point, and holdens of the state of leading (state);
point, and holdens of the state of leading (state);
point, and holdens of the state of leading (state);
point, and holdens of the state of leading (state);
point, and holdens of the state of leading (state);
point, and holdens of the state of leading (state);
point, and holdens of the state of leading (state);
point, and holdens of the state of leading (state);
point, and holdens of the state of leading (state);
point, and the state of leading (state);
point, 
                                                                                           poblic cold mappinespWritable k, Suss val.
                                                                                                                              gobio 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:

Yes a very congressable lier,

Organization and convertable comparable, writables or,

Reporter reporter) throws Discreption (

// 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, seriotypervalouras (Longertalas). (Lase);
group, seriotyperformat, depointed (Longertalas).
group, seriotyperformat, seriotyperformat, vilase);
group, seriotyperformat, seriotyperformat, seriotyperformat, vilase);
group, seriotyperformat, 
                                                                                           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

(4) Poil the bay one

(5) Poil the bay one

(6) Poil the bay one

(7) Poil the bay one

(8) Poil the bay one

(8) Poil the bay one

(9) Poil the bay one

(1) Poil the

(1) Poil the bay one

(1) 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

                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  Fath("/uses/gates/tep/joined");;
FilebatgutFormat, cetCotgutFathogroup, new
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               public static class LondClicks estando MagMediceMane
implements Mapper WritableComparable, Writable, toogWritable,
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Falls '/neer/apth/sount antocomput/actocycong
falls '/neer/apth/sounded'());
proup estNominadocetanks(b9);
/nh grouphon = nee /nh(group);
grouphon = neer/apth/sound();
grouphon = neer/apth/sound();
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             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 sians;
englis est. Document "Englis sians";
englis est input format (Regulatory | Lating | Provide to Class);
englis est. Sept format (Regulatory | Lating | Provide to Class);
englis est. Sept format (Regulatory | Sept | Lating | Provide to Class);
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          English one integral browned imagement is designated to intercept the tage of the property of the tage of the property of the 
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               public static class LimitClicks sotands HapdeduceBase implements RestoratelyLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryLindburgeryL
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                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());
store 11.
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              Authoritic je = new Authoritic (*find top 198 mitem for users 18 to 25°) and Authoritication (*find top 198 mitem for users 18 to 25°) and Authoritication (*j. 18 to 25°) and
                                                                                                                                             // accordingly.
Listostring: Siret = new ArreptietoStrings::;;
Listostring: second = new ArreptietoStrings::;;
                                                                                                                                             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

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

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.

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

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

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

HADOOP

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

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

"Compute nodes are the same as storage nodes."

This permits locality optimization (and fatolerance) 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

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

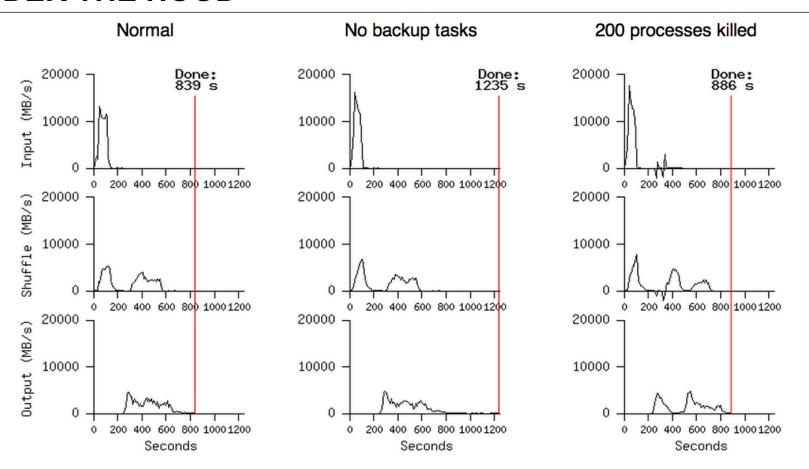
The Google File System (GFS) was developed alongside map-reduce to serve as the native file system for this type of processing.

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

If you use Amazon EMR, you can use their file



III. WORD COUNT EYAMDIE

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

```
input <k1, v1> mapper <k1, v1> \rightarrow <k2, v2> <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.

MAP-REDUCE EXAMPLE: MAPPER

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 in 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

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	[1,	1,	1,	1]			
is	[1,	1	,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...
```

75

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

```
MAP-REDUCE EXAMPLE: REDUCER
```

76

OUTPUT

Reducer output is aggregated & sorted by key.

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