Apache Hadoop

Large scale data processing

Speaker: Isabel Drost

Isabel Drost

Nighttime:

Came to nutch in 2004.
Co-Founder Apache Mahout.
Organizer of Berlin Hadoop Get Together.

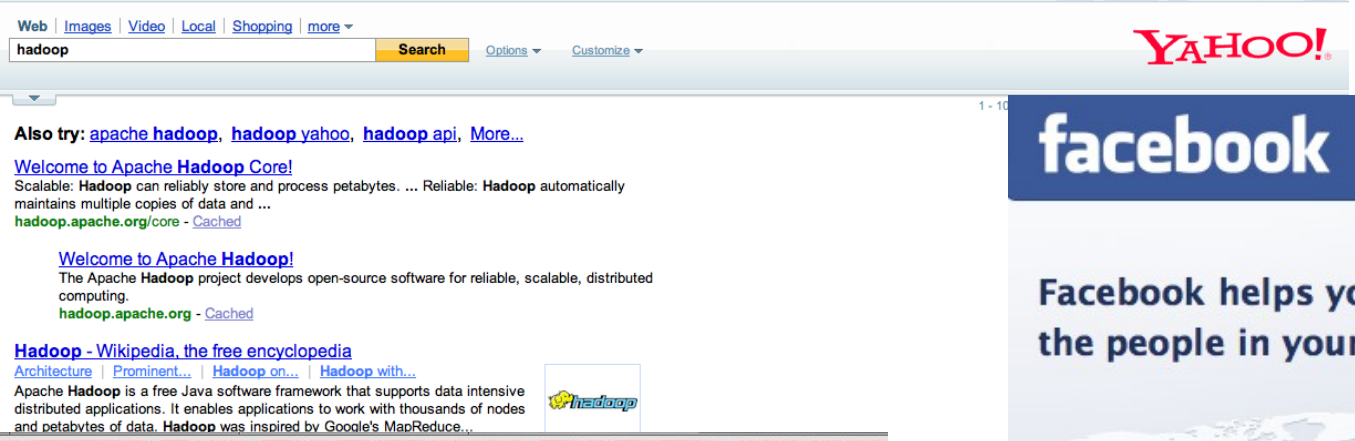
Daytime:

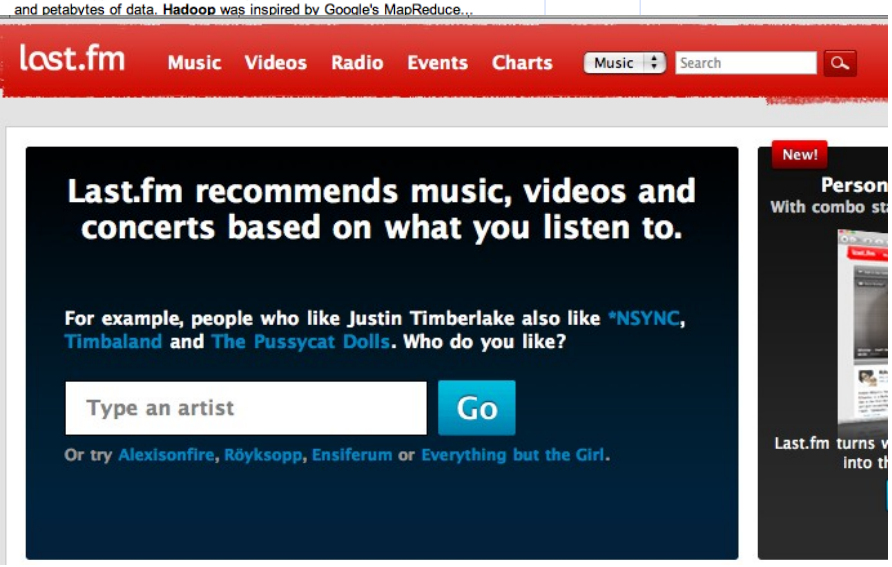
Software developer @ Berlin

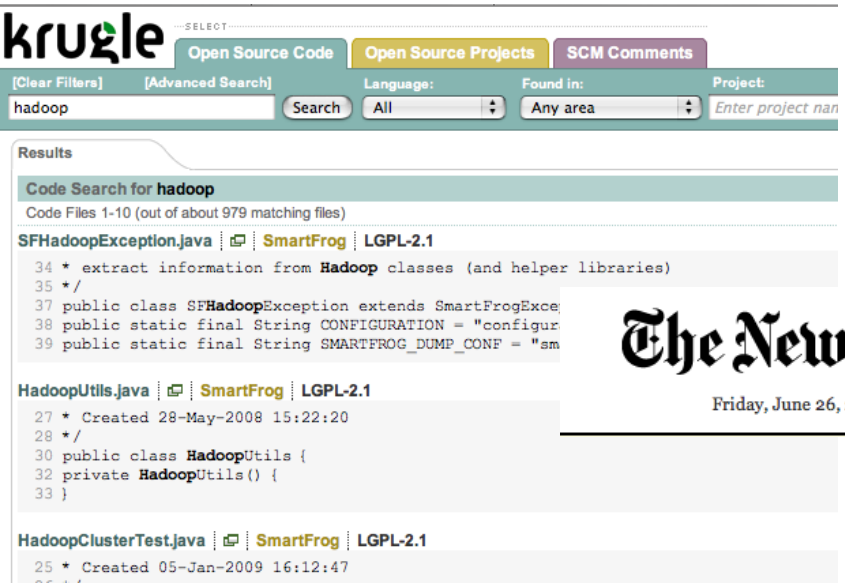
Hello Information Retrieval course!

Agenda

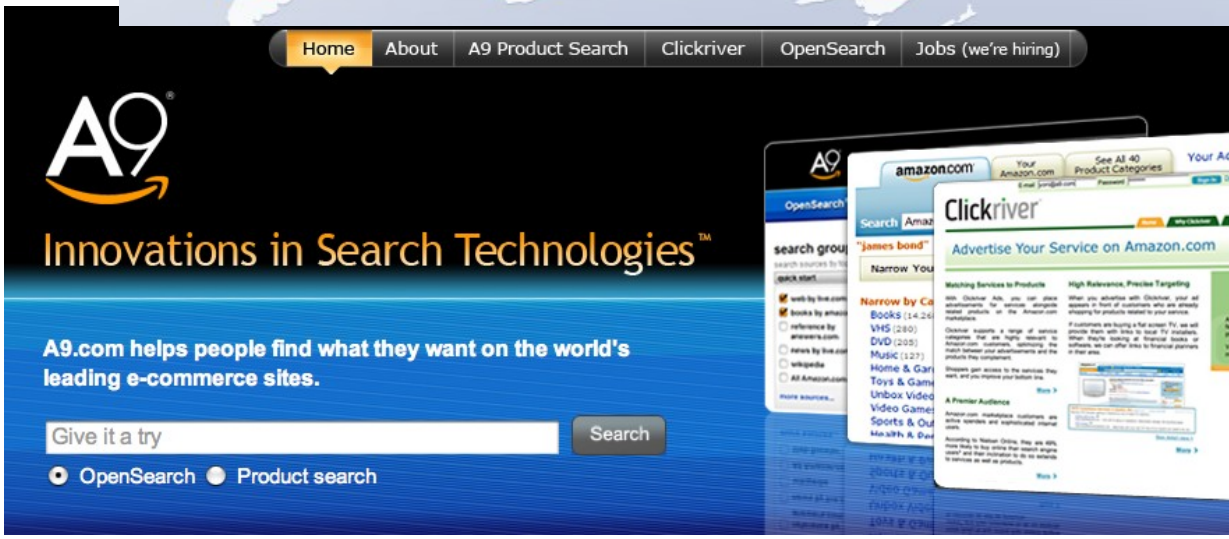
- Motivation.
- A short tour of Map Reduce.
- Introduction to Hadoop.
- Hadoop ecosystem.


Web | Images | Video | Local | Shopping | more
hadoop
Search
Options
Customize
YAHOO!
Also try: apache hadoop, hadoop yahoo, hadoop api, More...
Welcome to Apache Hadoop Core!
Scalable: Hadoop can reliably store and process petabytes. ... Reliable: Hadoop automatically maintains multiple copies of data and ...
hadoop.apache.org/core - Cached
Welcome to Apache Hadoop!
The Apache Hadoop project develops open-source software for reliable, scalable, distributed computing.
hadoop.apache.org - Cached
Hadoop - Wikipedia, the free encyclopedia
Architecture | Prominent... | Hadoop on... | Hadoop with...
Apache Hadoop is a free Java software framework that supports data intensive distributed applications. It enables applications to work with thousands of nodes and petabytes of data. Hadoop was inspired by Google's MapReduce...

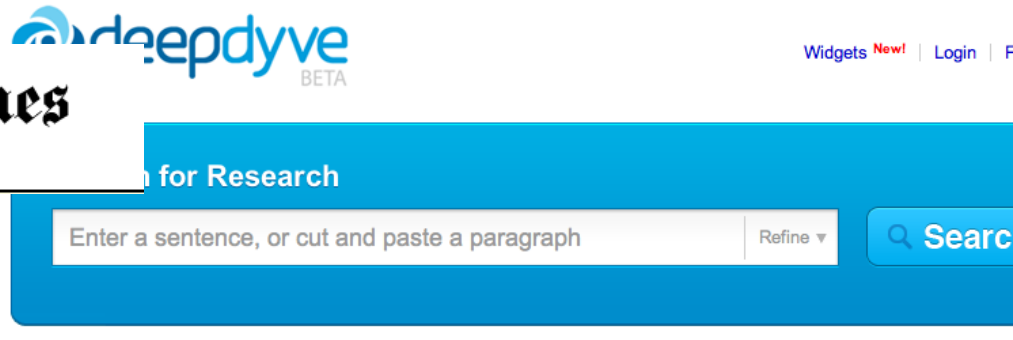

last.fm
Music
Videos
Radio
Events
Charts
Music
Search
Last.fm recommends music, videos and concerts based on what you listen to.
For example, people who like Justin Timberlake also like *NSYNC, Timbaland and The Pussycat Dolls. Who do you like?
Type an artist
Go
Or try Alexisonfire, Röyksopp, Ensiferum or Everything but the Girl.
New!
Person
With combo st
Last.fm turns
into th


krugle
SELECT
Open Source Code
Open Source Projects
SCM Comments
[Clear Filters]
[Advanced Search]
Language:
Found in:
Project:
hadoop
Search
All
Any area
Enter project nam
Results
Code Search for hadoop
Code Files 1-10 (out of about 979 matching files)
SFHadoopException.java | SmartFrog | LGPL-2.1
34 * extract information from Hadoop classes (and helper libraries)
35 */
37 public class SFHadoopException extends SmartFrogExce
38 public static final String CONFIGURATION = "configur
39 public static final String SMARTFROG_DUMP_CONF = "sm
HadoopUtils.java | SmartFrog | LGPL-2.1
27 * Created 28-May-2008 15:22:20
28 */
30 public class HadoopUtils {
32 private HadoopUtils() {
33 }
HadoopClusterTest.java | SmartFrog | LGPL-2.1
25 * Created 05-Jan-2009 16:12:47


facebook
Email
Facebook helps you connect and share with the people in your life.
Sign Up
It's free a
Full
Your
New Pas

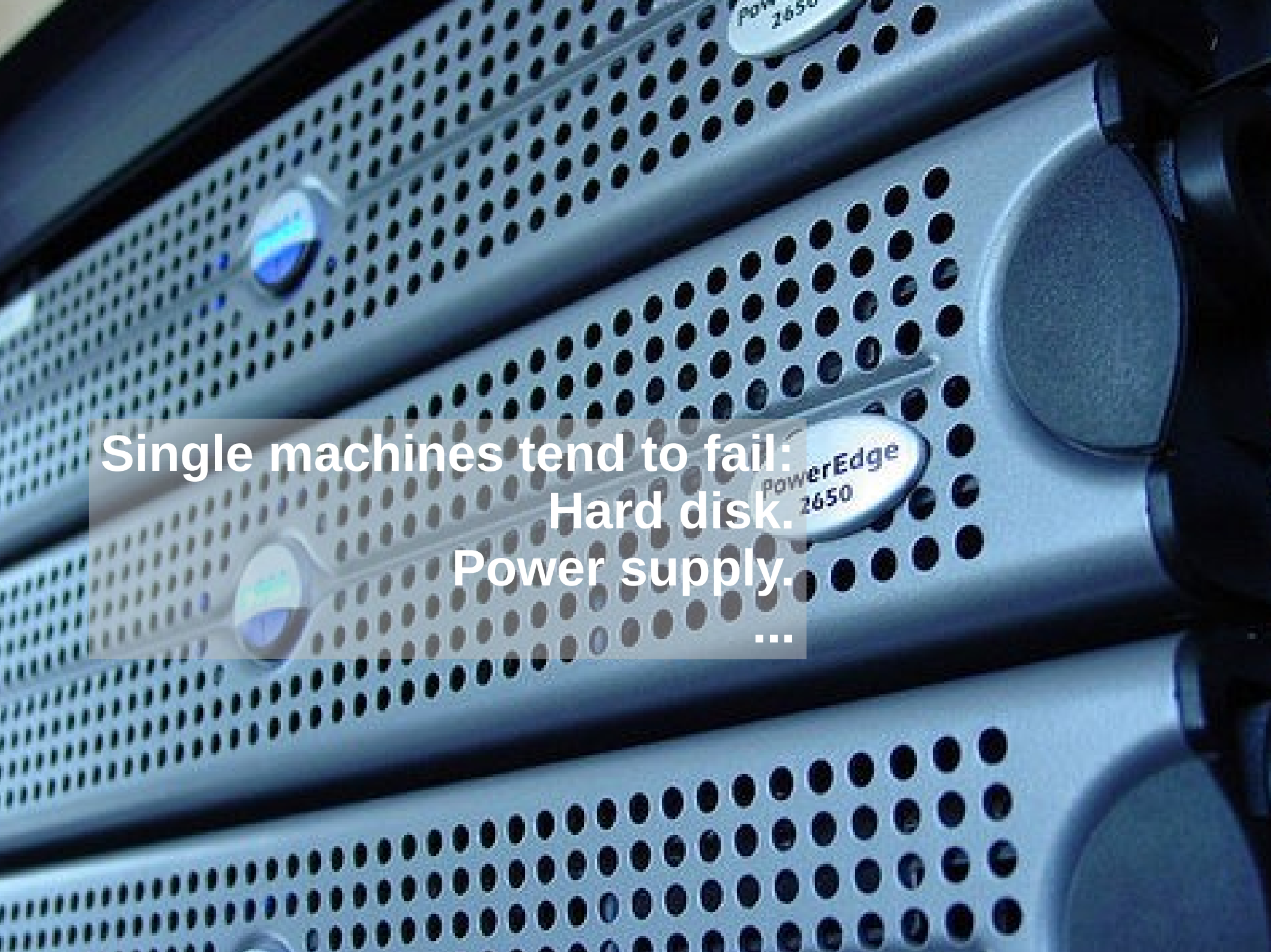

Home
About
A9 Product Search
Clickriver
OpenSearch
Jobs (we're hiring)
A9
Innovations in Search Technologies™
A9.com helps people find what they want on the world's leading e-commerce sites.
Give it a try
Search
OpenSearch
Product search
Clickriver
Advertise Your Service on Amazon.com


The New York Times
Friday, June 26, 2009
Last Update: 7:17 AM ET


deepdyve
BETA
Widgets New!
Login
for Research
Enter a sentence, or cut and paste a paragraph
Refine
Searc


January 8, 2008 by Pink Sherbet Photography
http://www.flickr.com/photos/pinksherbet/2177961471/
Massive data as in:
Cannot be stored on single machine.
Takes too long to process in serial.
Idea: Use multiple machines.

Challenges.



Single machines tend to fail:

Hard disk.

Power supply.

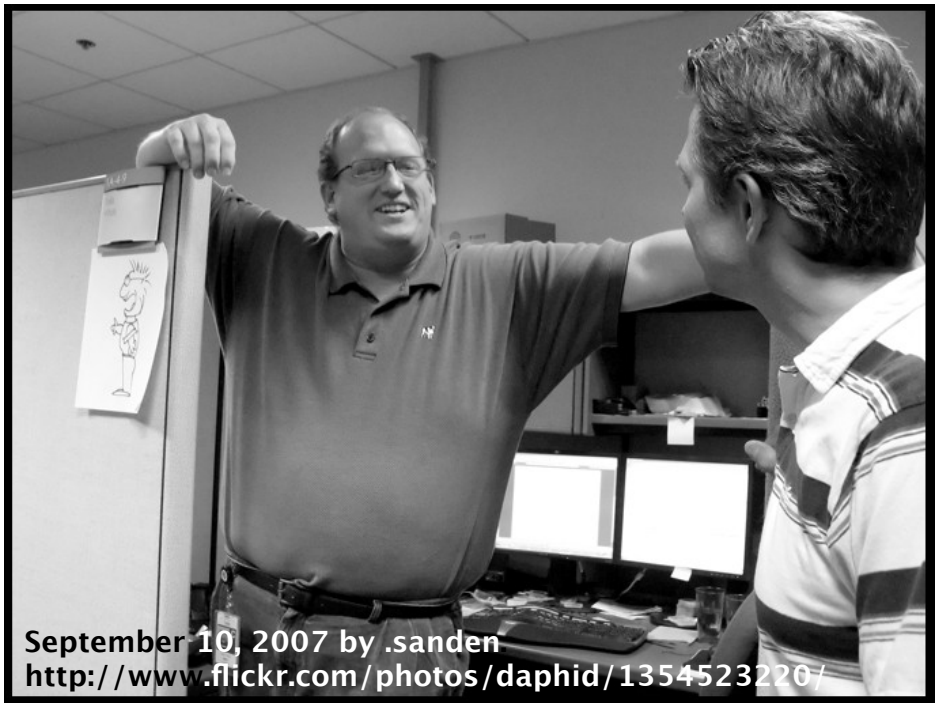
...


More machines – increased failure probability.
January 11, 2007, skreuzer
http://www.flickr.com/photos/skreuzer/354316053/

Requirements

- Built-in backup.
- Built-in failover.

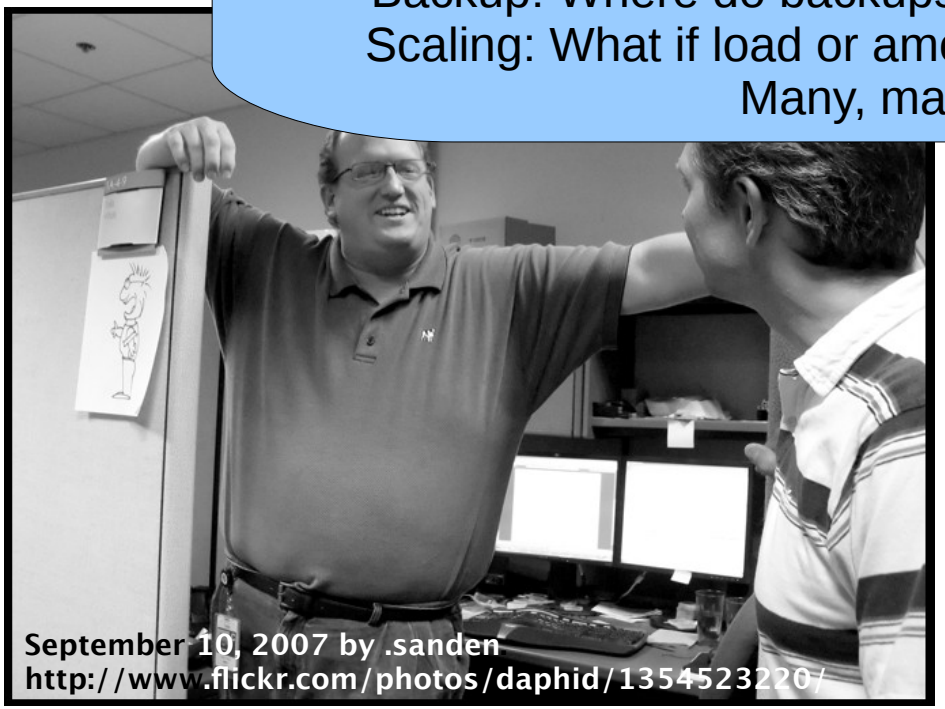
Typical developer



- Has never dealt with large (petabytes) amount of data.
- Has no thorough understanding of parallel programming.
- Has no time to make software production ready.

Typical developer

Failure resistant: What if service X is unavailable?
Failover built in: Hardware failure does happen.
Documented logging: Understand message w/o code.
Monitoring: Which parameters indicate system's health?
Automated deployment: How long to bring up machines?
Backup: Where do backups go to, how to do restore?
Scaling: What if load or amount of data double, triple?
Many, many more.



- Has no time to make software production ready.

Requirements

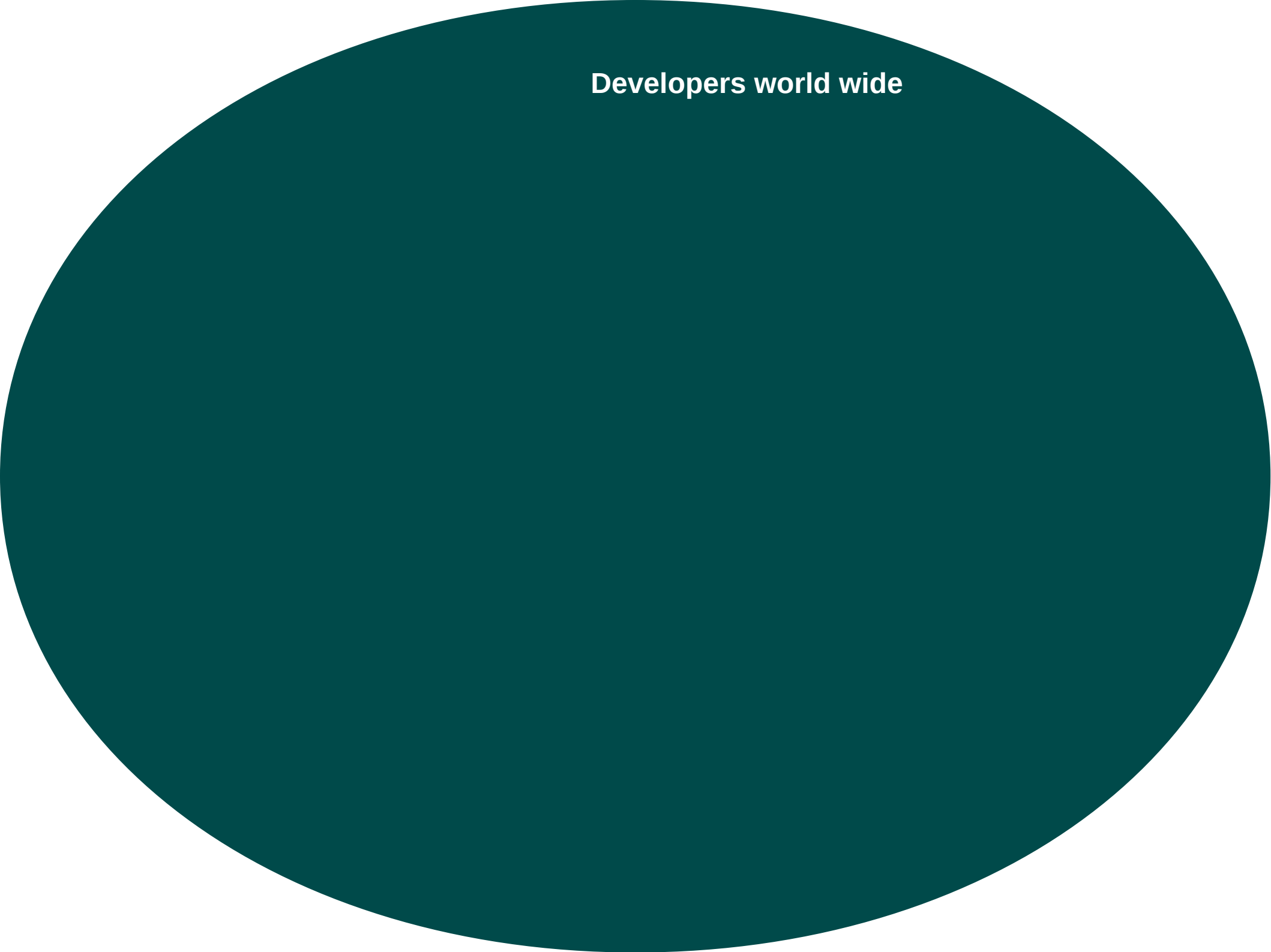
- Built-in backup.
- Built-in failover.
- Easy to use.
- Parallel on rails.

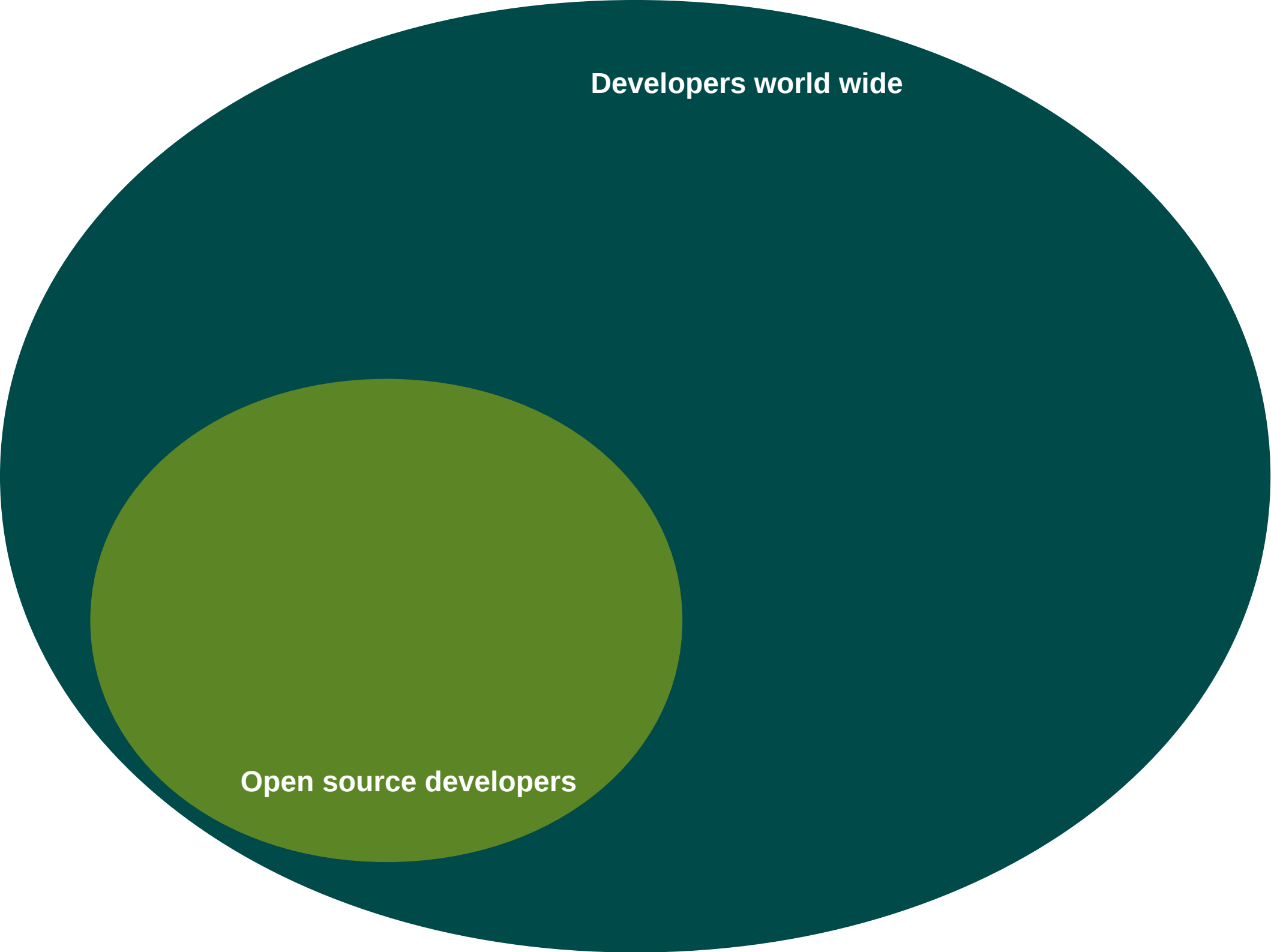
February 29, 2008 by Thomas Claveirole
http://www.flickr.com/photos/thomasclaveirole/2300932656/

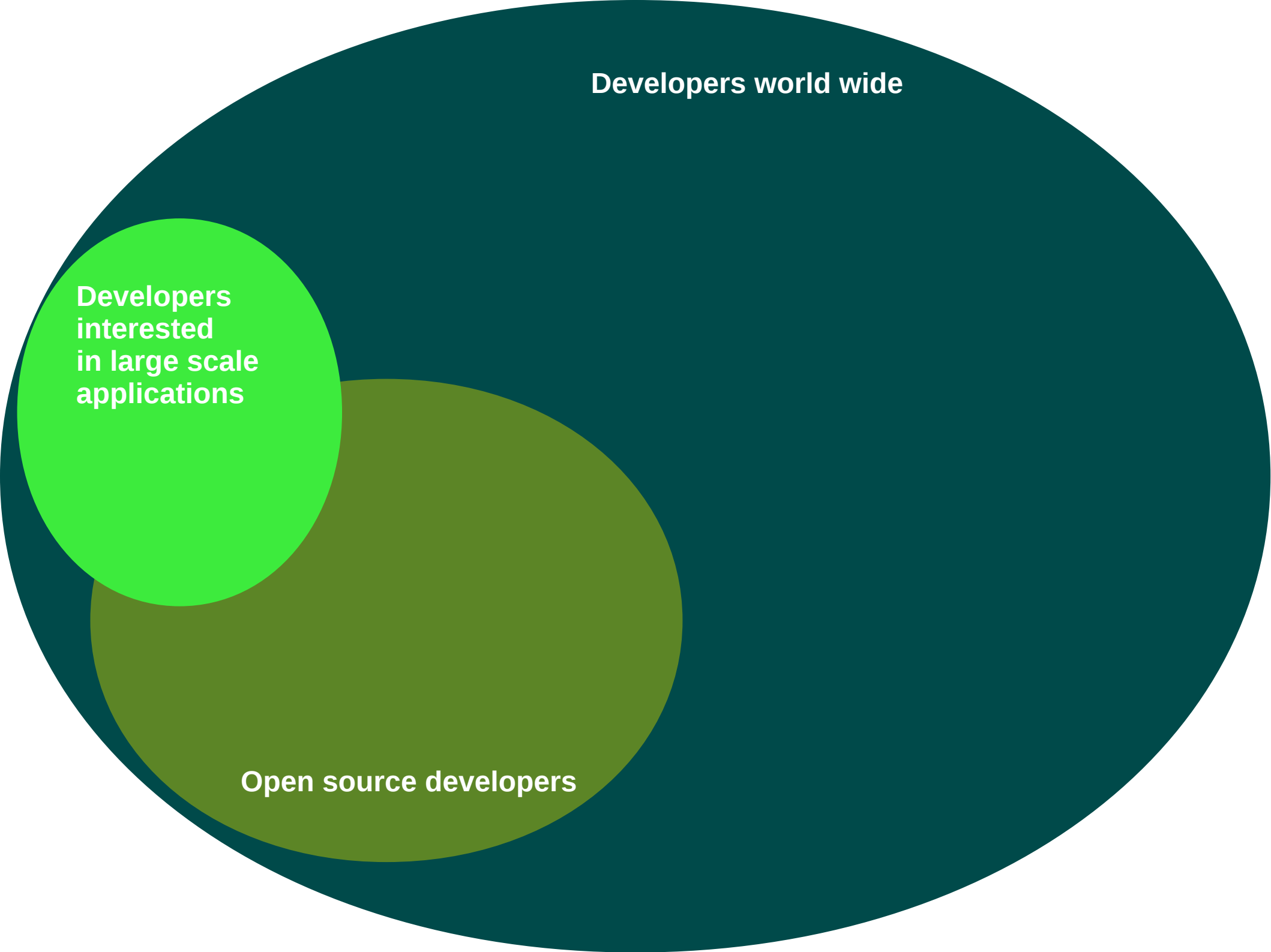
http://www.flickr.com/photos/jaaronfarr/3384940437/
March 25, 2009 by jaaron

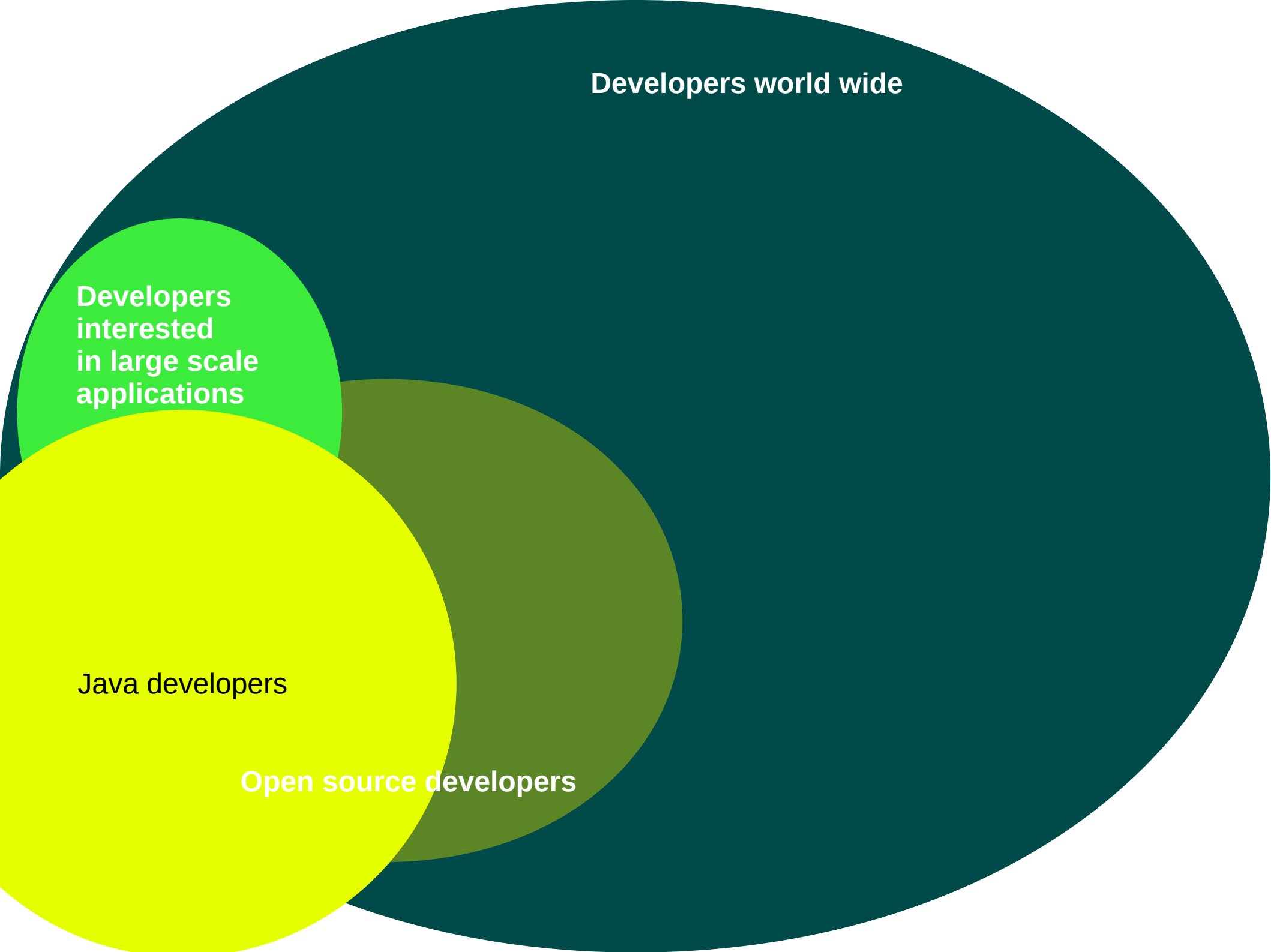
http://www.flickr.com/photos
/jaaronfarr/3385756482/
March 25, 2009 by jaaron

May 1, 2007 by danny angus
http://www.flickr.com/photos/killerbees/479864437/


Developers world wide

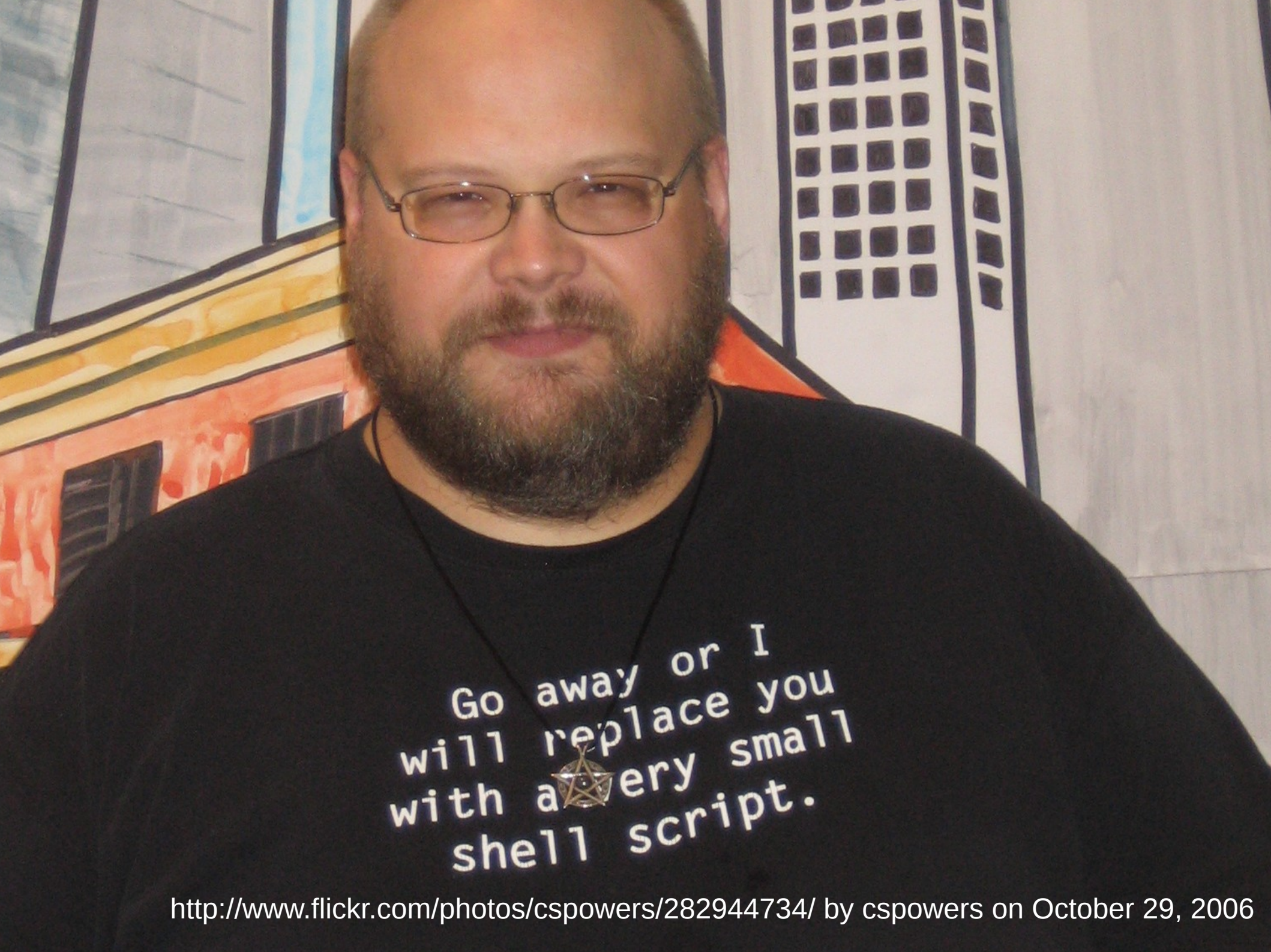

Developers world wide
Open source developers


Developers world wide
Developers interested in large scale applications
Open source developers


Developers world wide
Developers interested in large scale applications
Java developers
Open source developers

Requirements

- Built-in backup.
- Built-in failover.
- Easy to use.
- Parallel on rails.
- Java based.



http://www.flickr.com/photos/cspowers/282944734/ by cspowers on October 29, 2006

Requirements

- Built-in backup.
- Built-in failover.
- Easy to use.
- Parallel on rails.
- Easy to administrate.
- Single system.
- Java based.

We need a solution that:

Is easy to use.

Scales well beyond one node.

Java based implementation.

Easy distributed programming.

Well known in industry and research.

Scales well beyond 1000 nodes.

- 2008:
  - 70 hours runtime
  - 300 TB shuffling
  - 200 TB output
- In 2009
  - 73 hours
  - 490 TB shuffling
  - 280 TB output
  - 55%+ hardware
  - 2k CPUs (40% faster cpus)

- 2008
  - 2000 nodes
  - 6 PB raw disk
  - 16 TB RAM
  - 16k CPUs
- In 2009
  - 4000 nodes
  - 16 PB disk
  - 64 TB RAM
  - 32k CPUs (40% faster cpus)

Example use cases

- Distributed Grep.
- Distributed Sort.
- Link-graph traversal.
- Term-Vector per host.
- Web access log stats.
- Inverted index.
- Doc clustering.
- Machine learning.
- Machine translation.

Some history.

Feb '03 first Map Reduce library @ Google

Oct '03 GFS Paper

Dec '04 Map Reduce paper

Dec '05 Doug reports that nutch uses map reduce

Feb '06 Hadoop moves out of nutch

Apr '07 Y! running Hadoop on 1000 node cluster

Jan '08 Hadoop made an Apache Top Level Project

Hadoop assumptions

Assumptions:

Data to process does not fit on one node.
Each node is commodity hardware.
Failure happens.

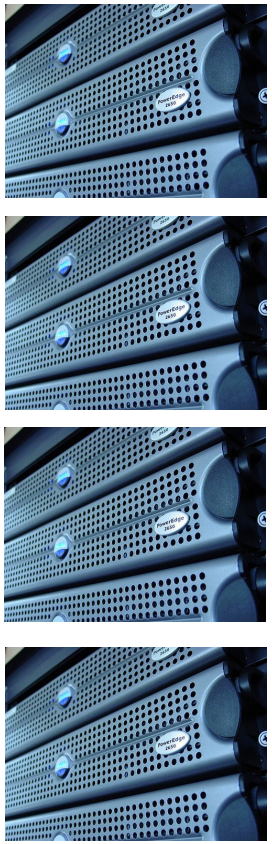



Ideas:

Distribute filesystem.
Built in replication.
Automatic failover in case of failure.

Assumptions:

Moving data is expensive.
Moving computation is cheap.
Distributed computation is easy.

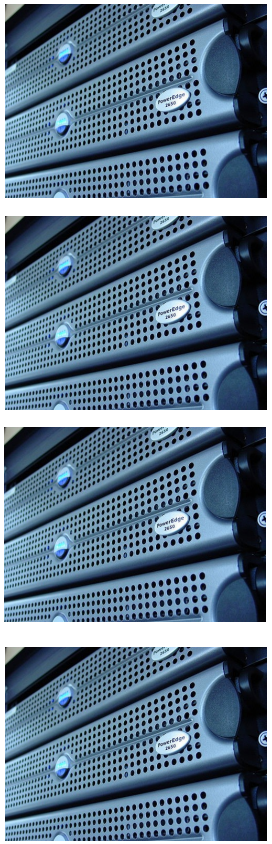


Ideas:

Move computation to data.
Write software that is easy to distribute.

Assumptions:

Systems run on spinning hard disks.
Disk seek >> disk scan.



Ideas:

Improve support for large files.
File system API makes scanning easy.

Hadoop by example

```
<?xml version="1.0" encoding="UTF-8"?>
<opml version="1.0" >
 <head>
  <text></text>
 </head>
 <body>
  <outline htmlUrl="http://eventseer.net" title="EventSeer - A Digital Library of Call for Papers" useC
alDefault" version="RSS" type="rss" xmlUrl="http://eventseer.net/feeds/main/rss.xml" id="312053548" tex
tseer.net" />
  <outline isOpen="false" id="669809145" text="Silent" >
   <outline htmlUrl="http://www.theserverside.com" title="TheServerSide.com: Patterns" useCustomFetchIn
ersion="RSS" type="rss" xmlUrl="http://www.theserverside.com/rss/theserverside-j2eepatterns-rss2.xml" i
taining up-to-date news, discussions, patterns, resources, and media" />
   <outline htmlUrl="http://chadwa.wordpress.com" title="Chad's Search Blog" useCustomFetchInterval="fa
S" type="rss" xmlUrl="http://chadwa.wordpress.com/feed/" id="545368194" text="Chad's Search Blog" descr
" />
   <outline htmlUrl="http://www.find23.net/Site/Blog/Blog.html" title="My Blog" useCustomFetchInterval=
"RSS" type="rss" xmlUrl="http://www.find23.net/Site/Blog/rss.xml" id="1620106192" text="My Blog" descri
   <outline htmlUrl="http://emotion.inrialpes.fr/~dangauthier/blog" title="Yet Another Machine Learning
eMode="globalDefault" version="RSS" type="rss" xmlUrl="http://emotion.inrialpes.fr/~dangauthier/blog/fe
g" />
   <outline htmlUrl="http://ml.typepad.com/machine_learning_thoughts/" title="Machine Learning Thoughts
="globalDefault" version="RSS" type="rss" xmlUrl="http://ml.typepad.com/machine_learning_thoughts/rss.x
etical and practical aspects of Machine Learning." />
   <outline htmlUrl="http://yaroslavvb.blogspot.com/" title="Machine Learning, etc" useCustomFetchInter
ion="RSS" type="rss" xmlUrl="http://yaroslavvb.blogspot.com/feeds/posts/default" id="805998569" text="M
   <outline htmlUrl="http://ptufts.blogspot.com/" title="Pinhead's Progress" useCustomFetchInterval="fa
S" type="rss" xmlUrl="http://ptufts.blogspot.com/feeds/posts/default" id="1019393988" text="Pinhead's P
   <outline htmlUrl="http://resnotebook.blogspot.com/" title="Misc Research Stuff" useCustomFetchInterv
on="RSS" type="rss" xmlUrl="http://resnotebook.blogspot.com/feeds/posts/default" id="216193226" text="M
   <outline htmlUrl="http://absolutely-regular.blogspot.com/" title="Absolutely Regular" useCustomFetch
 version="RSS" type="rss" xmlUrl="http://absolutely-regular.blogspot.com/feeds/posts/default" id="17850
   <outline htmlUrl="http://atomai.blogspot.com/" title="Data Mining, Analytics and Artificial Intelli
Mode="globalDefault" version="RSS" type="rss" xmlUrl="http://atomai.blogspot.com/feeds/posts/default" i
nt in data mining, artificial intelligence, analytics, intelligent agents, semiconductors, distributing
siness Objects, Oracle, Intel, AMD, or Pentaho. Heuristic, Six Sigma, or CMM. Contractor or in-house. H
ail.com" />
```

```
isabel@h1349259:~$ more data/feeds.opml | grep -o "http://[0-9A-Za-z\-_\.]*" | s
ort | uniq --count | sort | tail
      3 http://agbs.kyb.tuebingen.mpg.de
      3 http://irgupf.com
      3 http://jeffsutherland.com
      4 http://ml.typepad.com
      4 http://weblogs.java.net
      4 http://www.gridvm.org
      4 http://yaroslavvb.blogspot.com
      5 http://feeds.feedburner.com
      6 http://blogsearch.google.com
     10 http://arxiv.org
```

```
pattern=”http://[0-9A-Za-z\-_\.]*”

grep -o "$pattern" feeds.opml      | sort            | uniq --count
```

pattern=”http://[0-9A-Za-z\-_\.]*”

```
grep -o "$pattern" feeds.opml
M  A   P
```

```
| sort
| SHUFFLE
```

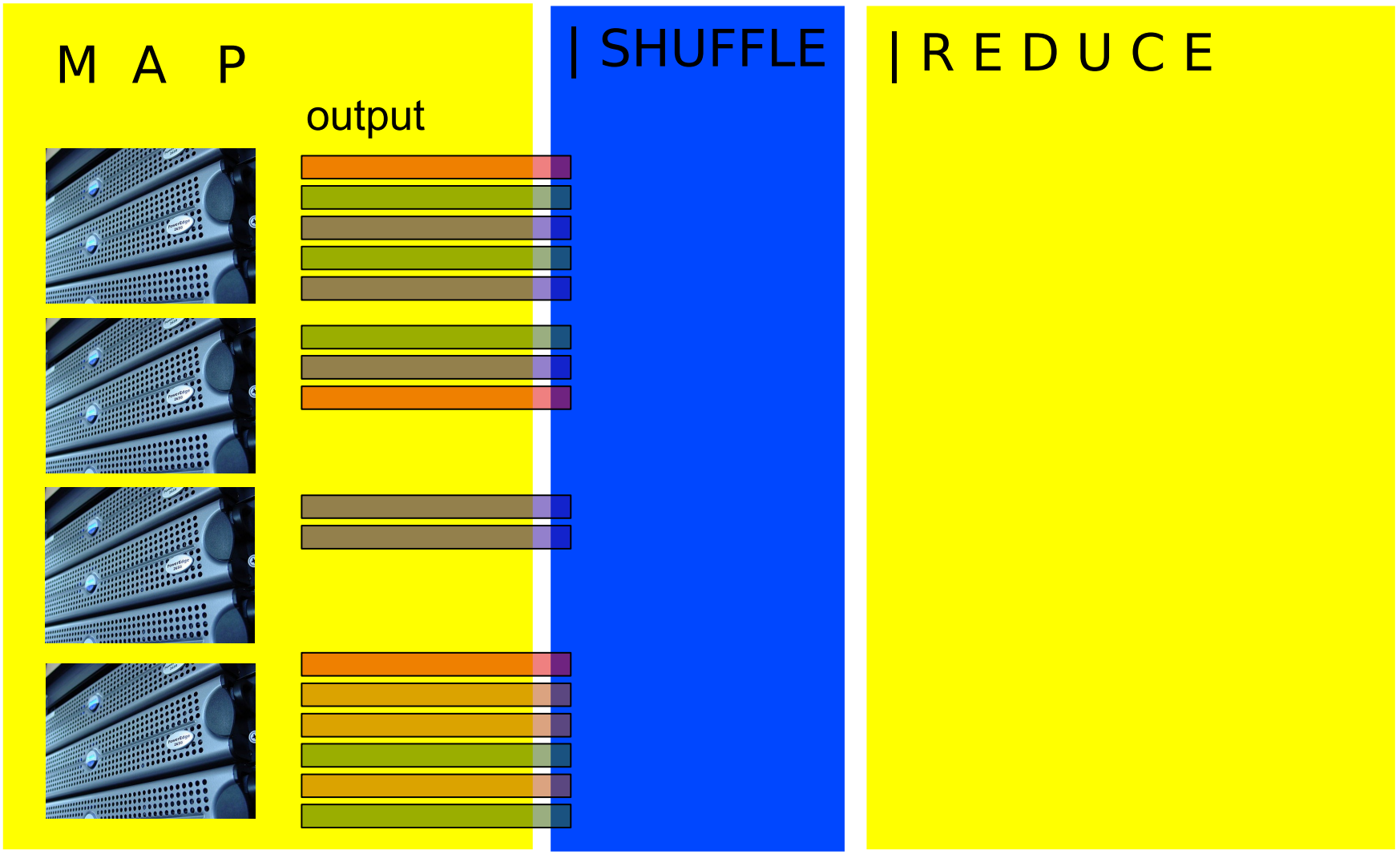
```
| uniq --count
| R E D U C E
```

M A P

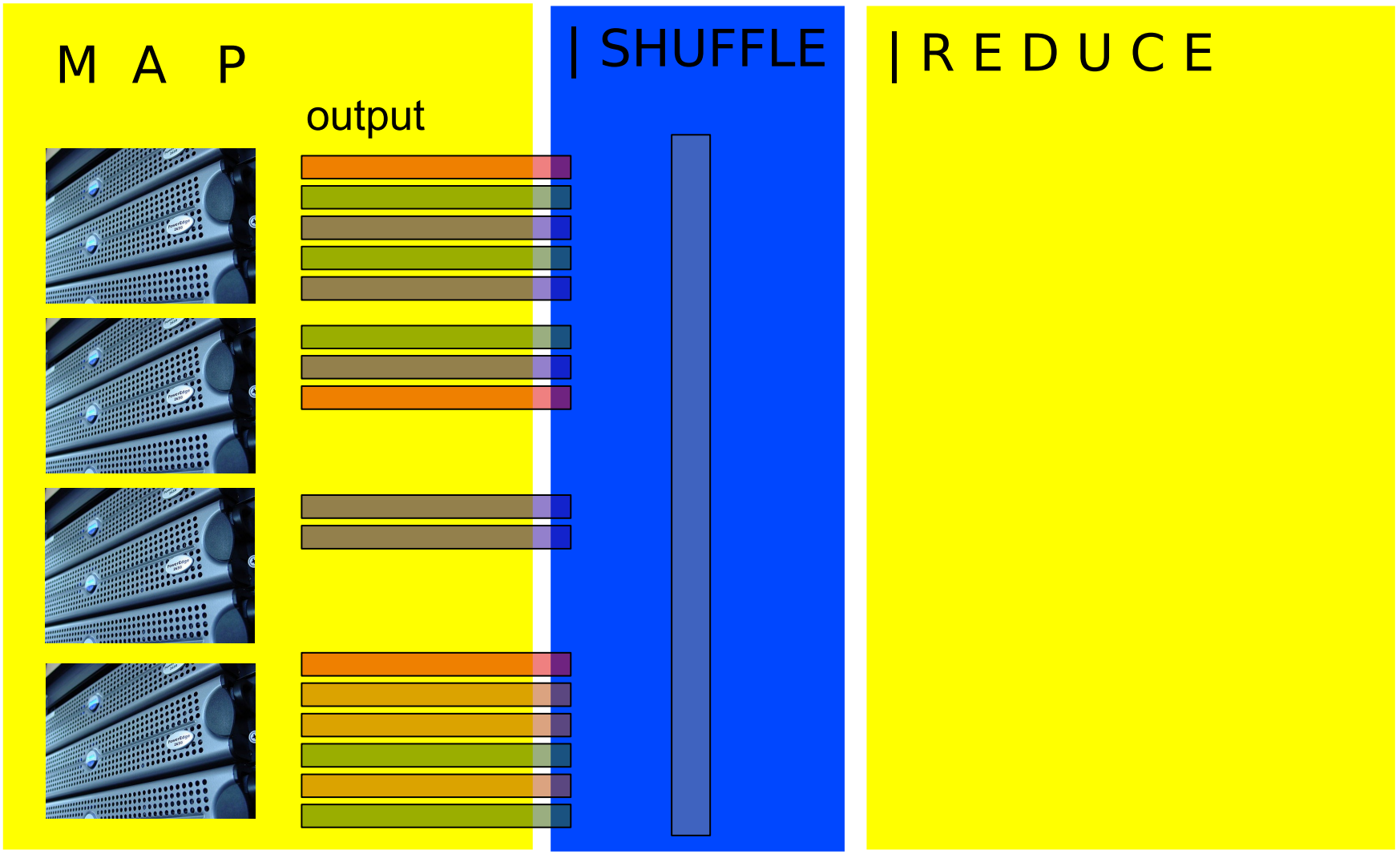
| SHUFFLE

| R E D U C E

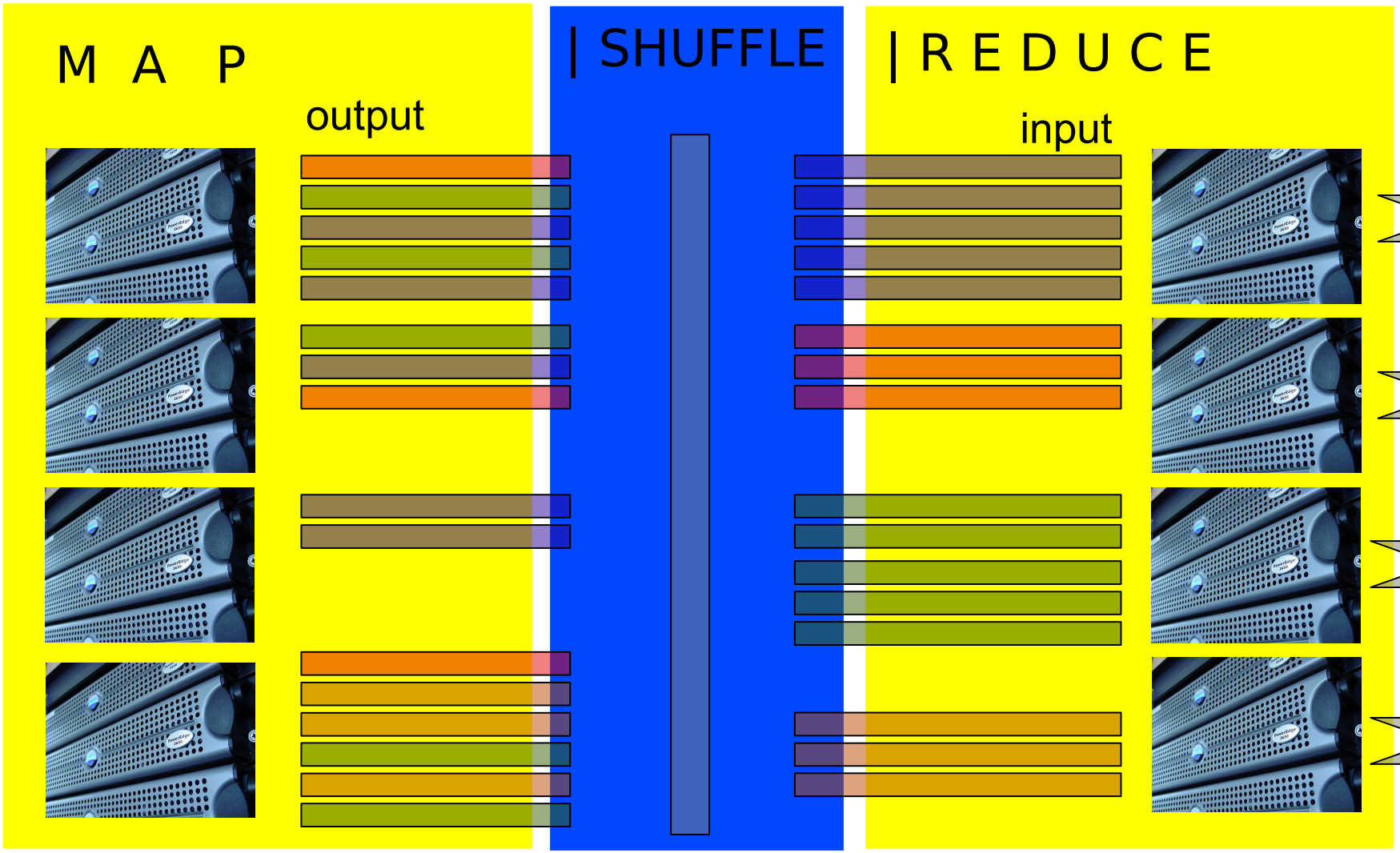
Local to data.



Local to data.
Outputs a lot less data.
Output can cheaply move.



Local to data.
Outputs a lot less data.
Output can cheaply move.



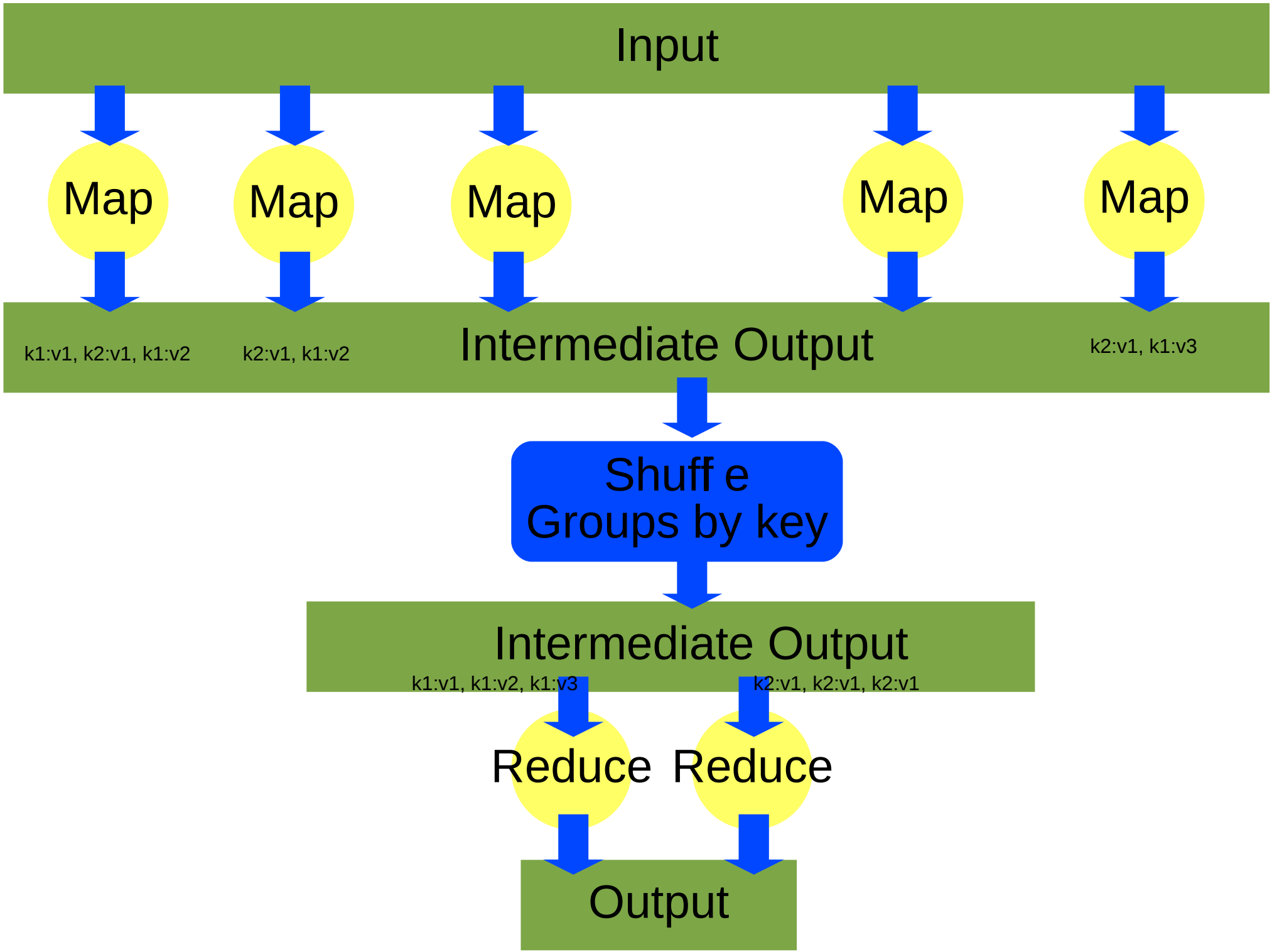
Local to data.
Outputs a lot less data.
Output can cheaply move.

Shuffle sorts input by key.
Reduces output significantly.

```
private IntWritable one = new IntWritable(1);
private Text hostname = new Text();


public void map(LongWritable key, Text value,
OutputCollector<Text, IntWritable> output,
Reporter reporter) throws IOException {
  String line = value.toString();
  StringTokenizer tokenizer = new StringTokenizer(line);
  while (tokenizer.hasMoreTokens()) {
    hostname.set(getHostname(tokenizer.nextToken()));
    output.collect(hostname, one);
  }
}


public void reduce(Text key, Iterator<IntWritable>
values, OutputCollector<Text, IntWritable> output,
Reporter reporter) throws IOException {
  int sum = 0;
  while (values.hasNext()) {
    sum += values.next().get();
  }
  output.collect(key, new IntWritable(sum));
}
```


Input
Map
Map
Map
Map
Map
k1:v1, k2:v1, k1:v2
k2:v1, k1:v2
Intermediate Output
k2:v1, k1:v3
Shuff e
Groups by key
Intermediate Output
k1:v1, k1:v2, k1:v3
k2:v1, k2:v1, k2:v1
Reduce
Reduce
Output

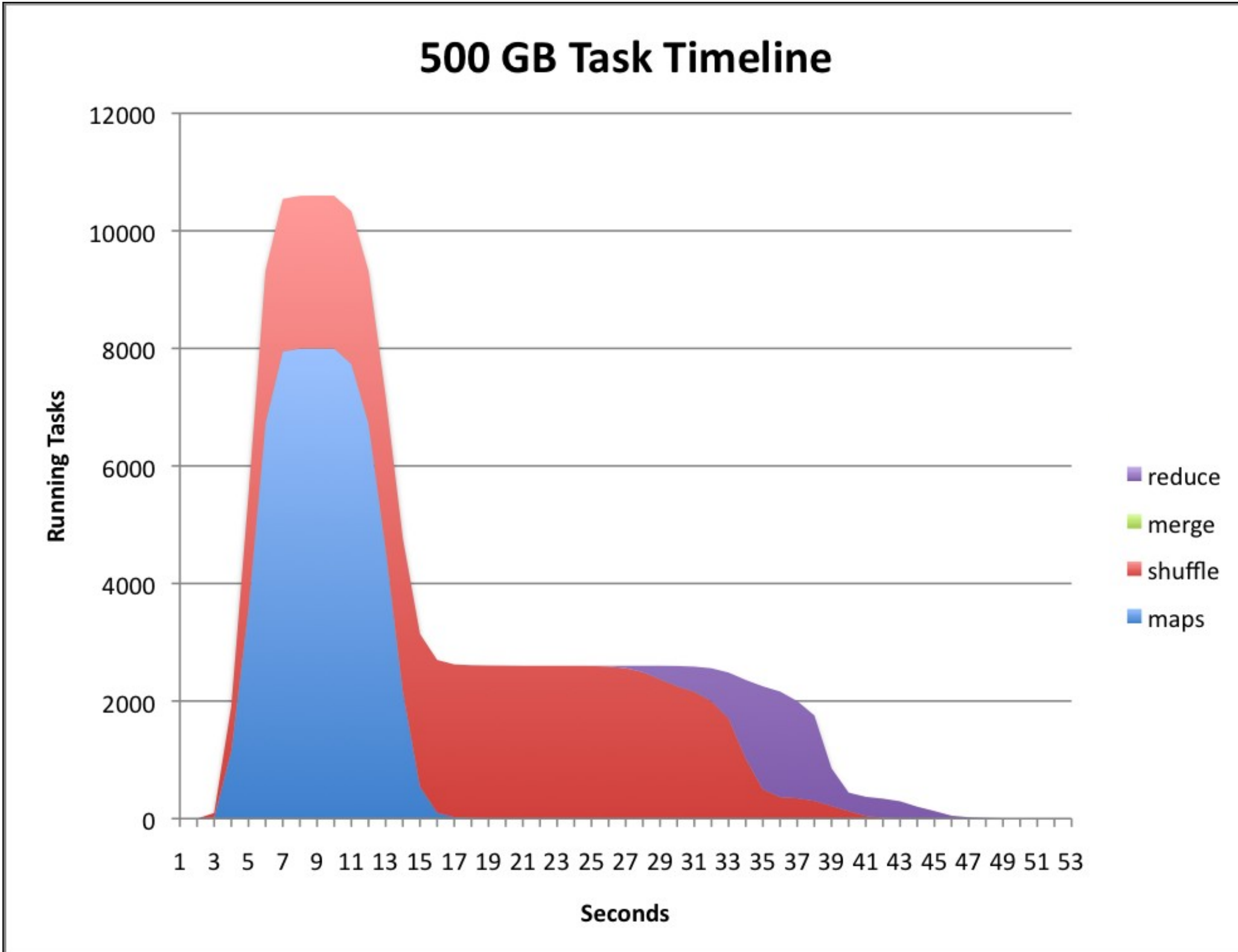
Petabyte sorting benchmark

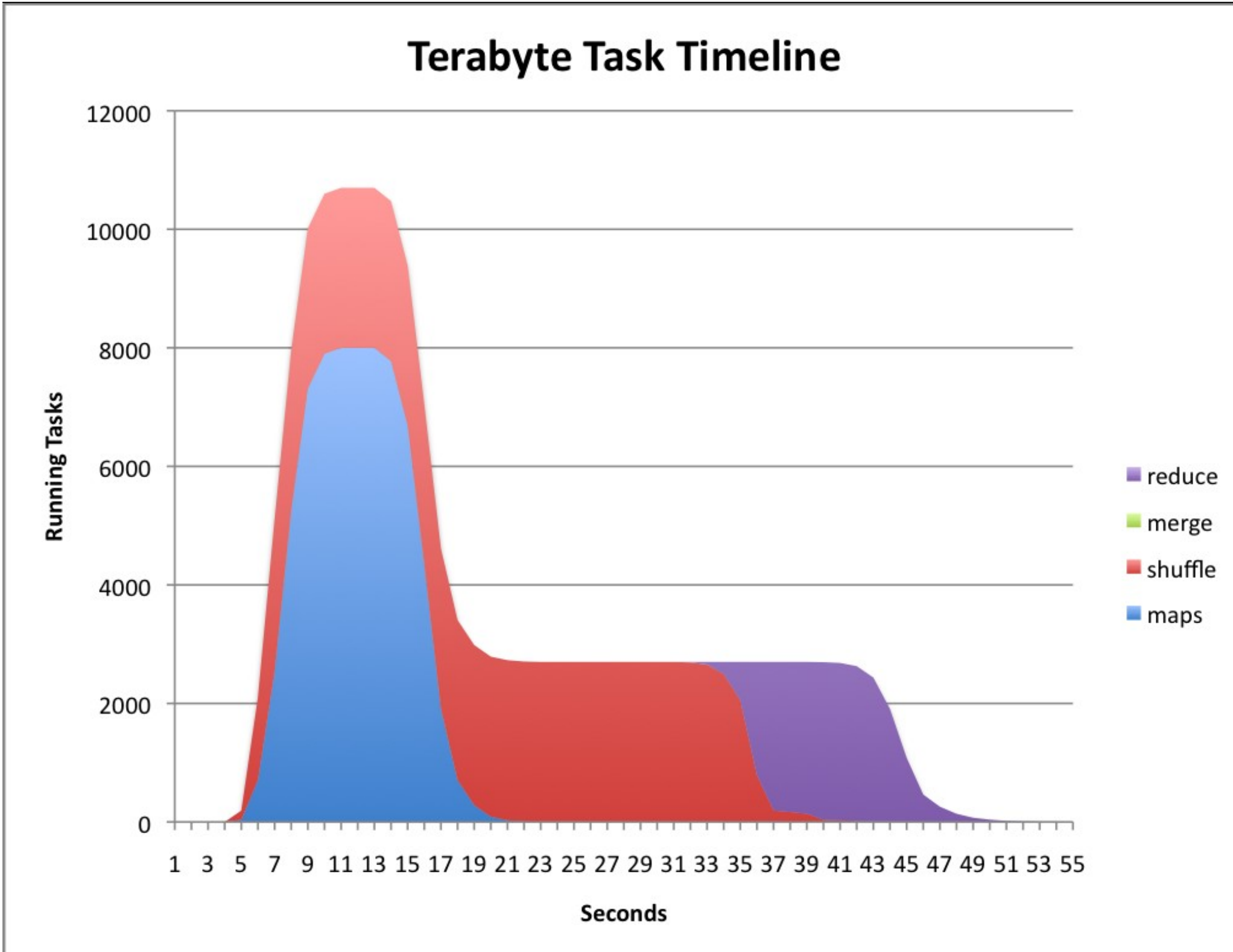
Bytes	Nodes	Maps	Reduces	Replication	Time
500,000,000,000	1406	8000	2600	1	59 seconds
1,000,000,000,000	1460	8000	2700	1	62 seconds
100,000,000,000,000	3452	190,000	10,000	2	173 minutes
1,000,000,000,000,000	3658	80,000	20,000	2	975 minutes

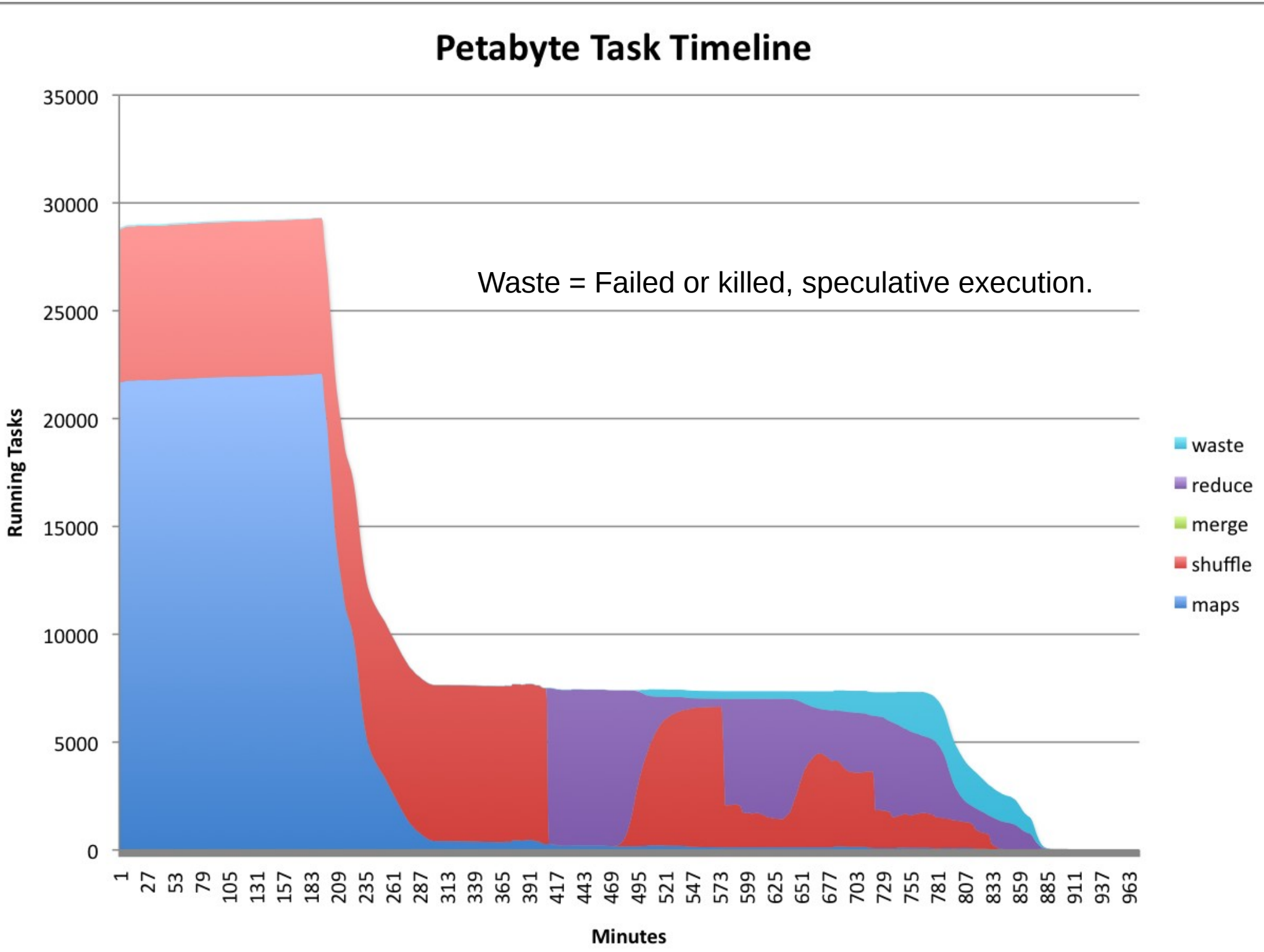
Per node: 2 quad core Xeons @ 2.5ghz, 4 SATA disks, 8G RAM (upgraded to

16GB before petabyte sort), 1 gigabit ethernet.

Per Rack: 40 nodes, 8 gigabit ethernet uplinks.


500 GB Task Timeline
12000
10000
8000
6000
4000
2000
0
Running Tasks
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53
Seconds
reduce
merge
shuffle
maps


Terabyte Task Timeline
12000
10000
8000
6000
4000
2000
0
Running Tasks
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55
Seconds
reduce
merge
shuffle
maps


Petabyte Task Timeline
Waste = Failed or killed, speculative execution.
Running Tasks
35000
30000
25000
20000
15000
10000
5000
0
1
27
53
79
105
131
157
183
209
235
261
287
313
339
365
391
417
443
469
495
521
547
573
599
625
651
677
703
729
755
781
807
833
859
885
911
937
963
Minutes
waste
reduce
merge
shuffle
maps

What was left out

- Combiners compact map output.
- Language choice: Java vs. Dumbo vs. PIG …
- Size of input files does matter.
- Facilities for chaining jobs.
- Logging facilities.
- Monitoring.
- Job tuning (number of mappers and reducers)
- ...

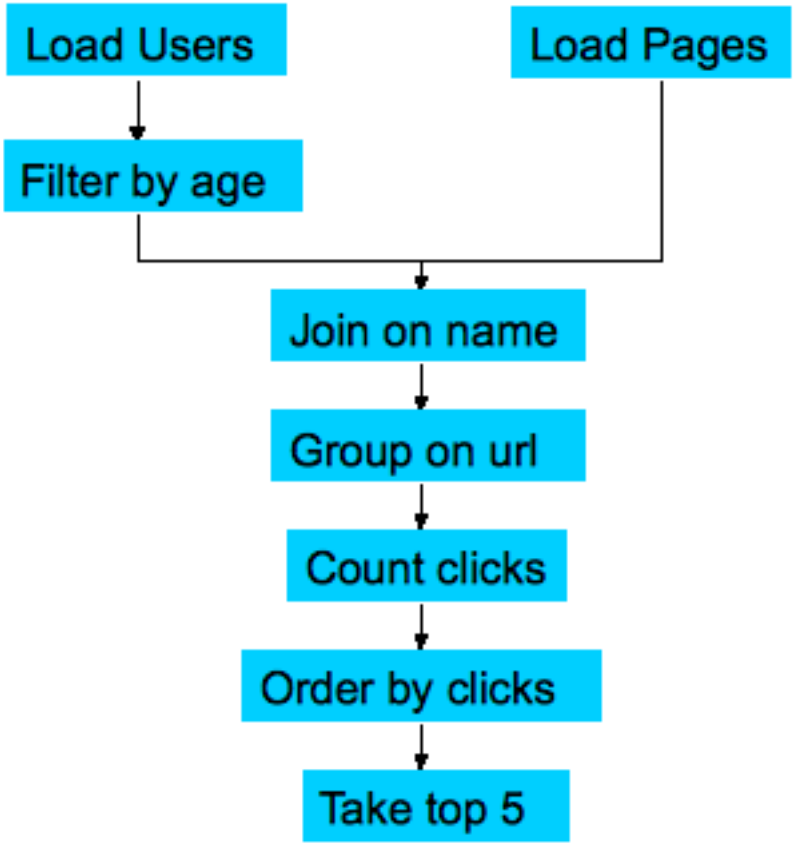
Hadoop ecosystem.

Higher level languages.

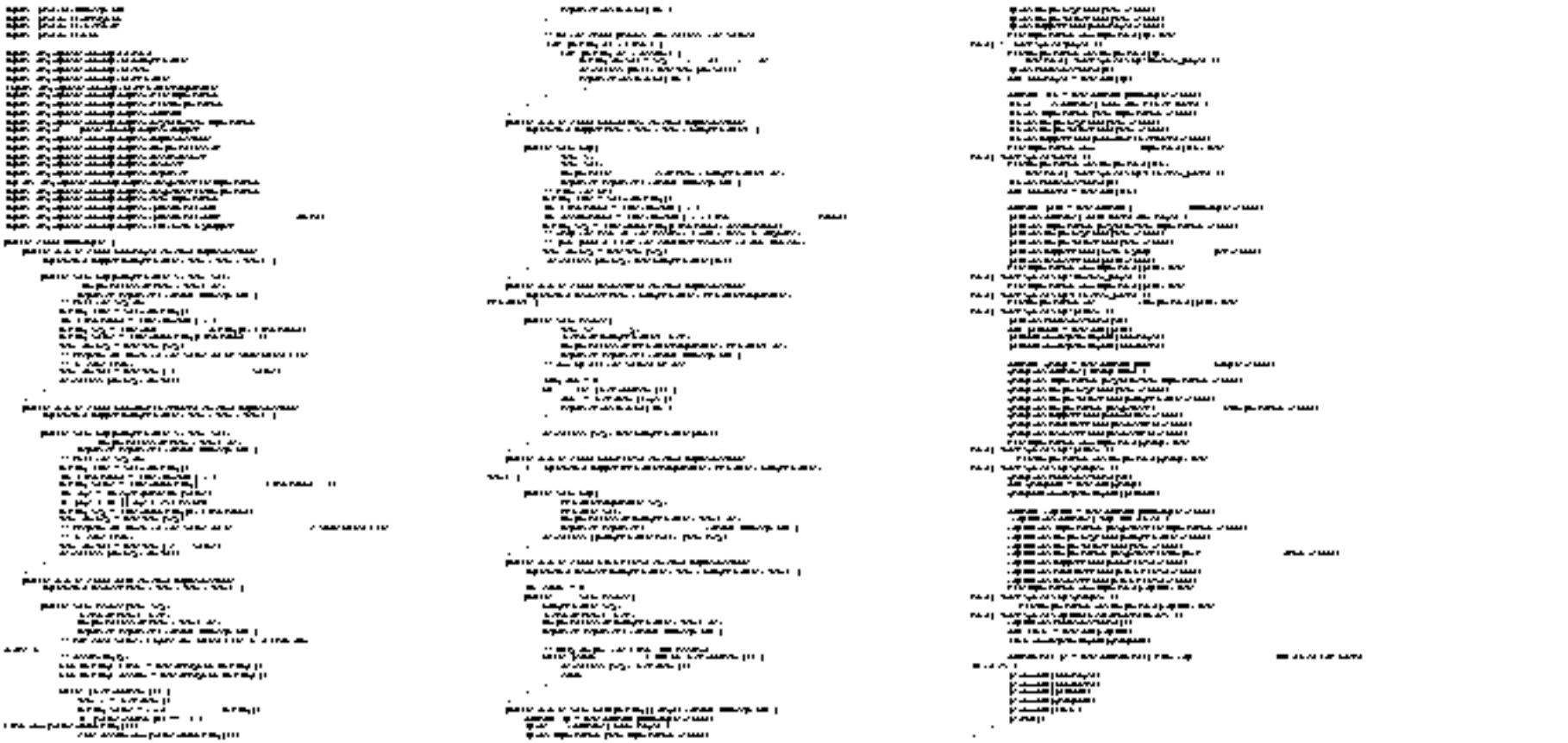
Cascading


Y!

Suppose you have user data in one file, website data in another, and you need to find the top 5 most visited pages by users aged 18 - 25.



Example from PIG presentation at Apache Con EU 2009



Example from PIG presentation at Apache Con EU 2009

```
Users = load 'users' as (name, age);
Fltrd = filter Users by
        age >= 18 and age <= 25;
Pages = load 'pages' as (user, url);
Jnd = join Fltrd by name, Pages by user;
Grpd = group Jnd by url;
Smmd = foreach Grpd generate group,
       COUNT(Jnd) as clicks;
Srtd = order Smmd by clicks desc;
Top5 = limit Srtd 5;
store Top5 into 'top5sites';
```

Example from PIG presentation at Apache Con EU 2009

```
[
  {publisher: 'Scholastic',
   author: 'J. K. Rowling',
   title: 'Deathly Hallows',
   year: 2007},

  {publisher: 'Scholastic',
   author: 'J. K. Rowling',
   title: 'Chamber of Secrets',
   year: 1999,
   reviews: [
     {rating: 10, user: 'joe', review: 'The best ...'},
     {rating: 6, user: 'mary', review: 'Average ...'}]},

  {publisher: 'Scholastic',
   author: 'J. K. Rowling',
   title: 'Sorcerers Stone',
   year: 1998},

  {publisher: 'Scholastic',
   author: 'R. L. Stine',
   title: 'Monster Blood IV',
   year: 1997,
   reviews: [
     {rating: 8, user: 'rob', review: 'High on my list...'},
     {rating: 2, user: 'mike', review: 'Not worth the paper ...',
      discussion:
        [{user: 'ben', text: 'This is too harsh...'},
         {user: 'jill', text: 'I agree ...'}]}]},

  {publisher: 'Grosset',
   author: 'Carolyn Keene',
   title: 'The Secret of Kane',
   year: 1930}
]
```

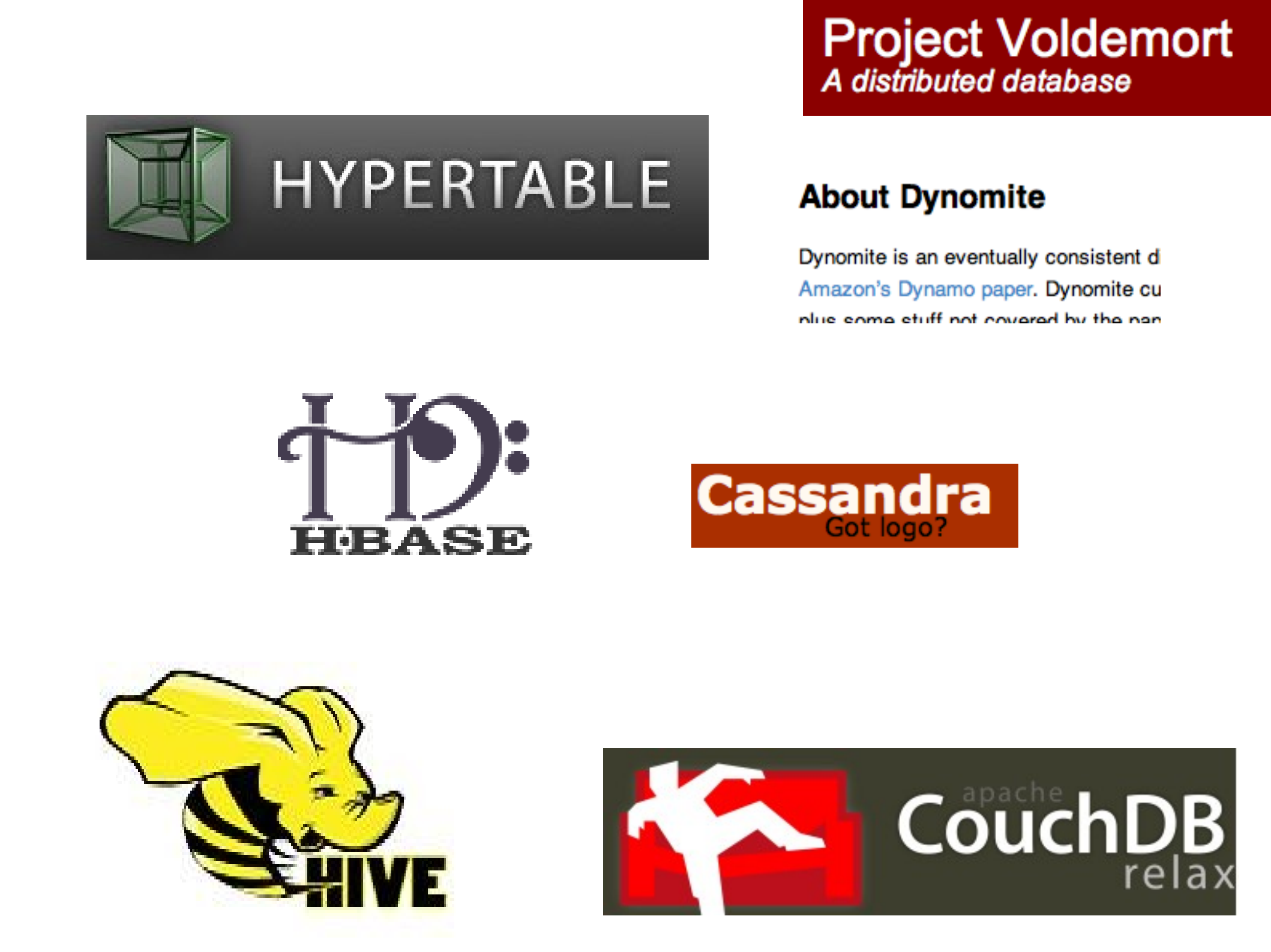
Example from JAQL documentation.

```
// Query 2. Find the authors and titles of books that have received
// a review.
for( $b in hdfsRead('books') )
 if( exists($b.reviews) )
  [{ $b.author, $b.title }];

// result...
[
  {author: 'J. K. Rowling', title: 'Chamber of Secrets'},
  {author: 'R. L. Stine',   title: 'Monster Blood IV'}
];
```

Example from JAQL documentation.

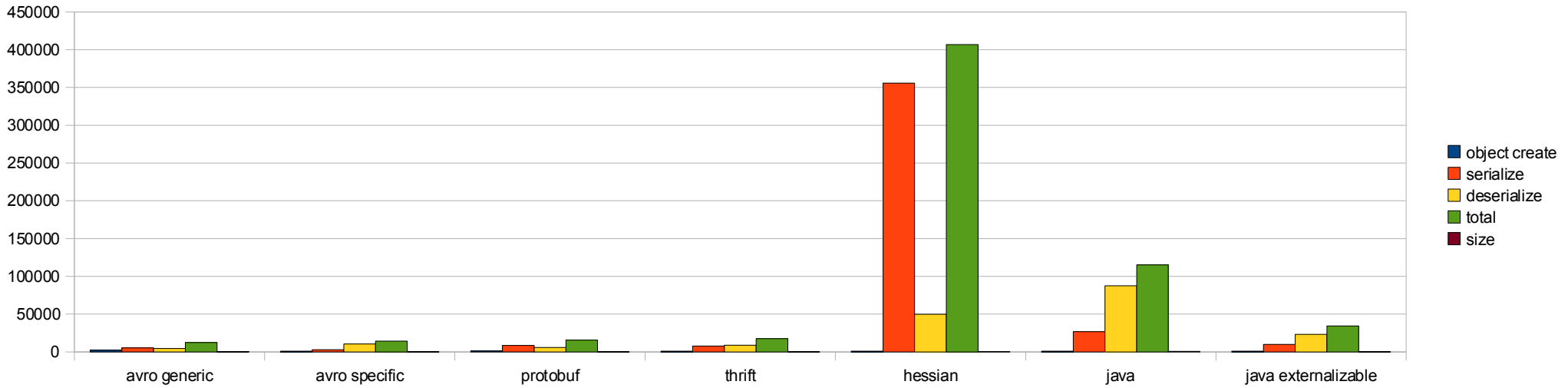
(Distributed) storage.


Project Voldemort
A distributed database
HYPERTABLE
About Dynomite
Dynomite is an eventually consistent d
Amazon's Dynamo paper. Dynomite cu
HBASE
Cassandra
Got logo?
HIVE
apache
CouchDB
relax

Libraries built on top.

Google code **protobuf**

Protocol Buffers - Google's data interchange format




MAHOUT

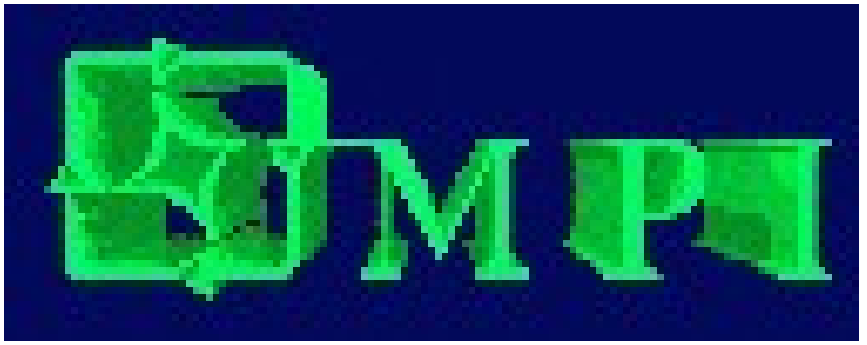

nutch


Apache
Solr


katta

Alternative approaches.


Greenplum


MPI


PVM
Parallel Virtual Machine


ERLANG


OpenMP
™

Get involved!

Do you love:

Do you love:

Solving hard problems?

Do you love:

Solving hard problems?

Communicating your solution?

Do you love:

Solving hard problems?

Communicating your solution?

Working with excellent teams?

Do you love:

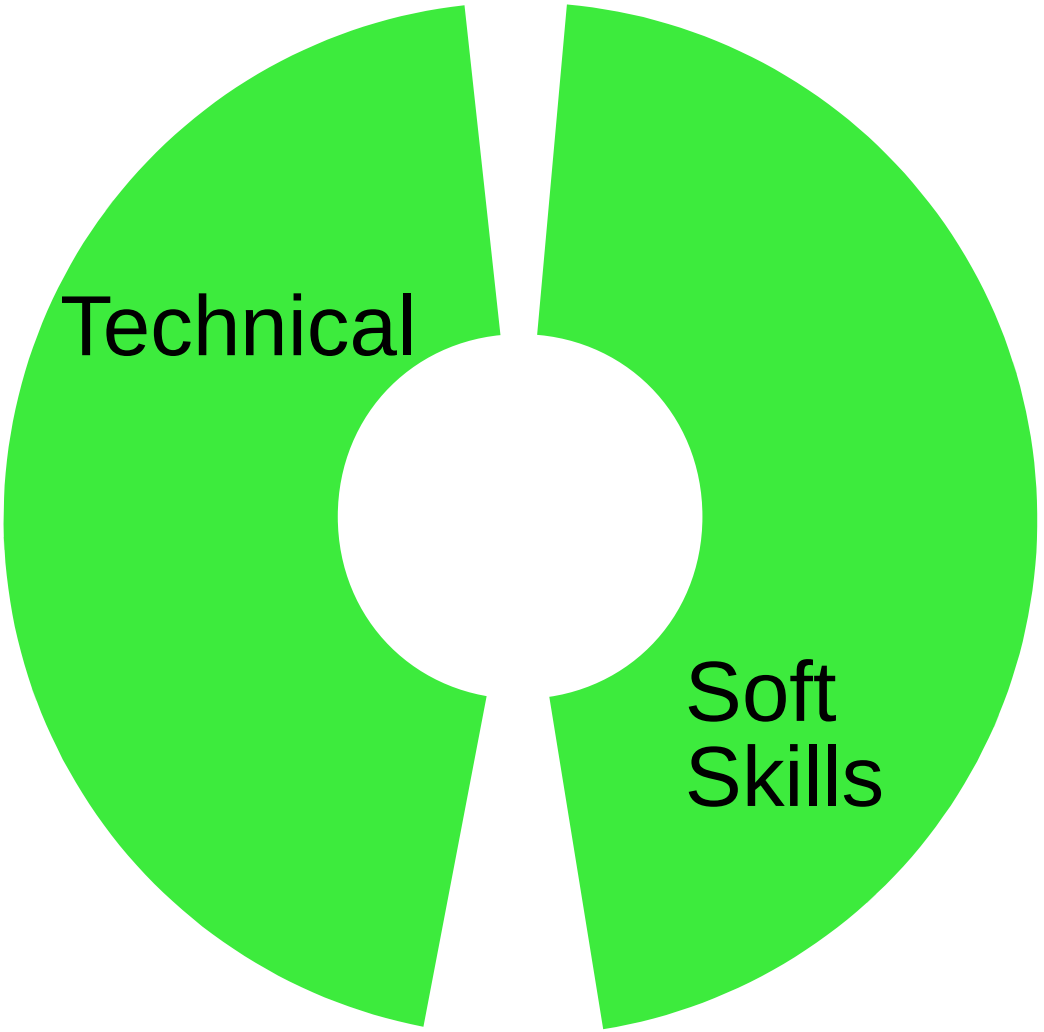
Solving hard problems?

Communicating your solution?

Working with excellent teams?

Picture by: July 9, 2006 by trackrecord, http://www.flickr.com/photos/trackrecord/185514449

Skills to learn:

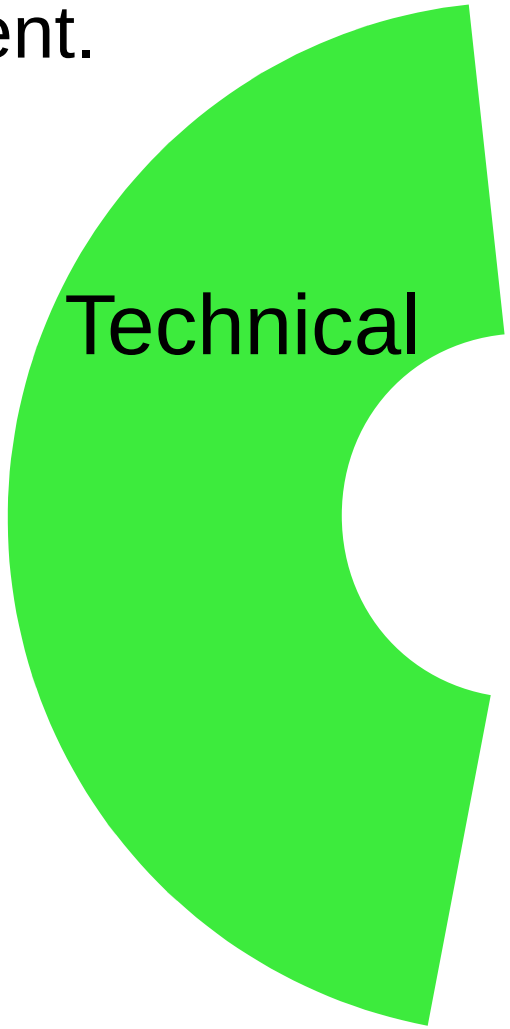


Source control system.

Continuous integration.

Test-first development.

Issue-tracker.



Create readable patches.

Communicate and discuss solutions.

Review others code.

Work in large, distributed teams.

How?

- First time users:
  - Documentation in wiki.
- Found a bug:
  - Go to JIRA, file a bug.
  - Describe the bug.
  - Create a test to show.
  - Provide a patch.
- Experimenting:
  - Write examples.
- Evaluating:
  - Test performance.
  - Provide comparison.
- Participate on-list.
  - Answer questions.
  - Discuss your use-

Recipe to Apache

- Download the release and use it.
- Subscribe to the mailing-list.
- Questions:
  - Documentation: Wiki.
  - Discussions: Mailing list.
  - Current status: JIRA.
  - History: JIRA for patches, mailing-list for votes.
- Checkout the code and built it.


Google™
SUMMER
OF CODE
2009

*-user@lucene.apache.org

*-dev@lucene.apache.org

July 9, 2006 by trackrecord
http://www.flickr.com/photos/trackrecord/185514449

Love for solving hard problems.

Interest in production ready code.

Interest in parallel systems.

Bug reports, patches, features.

Documentation, code, examples.

Contact Ross Gardler for more information on Apache at universities worldwide.

Message view

From	Grant Ingersoll <gsing...@apache.org>
Subject	Re: Lucene Branding: the TLP, and "Lucene Java"
Date	Wed, 11 Apr 2007 01:13:36 GMT

No, you are not the only one... Many a sleepless night spent on it... :-)

I usually try to refer to it as Lucene Java, but old habits die hard and often times I just call it Lucene. I think the name has a good brand at this point and is very strongly associated w/ the Java library. I seem to recall when they were forming the TLP, that the original proposal was search.a.o, but then changed b/c the ASF didn't like generic names (or at least that is how I recall it.) And, of course, with Hadoop and the potential for Tika/Lius, it isn't just search anymore. I have often thought about an Apache "Text" project, that could eventually hold a whole family of text based tools like Lucene, Tika, Hadoop, Solr, etc. plus things like part of speech taggers, clustering/classification algorithms, UIMA, etc. all under one roof. But that is just my two cents and I don't know if it fits with what other people have in mind. There are a lot of OSS tools out there for these things, but none bring together a whole suite under a brand like Apache.

-Grant

Why go for Apache?

Jumpstart your project with proven code.

January 8, 2008 by dreizehn28
http://www.flickr.com/photos/1328/2176949559


Discuss ideas and problems online.
November 16, 2005 [phil h]
http://www.flickr.com/photos/hi-phi/64055296

Become part of the community.