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

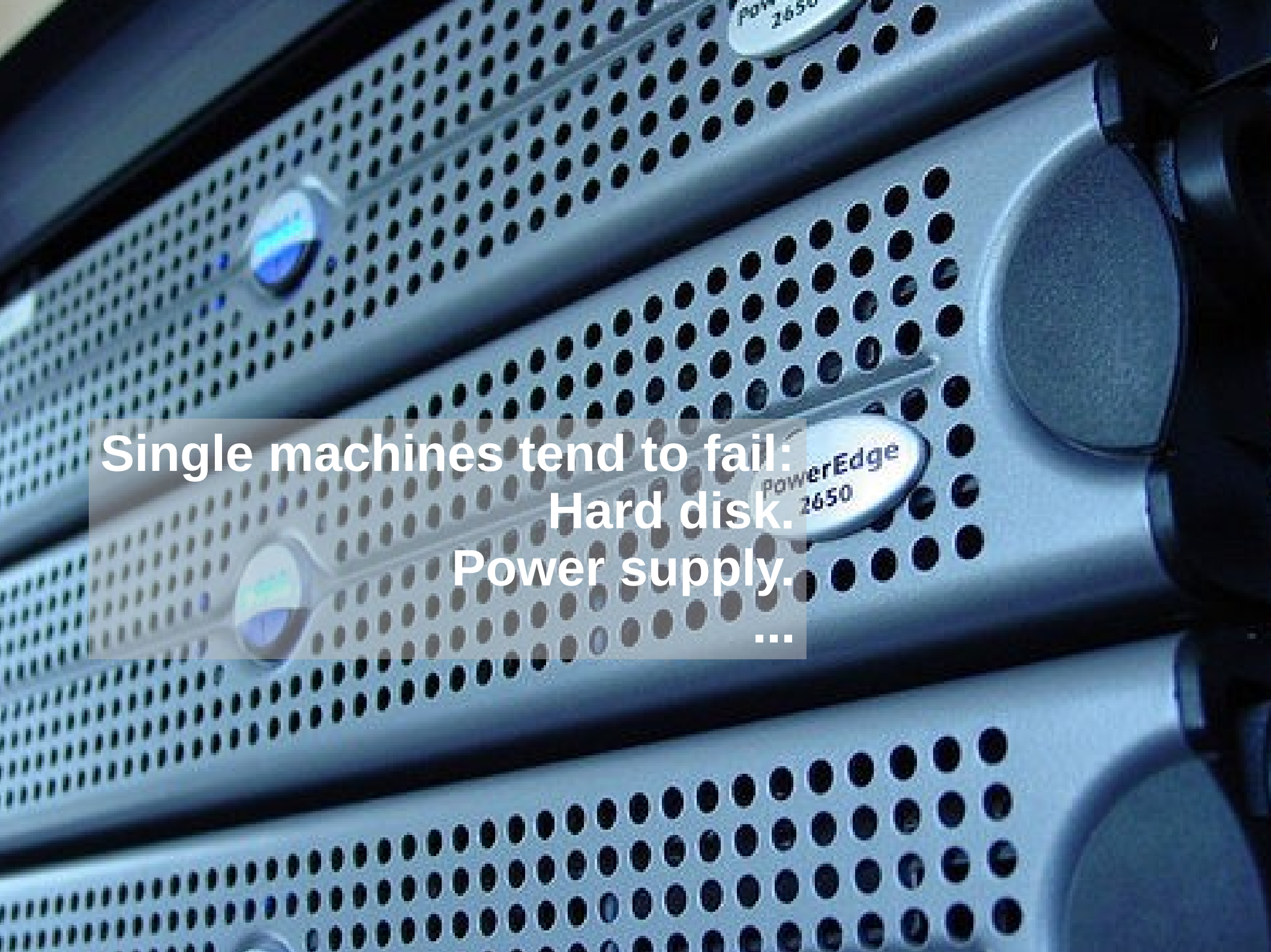
Hello Machine Learning /
Intelligent Data Analysis Group!

Agenda

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


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.

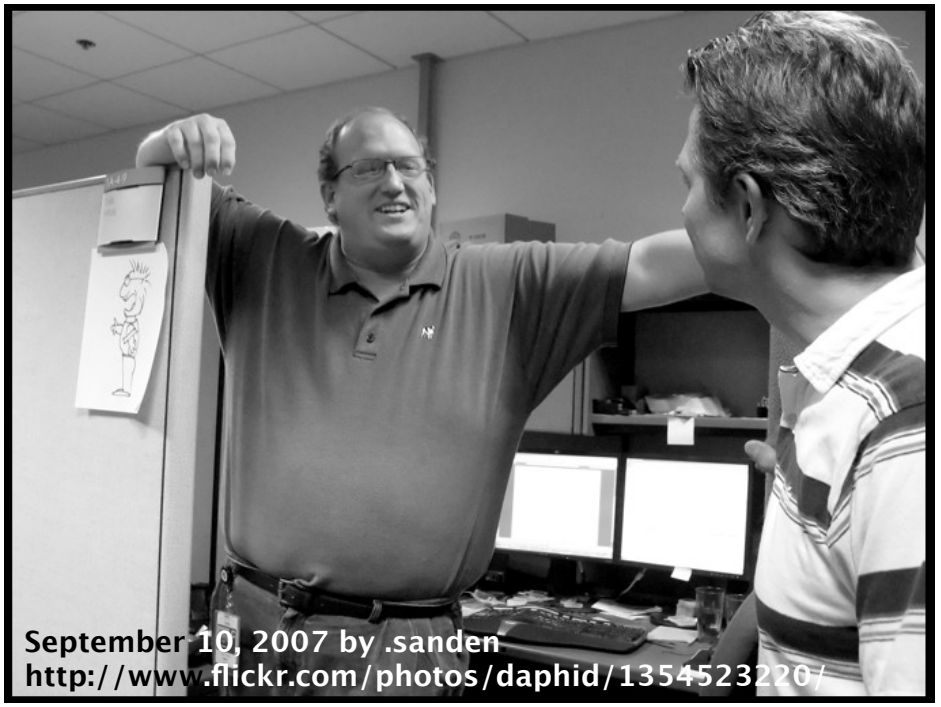
...


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

Requirements

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

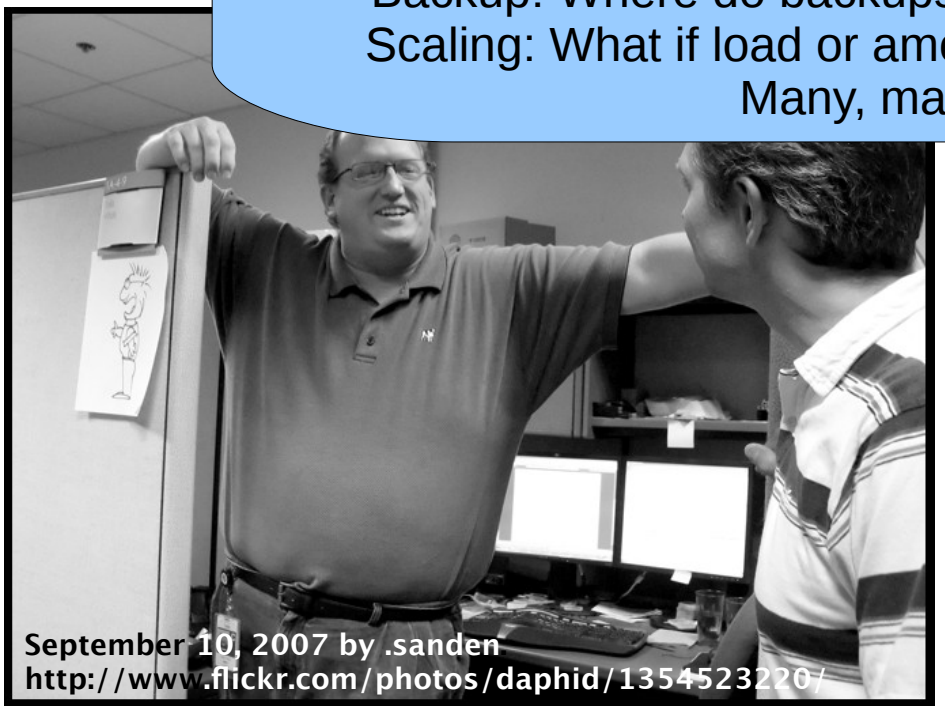
Typical developer



- 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

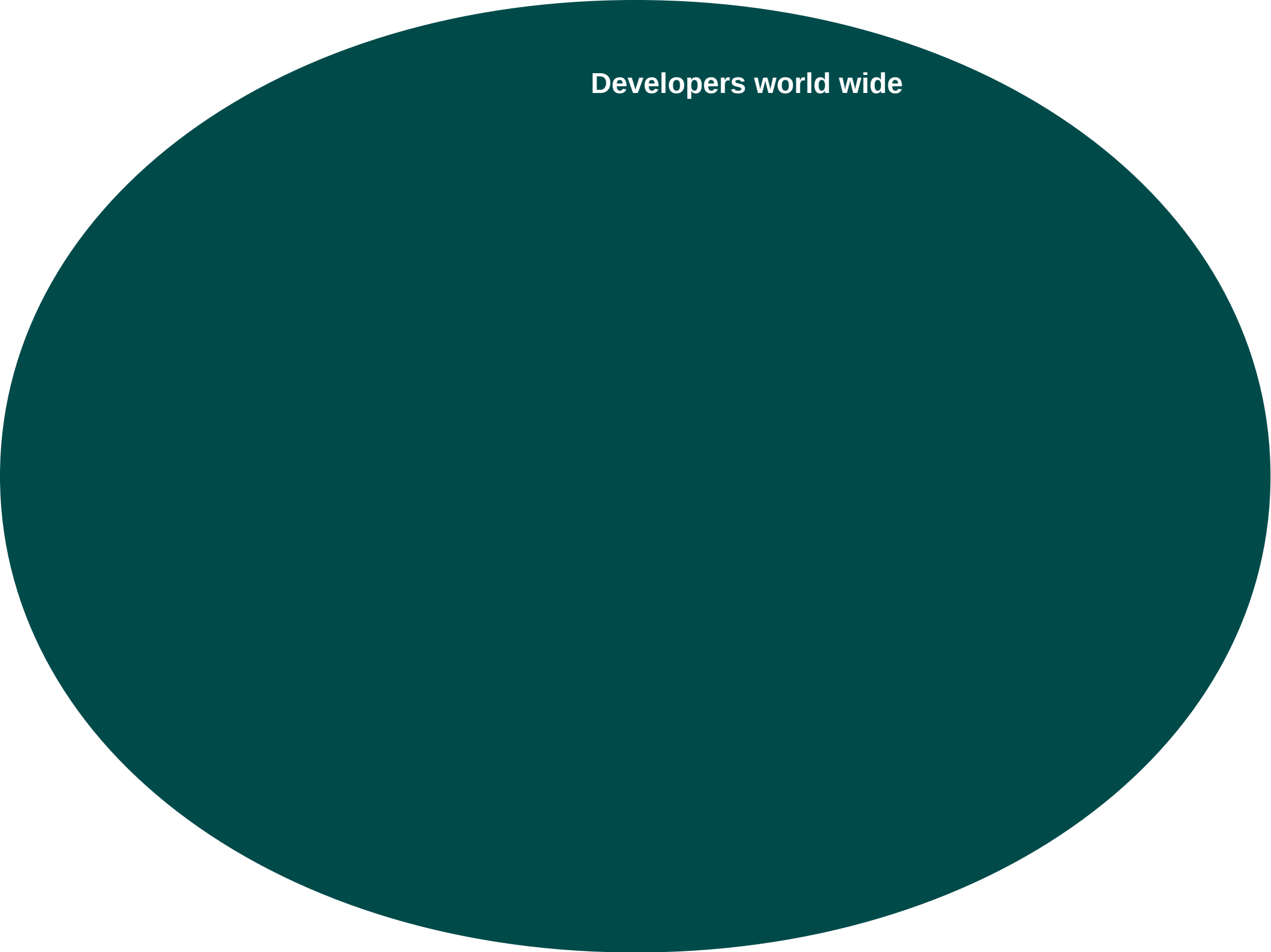
- Built-in backup.
- Built-in failover.
- Easy to use.
- Parallel on rails.

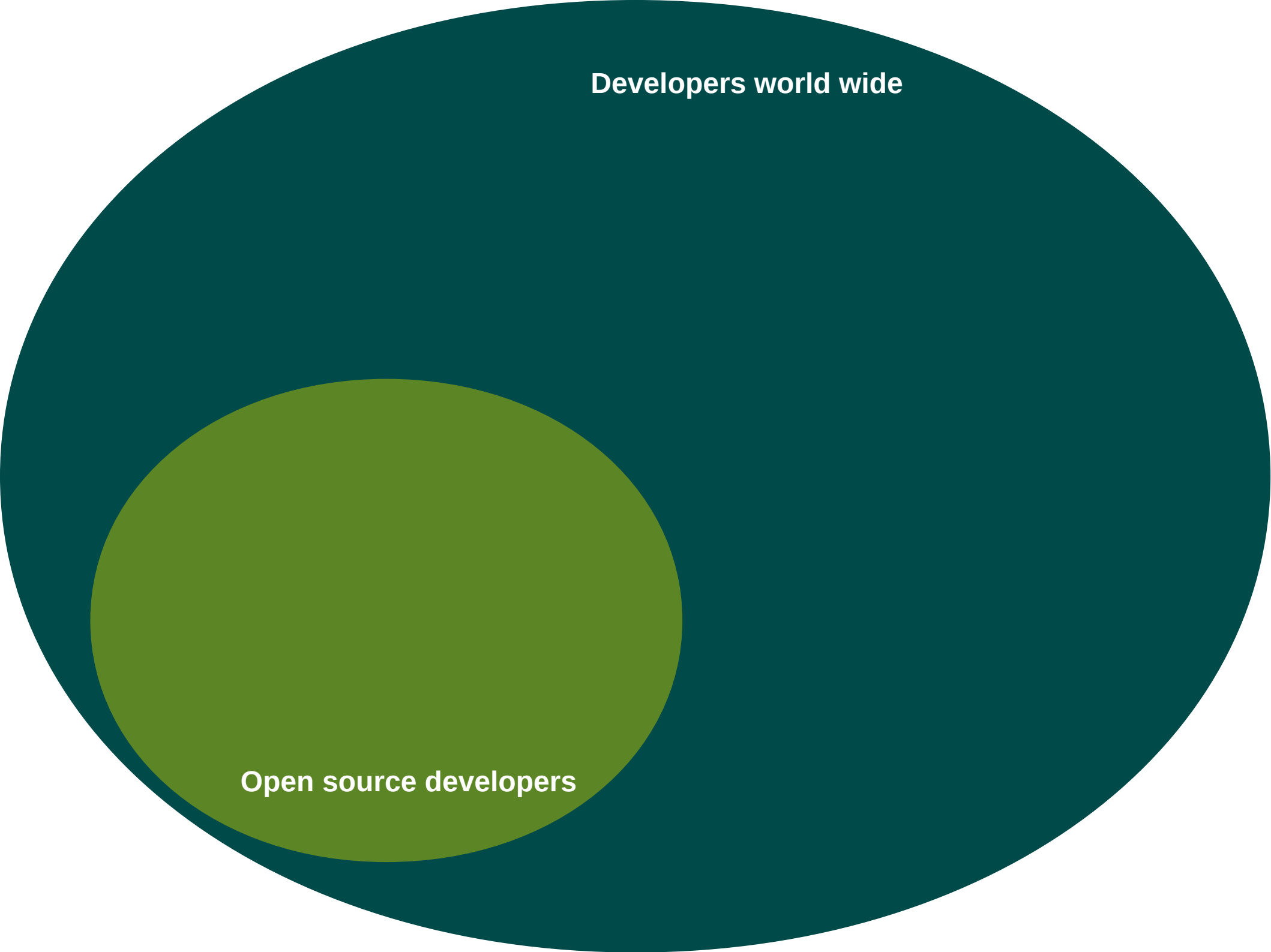
February 29, 2008 by Thomas Claveirole
http://www.flickr.com/photos/thomasclaveirole/2300932656/

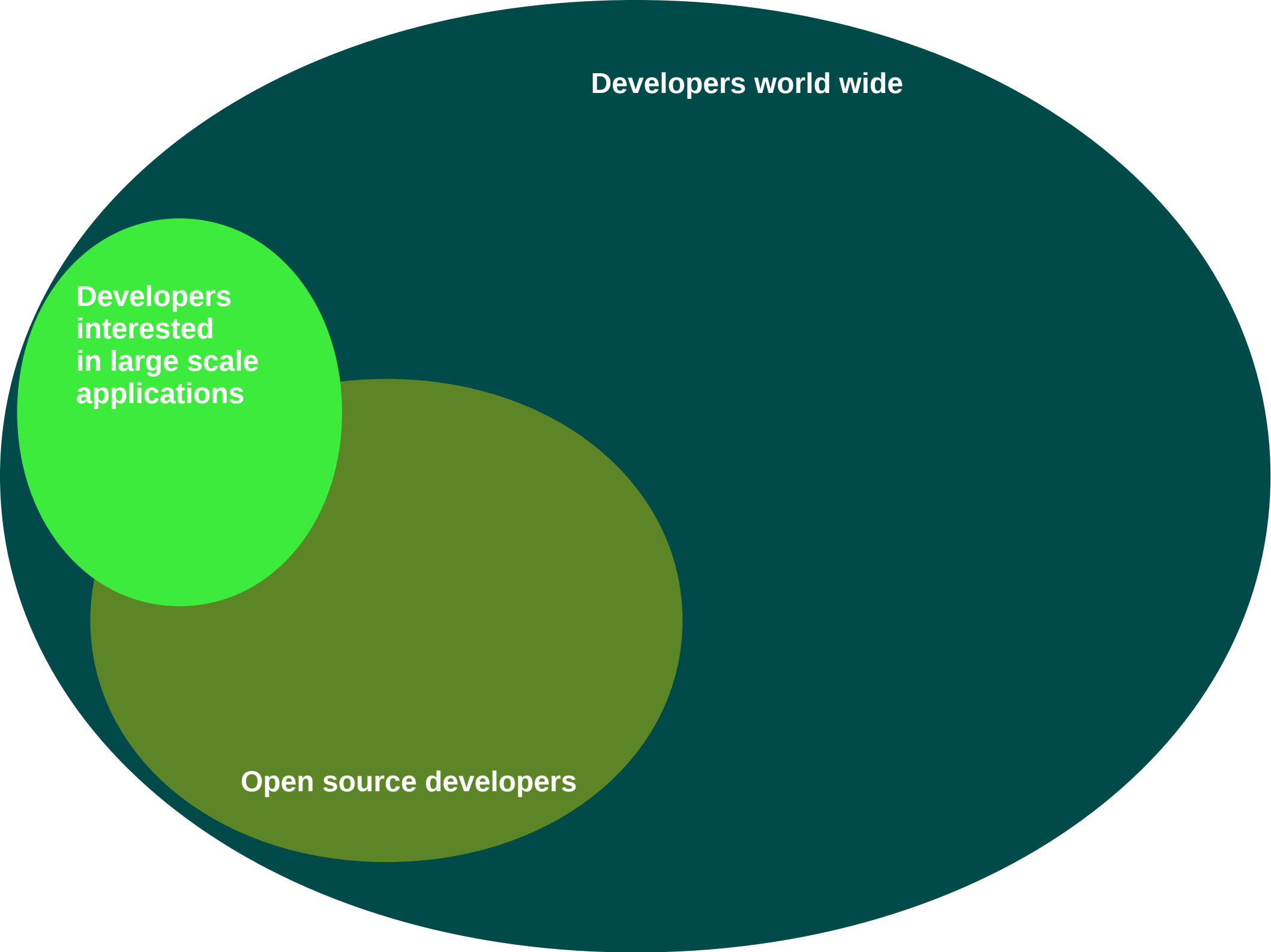
http://www.flickr.com/photos/jaaronfarr/3384940437/
March 25, 2009 by jaaron

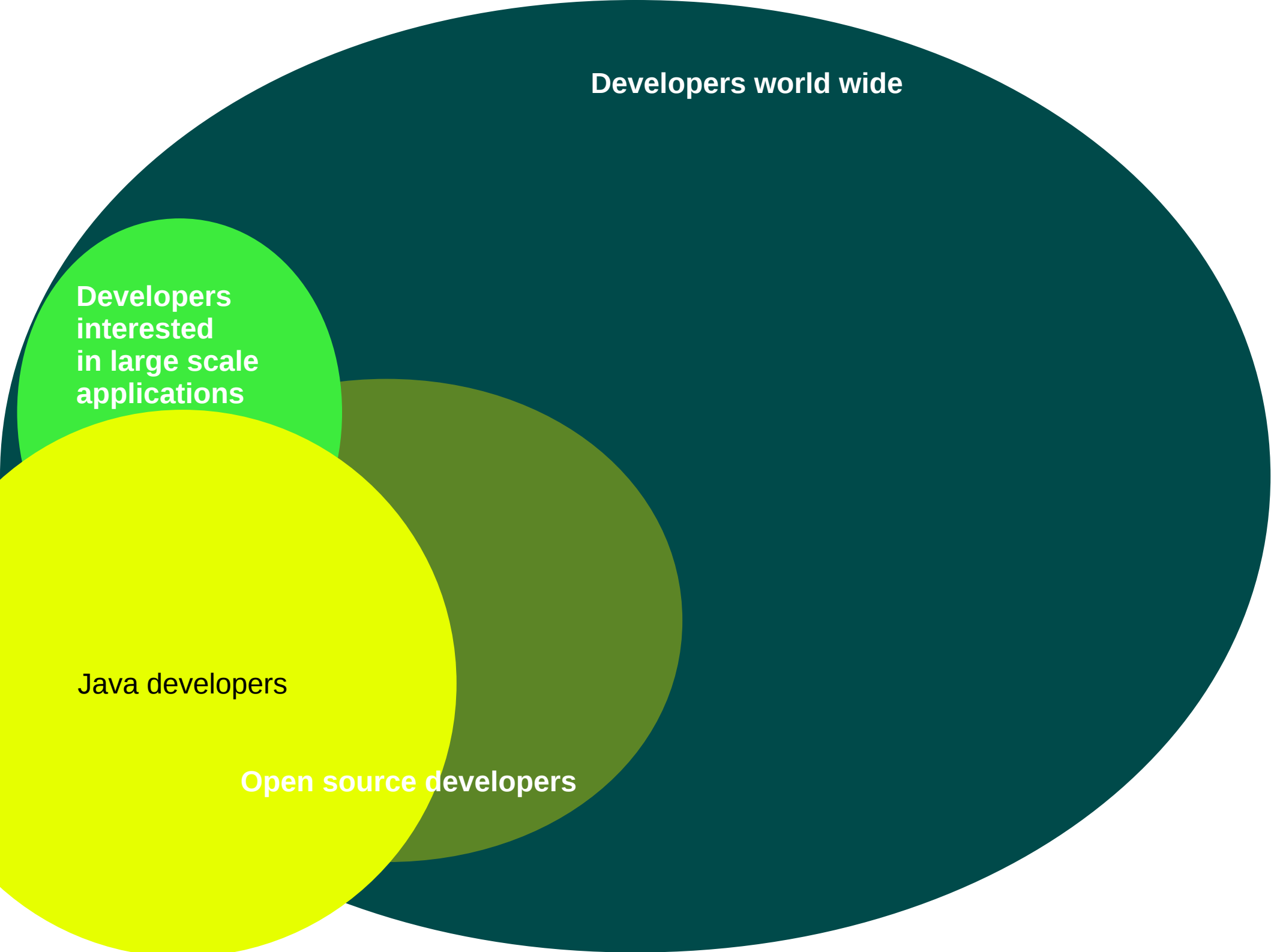
http://www.flickr.com/photos
/jaaronfarr/3385756482/
March 25, 2009 by jaaron

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


Developers world wide

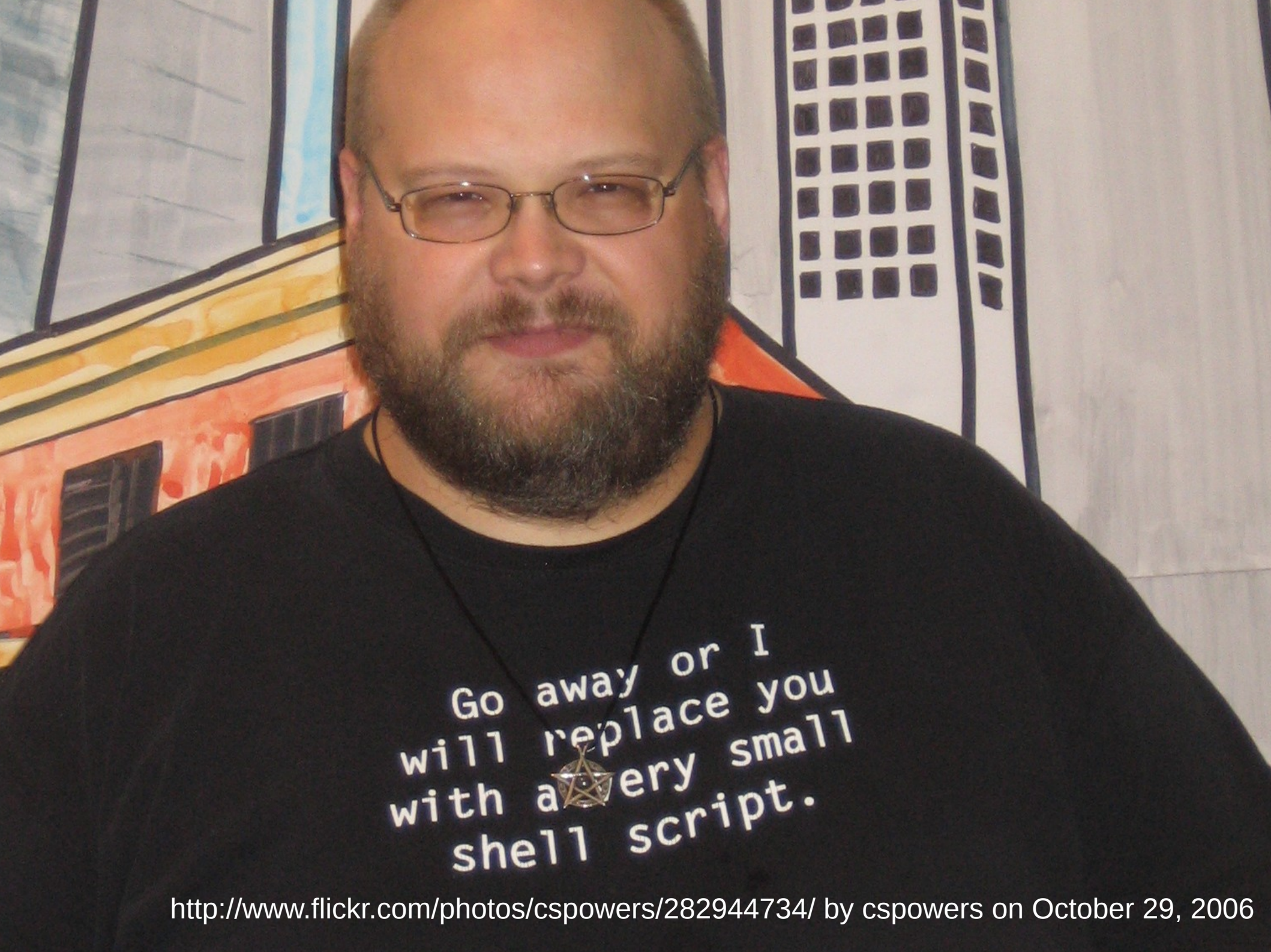

Developers world wide
Open source developers


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)

Some history.

Feb '03 first Map Reduce library @ Google

Oct '03 GFS Paper

Dec '04 Map Reduce paper

Jul '05 Doug reports that nutch uses map reduce

Feb '05 Hadoop moves out of nutch

Apr '07 Y! running Hadoop on 1000 node cluster

Jan '08 Hadoop made an Apache Top Level Project

Hadoop by example

```
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
```

Assumptions:

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

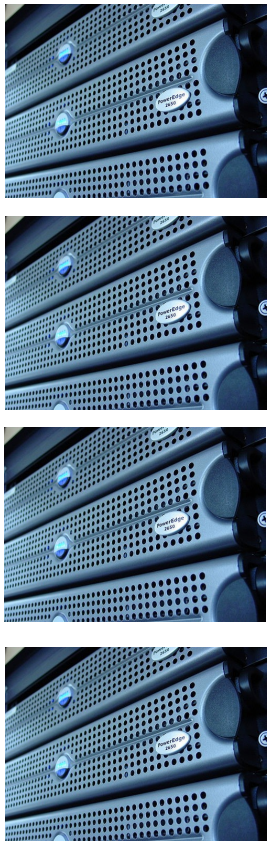



Ideas:

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

Assumptions:

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

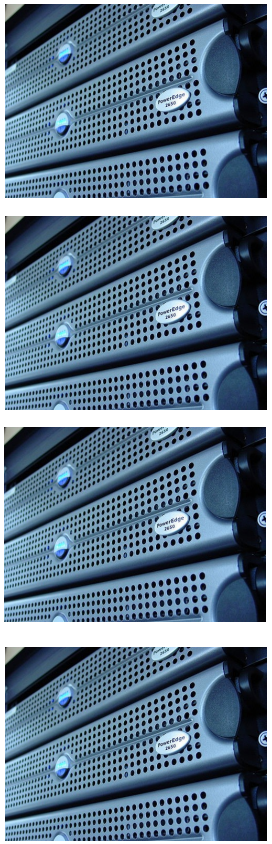


Ideas:

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

Assumptions:

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



Ideas:

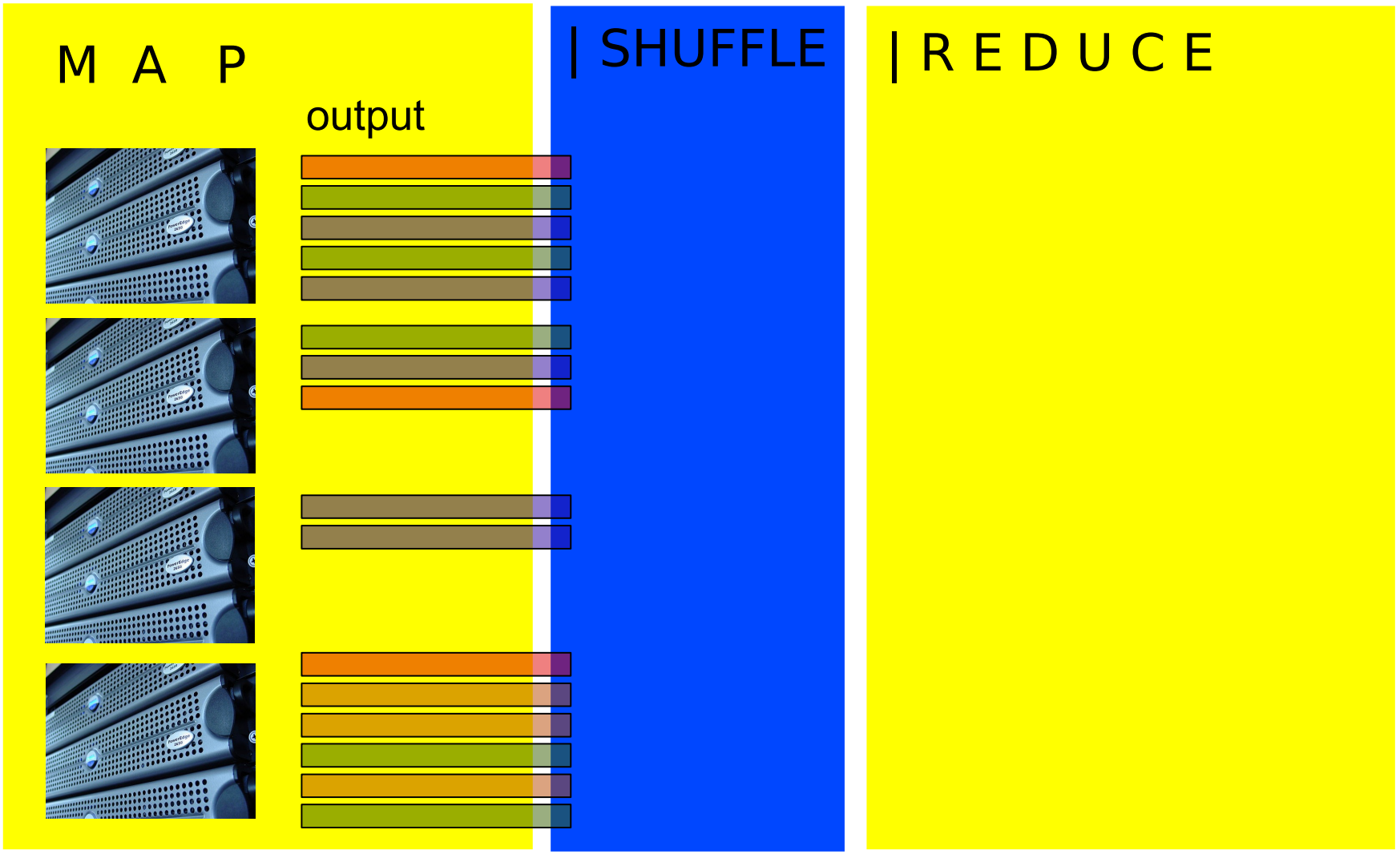
Improve support for large files.
File system API makes scanning easy.

M A P

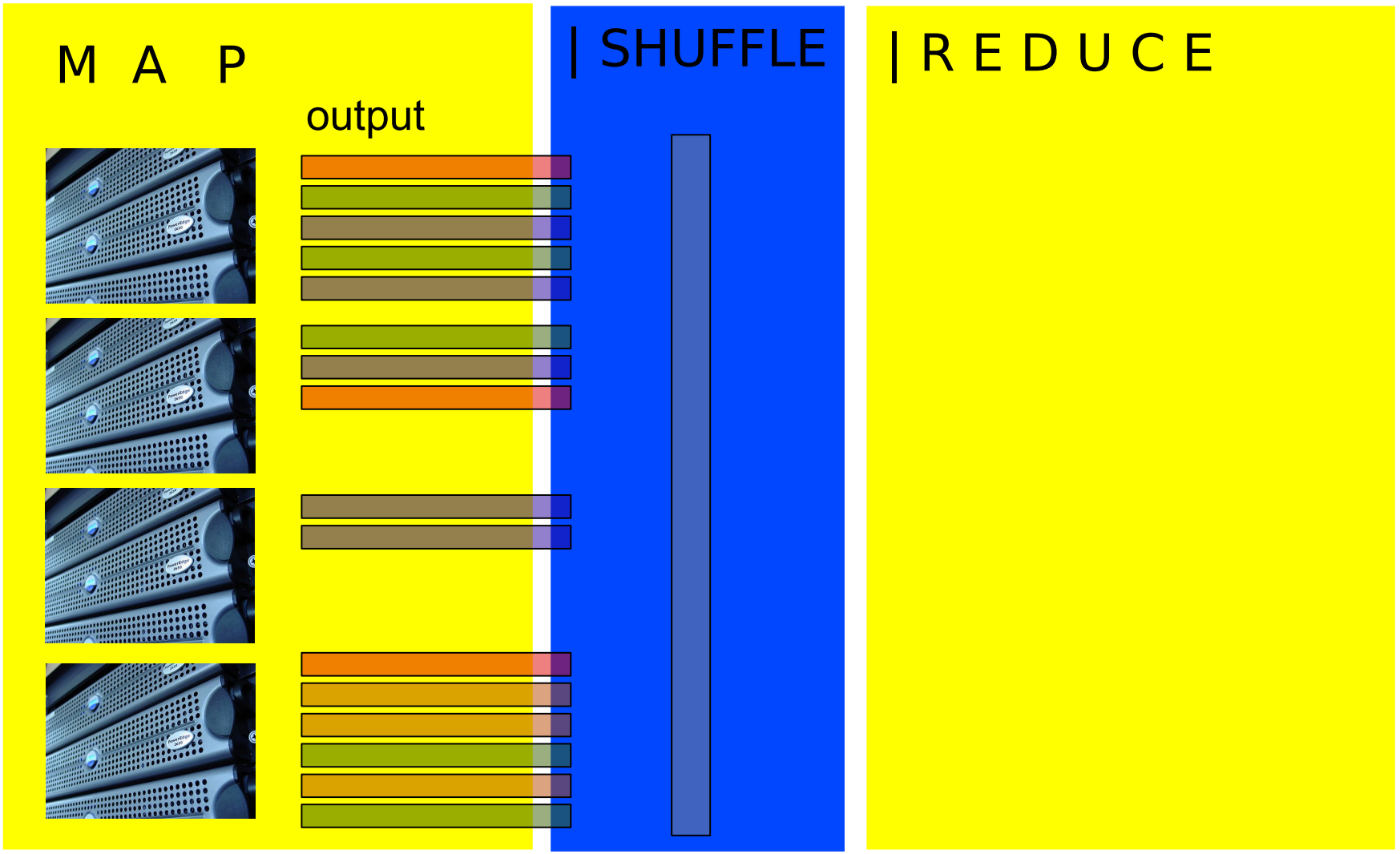
| SHUFFLE

| R E D U C E

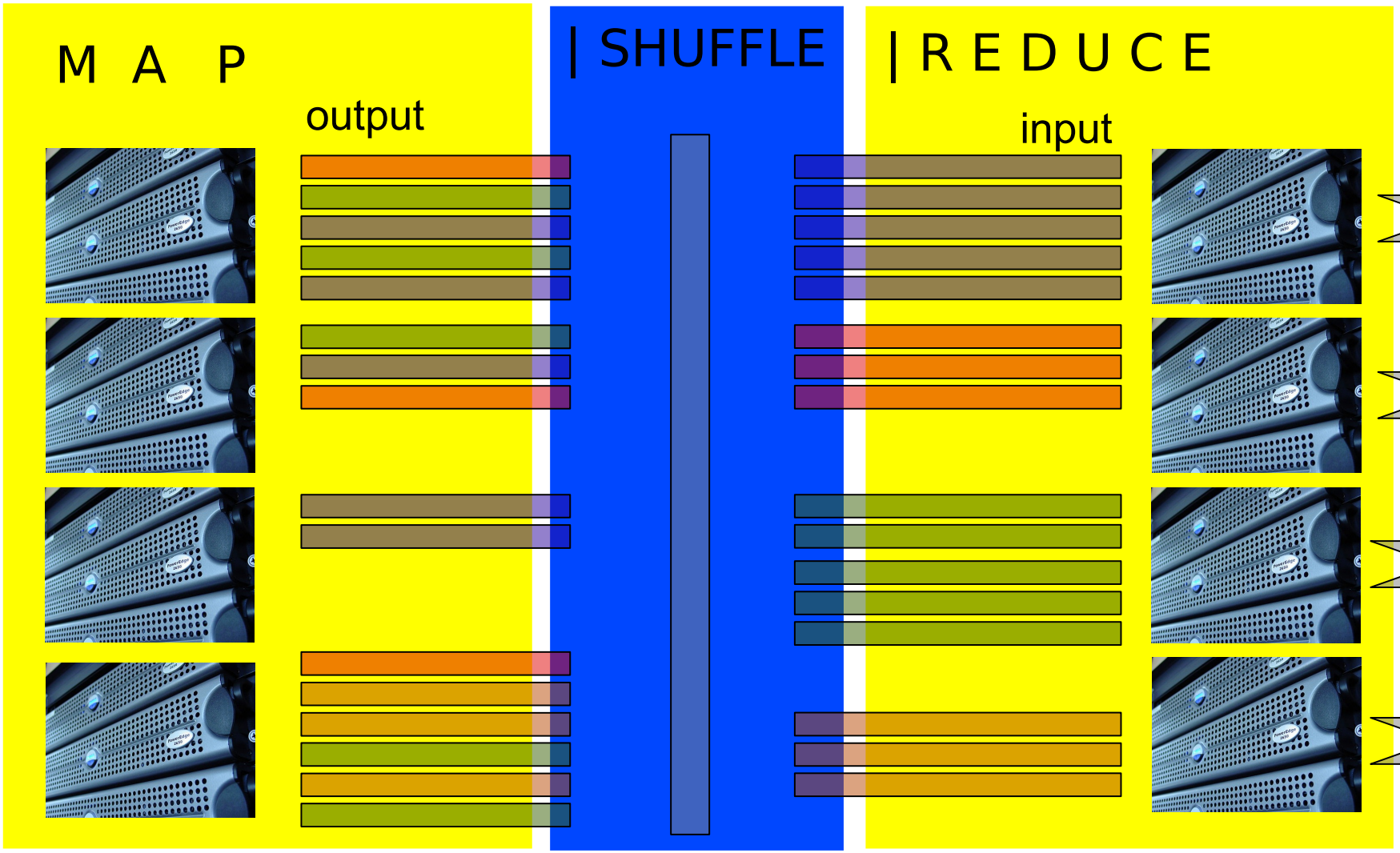
Local to data.



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));
}
```

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)
- ...

Options for running Hadoop

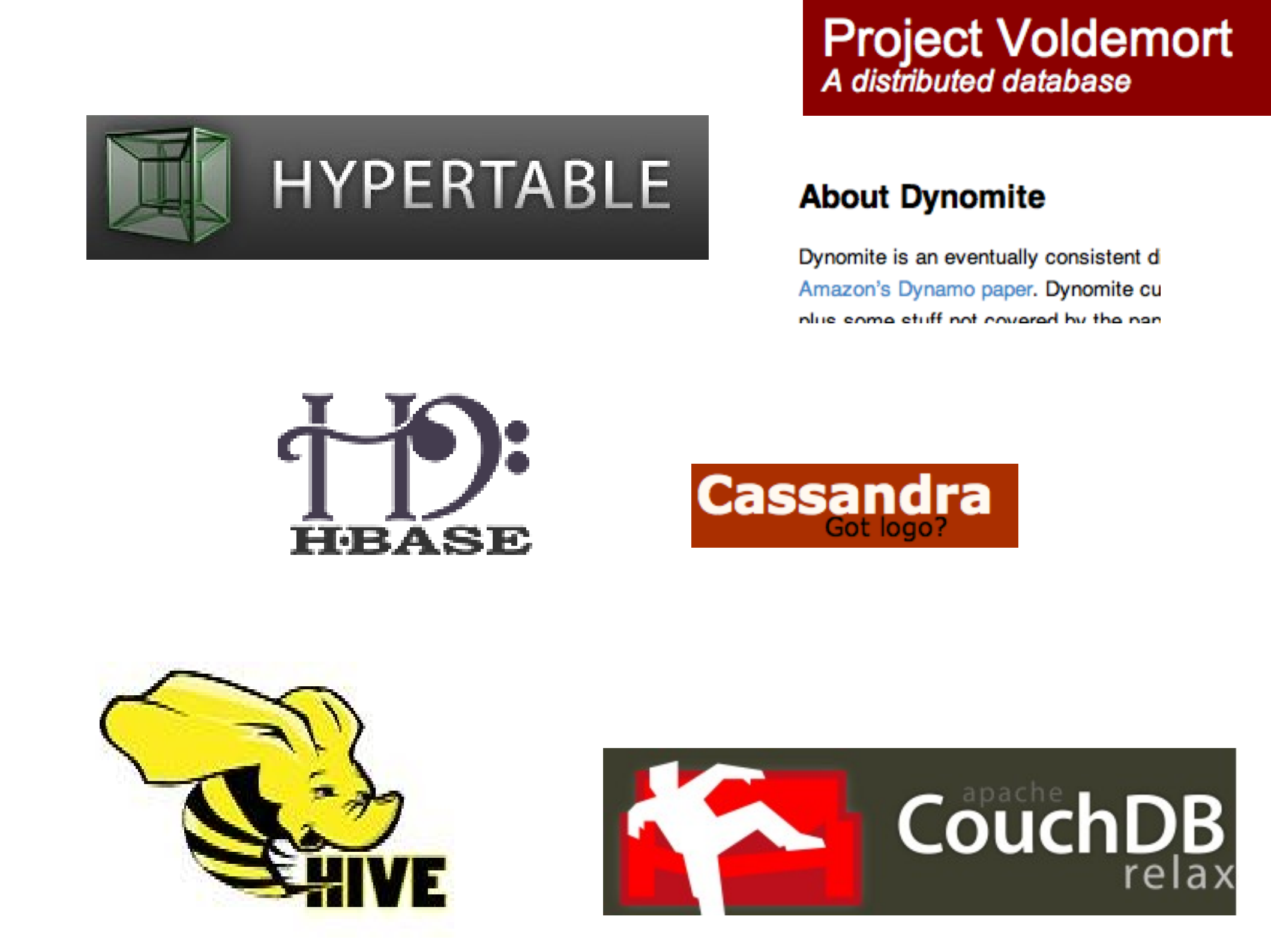
- Amazon Elastic Map Reduce (has drawbacks)
- Amazon EC2 with custom Hadoop cluster.
- Roll your own.

Hadoop ecosystem.

Higher level languages.

Cascading

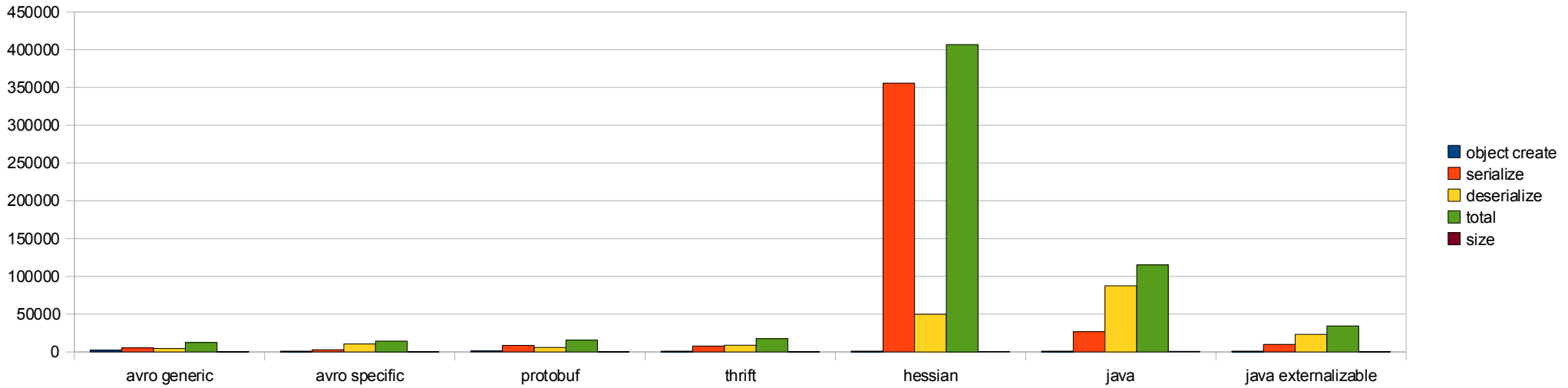
(Distributed) storage.


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

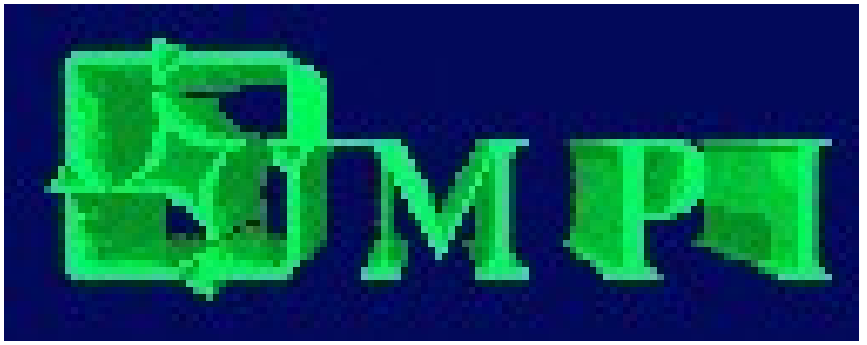


Apache Solr

nutch

Alternative approaches.


Greenplum


MPI

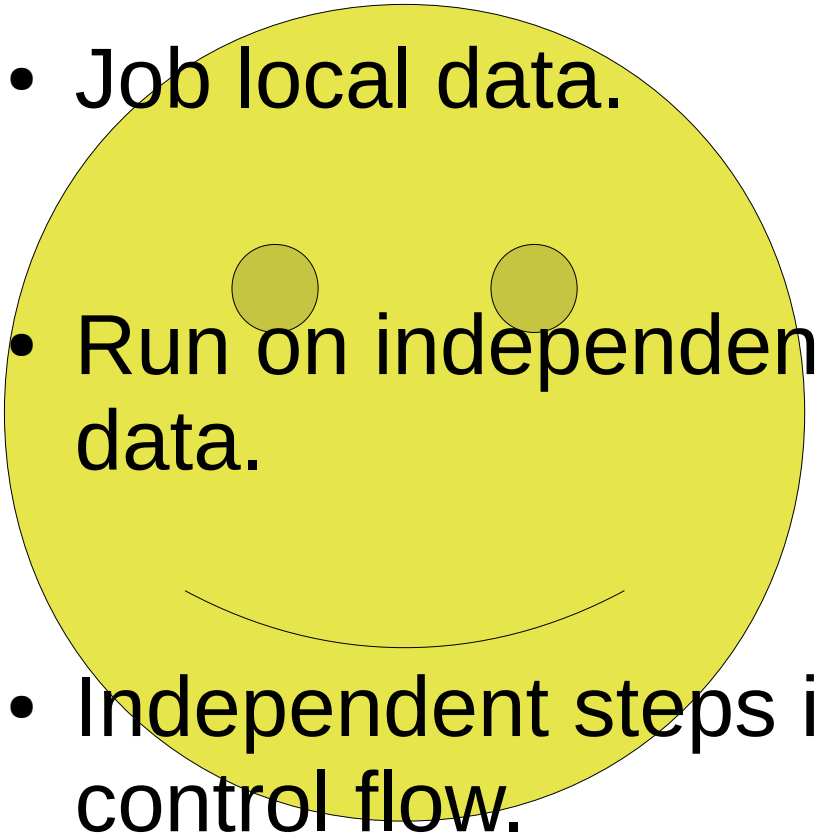

PVM
Parallel Virtual Machine


ERLANG


OpenMP
™

Role your own.

Algorithm properties



- Job local data.
- Run on independent data.
- Independent steps in control flow.

- Need for global data.
- Data dependencies.
- Control flow dependencies.

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.

Get involved!

*-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.

June, 25th 2009: Hadoop* Get Together in Berlin

- Torsten Curdt: “Data Legacy - the challenges of an evolving data warehouse.”
- Christoph M. Friedrich: “SCAIView - Lucene for Life Science Knowledge Discovery”
- Uri Boness, Bram Smeets: “Solr in production.”

newthinking store

Tucholskystr. 48

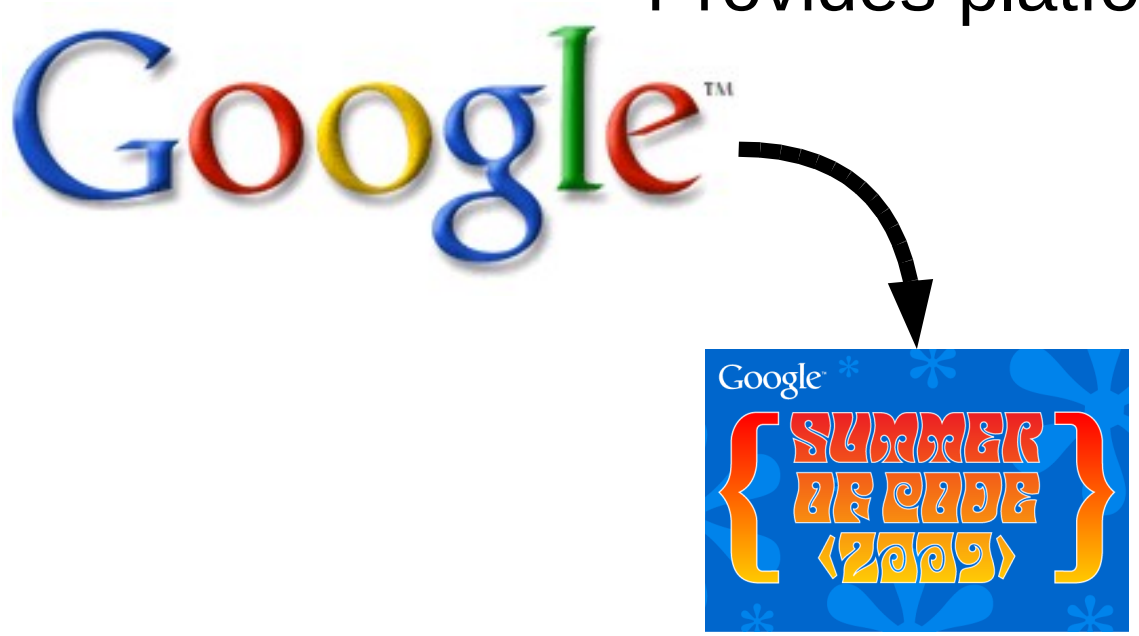
September, 29th 2009: Hadoop* Get Together in Berlin featuring a talk on UIMA by Thilo Götz.

* UIMA, Hbase, Lucene, Solr, katta, Mahout, CouchDB, pig, Hive, Cassandra, Cascading, JAQL, ... talks welcome as well.


Google™
SUMMER OF CODE
2009

Reasons: Promotion, recruiting, and others.

Provides platform and coordination.
Provides money.



Reasons: Paid, new developers.

Provides infrastructure.
Provides mentor.

Google™

Google Summer of Code 2009

Reasons: Paid open source job.
Gain experience in large projects.

Does all the work :)

GSoC Timeline

- February 8: Program announced. Life is good.
- March 9: Mentoring organizations can begin submitting applications to Google.
- March 23: Student application period opens.
- May 23: Students begin coding for their GSoC projects;
- July 6: Mentors and students can begin submitting mid-term evaluations.
- August 24: Final evaluation deadline;
- Google begins issuing student and mentoring organization payments.
- August 25: Final results of GSoC 2009 announced
- September 3: Students can begin submitting required code samples to Google
- October (date TBD): Mentor Summit at Google: Representatives from each successfully participating organization are invited to Google to greet, collaborate and code. Our mission for the

Ted Dunning: “[...] if I have a candidate at any level who has made **significant contributions** to a **major open source project**, I generally don't even drill much more on code hygiene issues.

The standards in most open source projects regarding **testing and continuous integration** are high enough that I don't have to worry about whether the applicant understands **how to code and how to code with others**.”

*-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.

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