Google App Engine for Java

Lars Vogel
http://www.vogella.de
Twitter: http://www.twitter.com/vogella

About Lars

Works for SAP as product owner of a SCRUM team.

Privately active in the Open Source Community

Eclipse committer, received the Eclipse Top Contributor community award 2010

Webmaster of http://www.vogella.de with more then 15 000 visitors per day

Lars Vogel
http://www.vogella.de
Twitter: @vogella

Why do I quality?

- Two popular tutorials in the internet about Google App Engine
  - http://www.vogella.de/articles/GoogleAppEngineJava/article.html
  - http://www.vogella.de/articles/GoogleAppEngine/article.html
- Published an article about robots on the App Engine in the (german) Eclipse Magazin
- Was technical editor for a Google App Engine Book (unpublished)

- What is cloud computing for me?
- What is the Google App Engine?
- Development for GAE with Java
- Persistence without SQL
- Tools - A fool with a tool is still a fool
- App Engine Services
- Testing on the cloud
- And of course coding

A typical system setup

OS +Application Server + DB

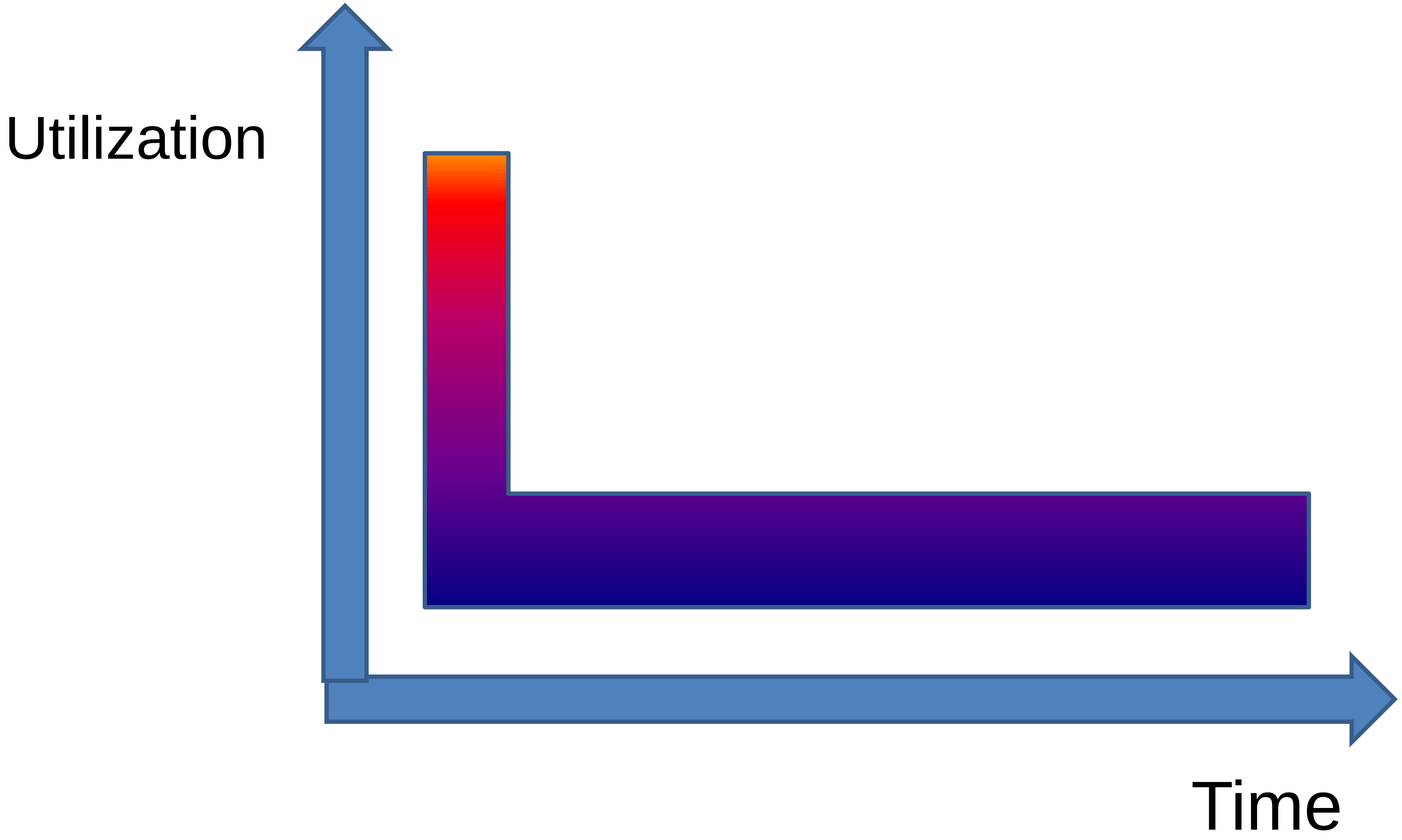
You need scaling...

Application Server Application Server Application Server

Database

So you are fine, right?...

Designed for peek consumption



Cloud computing tries to solve this

What is a cloud computing not?

Putting one or more computer in the internet is not necessary cloud computing

That is just a server in the internet

What is a cloud computing?

Cloud computing is Web-based processing, whereby shared resources, software, and information are provided to computers and other devices (such as smartphones) on demand over the Internet.

Cloud: ....some kind of abstracting from the hardware and providing resources on demand

What types of cloud computing do we have?

Infrastructure as a Service -> Amazon

Platform as a Service -> Google App Engine

Software as a service -> Salesforce.com, MyERP

Google App Engine

GAE allows you to host webapplications on the Google infrastructure.

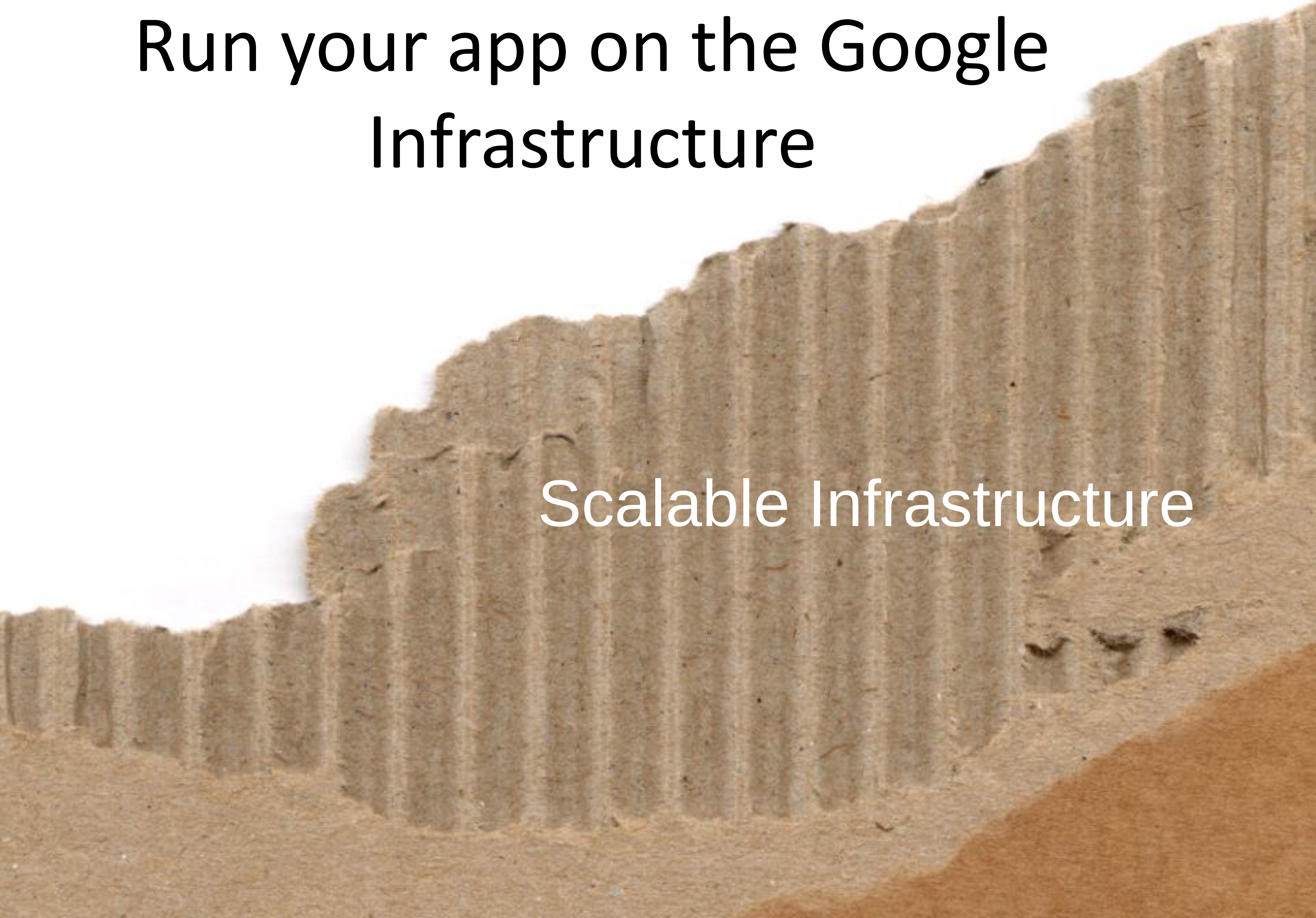
Difference to Amazons

Amazon provides virtual servers

App Engine provides an interface to program against

App Engine give no access to the underlying system / hardware

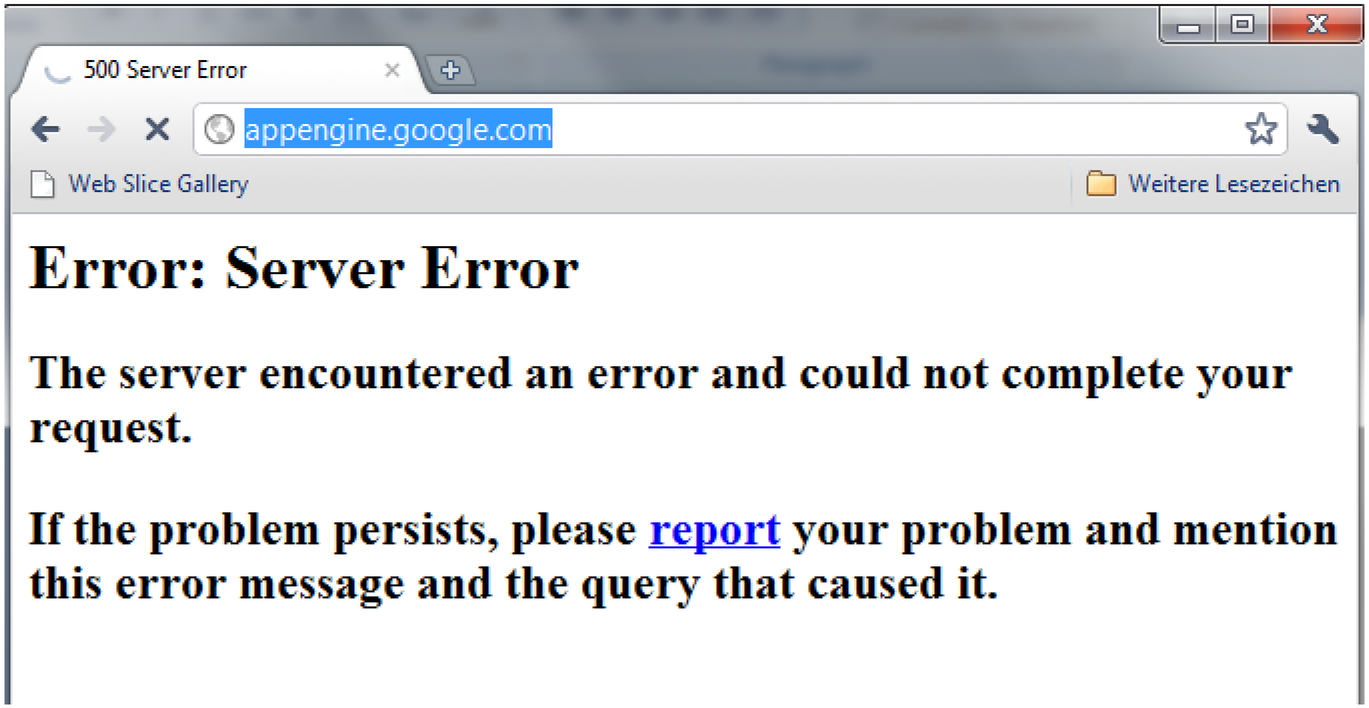
Run your app on the Google Infrastructure



Scalable Infrastructure

Google handles the infrastructure, e.g. hardware failures, security patches, OS upgrades

Sometimes there are issues



Error: Server Error

The server encountered an error and could not complete your request.

If the problem persists, please report your problem and mention this error message and the query that caused it.

and the usage of App Engine is free…. within limits

in any case you only pay for what your use

Google App Engine – Free Hosting

10 Applications per User

5 Million Pageviews are free per month.

Approx. 6.5 hours of CPU and 1 Gigabyte of inbound and outbound traffic.

100 hits per secs (non-billing) and 500 for billing enabled applications

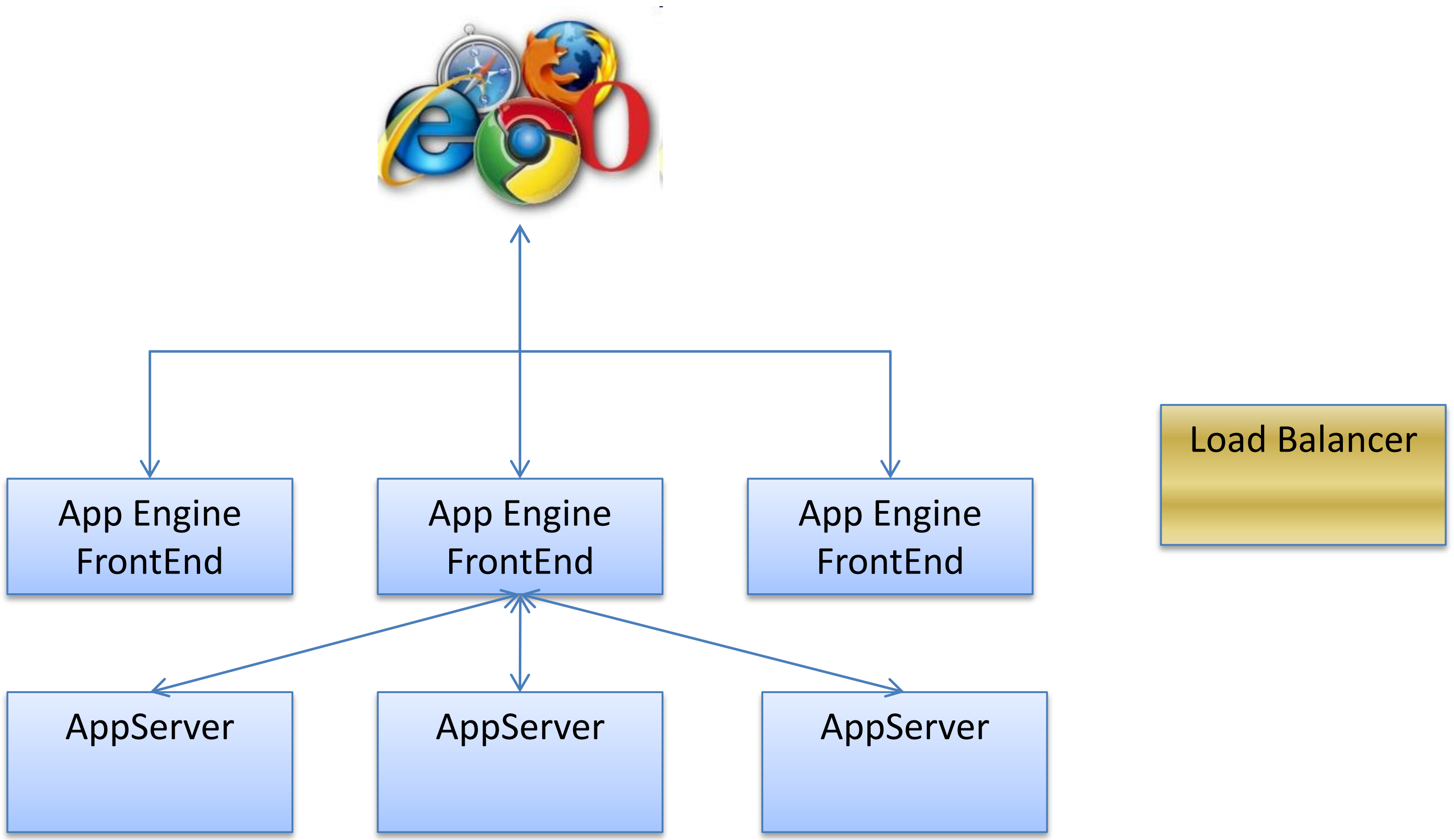
http://code.google.com/intl/en-EN/appengine/docs/billing.html

How does Google run thousands of thousands Java apps at the same time?

They don’t… applications which are not actively used will be frozen and saved to big table

→ Initial startup time

App Engine Architecture



So it scales but what about Platform as a service?

Writing Java Webs application is hard

CONFIGS
MYSQL
LOGS
CASH
APACHE
$$$$
$$$$
$$$$
MACHI
MONITORING
RSIONING
LINUX

from Guillaume Laforge and Patrick Chanezon http://www.slideshare.net/glaforge/google-app-engine-java-groovy-baby

To write a Java Web application

- Install / prepare the server
- Install / Configure the database
- Setup the webcontainer
- Develop your application
- Package your application into a WAR file
- Deploy it on your server

…developing a Java Web application from scratch is really, really slow due to the initial setup required

Google App Engine –

Programming Languages

Python

Java-isch

Scala

Groovy

JRuby

JPython

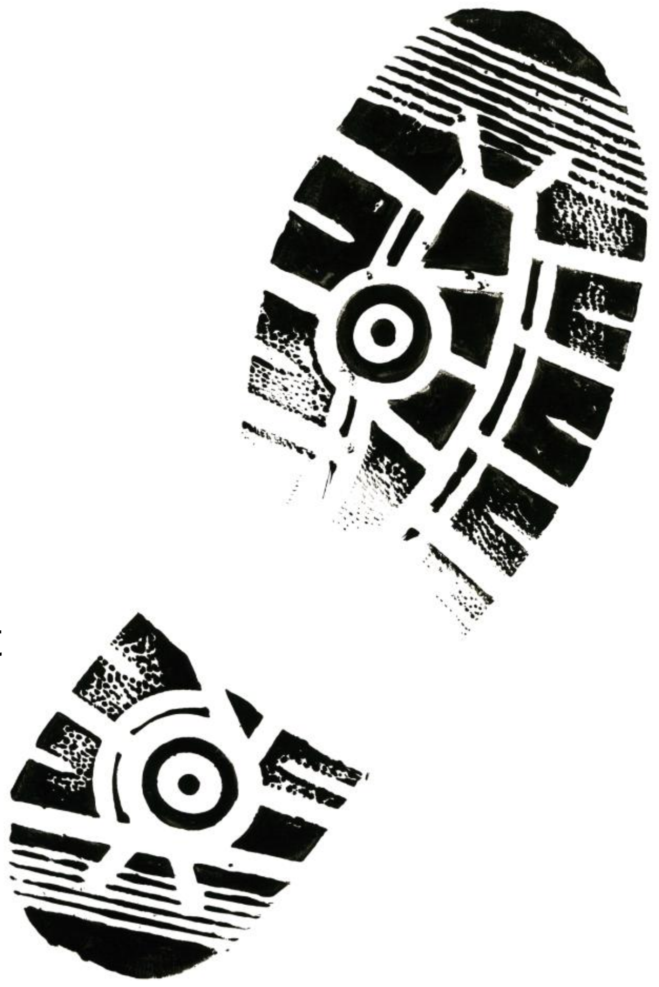
…

Still some issues with Grails....

Servlets and JSPs

A servlet is a Java class which answers a HTTP request within a web container.

JavaServer Pages (JSP) are files which contains HTML and Java code. The web container compiles the JSP into a servlet at the first time of accessing this JSP



Possible Web Frameworks on GAE

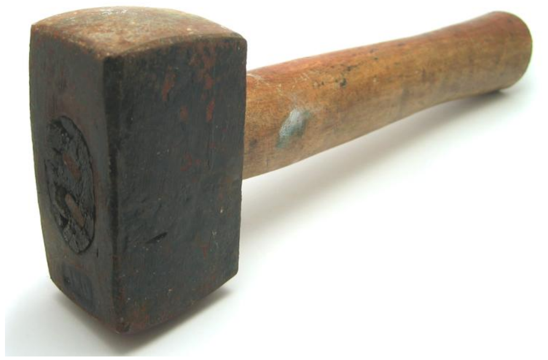
Basically all Java Web frameworks, e.g. JSP, Servlets based

GWT, JSF, Struts, Wicket

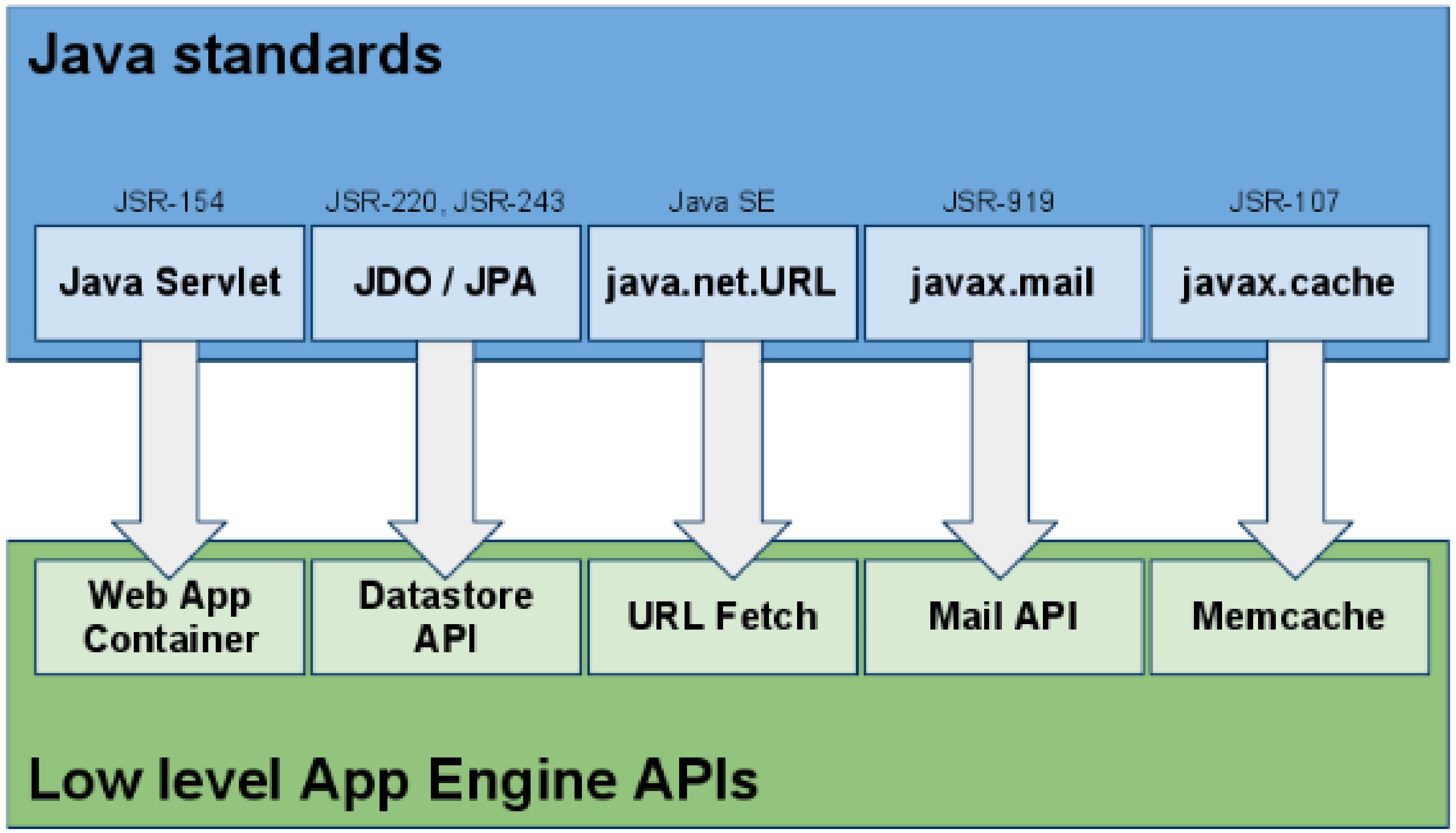
App Engine can be more then just a web application platform

Doesn‘t have to be a webapplication, can be a backend for Android or iPhones...

Can be used to do some serious number crunching



Develop according to Java standards



from Guillaume Laforge and Patrick Chanezon http://www.slideshare.net/glaforge/google-app-engine-java-groovy-baby

Session Support

Turned off by default

```
appengine-web.xml
<sessions-enabled>true</sessions-enabled>
```

Persistent and distributed

Configuration

web.xml

appengine-web.xml

- allows several versions of your app
- can define static content -> super fast
- can define system variables
- enable ssl and http sessions

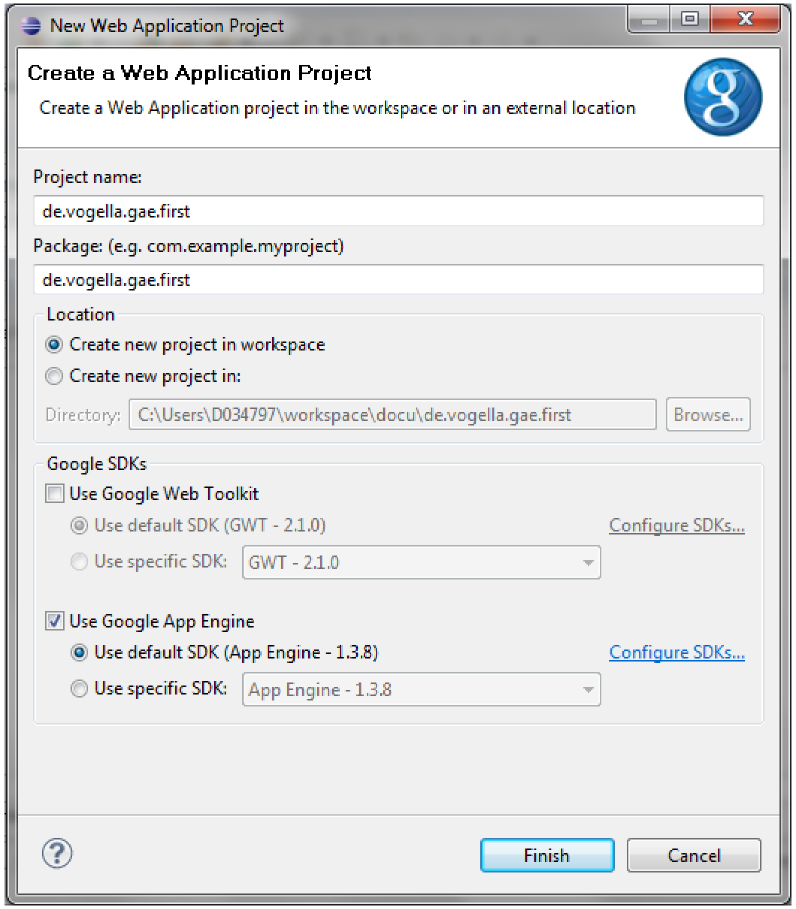
Development Tools

DevApp Server

Emulates the Google App Engine, its service and restrictions

Based on Jetty

Eclipse



New Web Application Project

Create a Web Application Project

Create a Web Application project in the workspace or in an external location

Project name: de.vogella.gae.first

Package: (e.g. com.example.myproject) de.vogella.gae.first

Location

[✓ Create new project in workspace]

Create new project in:

Directory: C:\Users\D034797\workspace\docu\de.vogella.gae.first

Google SDKs

Use Google Web Toolkit

Use default SDK (GWT - 2.1.0)

Use specific SDK: GWT - 2.1.0

[✓ Use Google App Engine]

[✓ Use default SDK (App Engine - 1.3.8)]

Use specific SDK: App Engine - 1.3.8

Deployment

Run your application on appplication-id.appspot.com or on your own domain

Command line or Eclipse based

Logging

java.util.logging.Logger

System.out and System.err are also logged

Performance

Be fast

You have to be fast:

- 30 seconds to respond!
- Otherwise com.google.apphosting.api.DeadlineExceededException

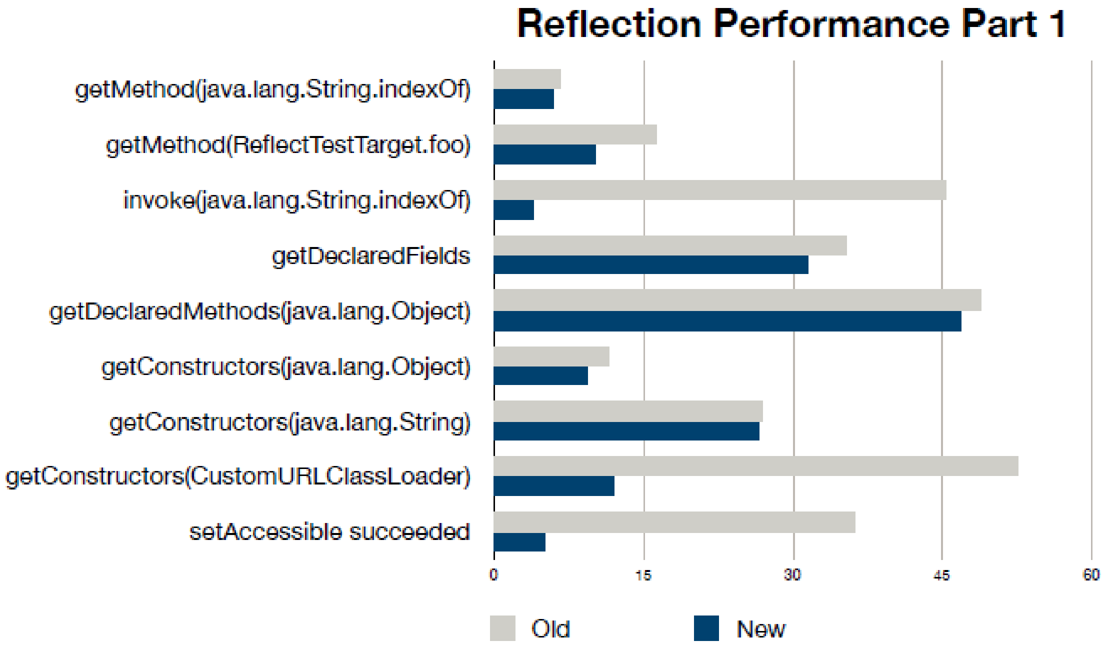
Google helps you that as its optimize its infrastructure all the time

Example for recent performance work

Byte code used to enhanced dynamically, now App Engine does this this statically.

JIT and GC improvements

Example Work on reflection



Data from Google I/O 2010

Limits

Limits

No native threads – but Tasks

No sockets

No write access to file system

No Native code

Not all standard Java classes available (white list)

Processes must finish without 30 secs (Tasks Chains…)

Compartibility is increasing

Whitelist is growing.

More and more libraries are supported

- Hessian (4.0.6)
- JDOM (1.1)
- Jersey (1.1.5)
- MyFaces (2.0.0)
- OpenAMF
- PureMVC
- Restlet (2.0M5)
- Struts 2
- Tapestry (5.1)
- VRaptor (3)
- Vaadin (6.1)
- Wicket
- ...

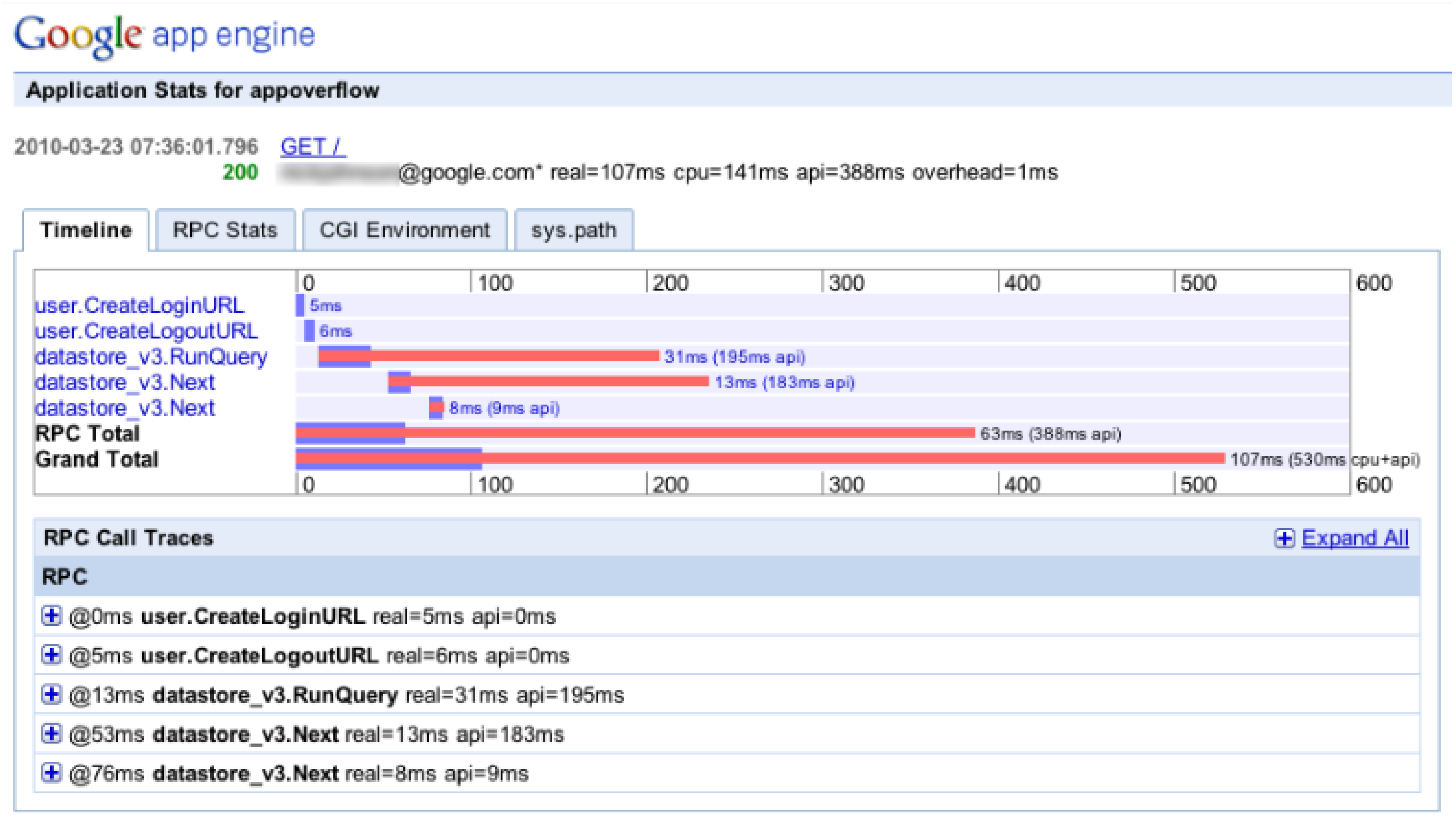
AppStat - Servlet filter in web.xml

```
<filter>
        <filter-name>appstats</filter-name>
        <filter-
class>com.google.appengine.tools.appstats.Appstats
Filter</filter-class>
        <init-param>
            <param-name>logMessage</param-name>
            <param-value>Appstats available:
/appstats/details?time={ID}</param-value>
        </init-param>
    </filter>
    <filter-mapping>
        <filter-name>appstats</filter-name>
        <url-pattern>/*</url-pattern>
    </filter-mapping>
```

Access AppStats

```
<servlet>
        <servlet-name>appstats</servlet-name>
        <servlet-
  class>com.google.appengine.tools.appstats.Appstats
  Servlet</servlet-class>
    </servlet>
    <servlet-mapping>
        <servlet-name>appstats</servlet-name>
        <url-pattern>/appstats/*</url-pattern>
    </servlet-mapping>
<security-constraint>
....
```

AppStat



Google app engine

Application Stats for appoverflow

2010-03-23 07:36:01.796 GET /
200 [redacted]@google.com* real=107ms cpu=141ms api=388ms overhead=1ms

Timeline | RPC Stats | CGI Environment | sys.path

0
100
200
300
400
500
600
user.CreateLoginURL 5ms
user.CreateLogoutURL 6ms
datastore_v3.RunQuery 31ms (195ms api)
datastore_v3.Next 13ms (183ms api)
datastore_v3.Next 8ms (9ms api)
RPC Total 63ms (388ms api)
Grand Total 107ms (530ms cpu+api)

RPC Call Traces Expand All

RPC

@0ms **user.CreateLoginURL** real=5ms api=0ms
@5ms **user.CreateLogoutURL** real=6ms api=0ms
@13ms **datastore_v3.RunQuery** real=31ms api=195ms
@53ms **datastore_v3.Next** real=13ms api=183ms
@76ms **datastore_v3.Next** real=8ms api=9ms

Source: http://googleappengine.blogspot.com/2010/03/easy-performance-profiling-with.html

Datastore

Storing data

Application can store data in memory, memcache or the **datastore**

Datastore

The datastore is App Engine's non-relational database

Uses internally Bigtable which is a key-value store

Read / Write Consistent and distributed

Datastore

The performance of a query depends only on the size of your result set. It does not depend on the total number of data entries.

Data Store

Based on BigTable

Sorted Map, no Joins

Schemaless

Transactional

Low-level APIs

JDO and JPA

Blobstore

Datastore via JDO

Data Classes

```
@PersistenceCapable
public class Employee {
   @Persistent
   private String name
}
```

PersistenceManager

```
PersistenceManager pm = PMF.get().getPersistenceManager();
pm.makePersistent(employee);
pm.close;
```

Retrieval and Queries

```
employee = pm.getObjectById(Employee.class, key);
Query query = pm.newQuery(Employee.class, “name = :name”);
List<Employee> results = (List<Employee>) query.execute("Smith")
```

Relations

```
@Persistent(mappedBy = "employee“)
private List<ContactInfo> contactInfoSets;
```

Indexes

- App Engine builds indexes for several simple queries by default.
- If you run a query that filters on multiple entity properties or orders results by multiple properties, you will need an index for that query.
- datastore-indexes.xml in WEB-INF/

Indexes

- Every query you run in the SDK automatically gets indexed.
- Stored in application's datastore-indexes.xml (Java)

Other storage possibilities

Blobstore

Memcache

Memcache

Cache data to avoid re-doing expensive operations

Fast

Developer can define a time-limit but App Engine may remove the data earlier -> do not rely on the availability

Blogstore

Store, Upload and serve large files (<=2GB per blob)

Read only

Application can read blobs as if there are local files.

Requires billing enabled

App Engine Services

Caching

Channel API (push to client, come

URL Fetching

Mail

Instant Messaging (XMPP)

Image Manipulation

User Management

Background Tasks

Map (no Reduce)

XMPP (Chat)

Testing…

Testing

Local Testing
 - equal functionality as the cloud
 - not equal in scale
- LocalServiceTestHelper(….services to be tested)

Cloud Testing
 - serving HTTP requests
 - scalable and fast
 - all limits apply which apply for standard application
-

Local Testing

Similar to standard Java testing
 - JUnit, TestNG, Selenium

Difference: Some classes require a certain App Engine setup, e.g. the datastore

AppEngine Testing API

Datastore: Initialize your local environment

```
private final
  LocalServiceTestHelper helper =
          new
  LocalServiceTestHelper(new
  LocalDatastoreServiceTestConfig
  ());
```

Use the local test Datastore

```
DatastoreService ds = DatastoreServiceFactory.getDatastoreService();
assertEquals(0, ds.prepare(new
    Query("yam")).countEntities(withLimit(10)));
ds.put(new Entity("yam"));
ds.put(new Entity("yam"));
assertEquals(2, ds.prepare(new
    Query("yam")).countEntities(withLimit(10)));
```

API.Proxy

Register your own class as proxy for the GAE API calls and interfere them.

Future

GAE 1.4

App Engine Future I

Running JVM (3 instances)

WarmupRequest via appengine-web.xml

The Channel API is now available for all users.

Task Queue has been officially released,

Deadline for Task Queue and Cron requests has been raised to 10 minutes.

Source: http://groups.google.com/group/google-appengine/browse_thread/thread/0f90ef8dda3b8400

App Engine Future II

URL Fetch allowed response size has been increased, up to 32 MB. Request size is still limited to 1 MB.

Added a low-level AysncDatastoreService for making calls to the datastore
asynchronously.

The whitelist has been updated to include all classes from javax.xml.soap.

Source: http://groups.google.com/group/google-appengine/browse_thread/thread/0f90ef8dda3b8400

MySQL planned

Source: http://groups.google.com/group/google-appengine/browse_thread/thread/0f90ef8dda3b8400

Photo credits

Please add http://www.sxc.hu/photo/ to the number if not specified

- Cloud 1309726
- Agenda 187747
- Guy pulling the white wall http://www.sxc.hu/photo/702367
- Balloon http://www.sxc.hu/photo/566242
- Critical guy http://www.sxc.hu/photo/1173019
- Question mark in the box 1084632 and 1084633
- Hammer 604247
- Server 175983
- Turtle 1198861
- Thinking Guy 130484
- Picked Fence 695054
- Performance 765733
- Tools 1197009
- Open Door http://www.sxc.hu/photo/1228296
- Paper Chart http://www.sxc.hu/photo/565681
- Binary http://www.sxc.hu/photo/1072645
- Footprint http://www.sxc.hu/photo/442696
- Old Computer http://www.sxc.hu/photo/1028528
- Carton http://www.sxc.hu/photo/502161
- Eye http://www.sxc.hu/photo/933394
- Guitar playing man http://www.sxc.hu/photo/ 894247
- Brown bag 250762
- Future 1139530
- Guy reading book 406547
- Money House 1097251

Thank you

For further questions:

Lars.Vogel@gmail.com

http://www.vogella.de

http://www.twitter.com/vogella

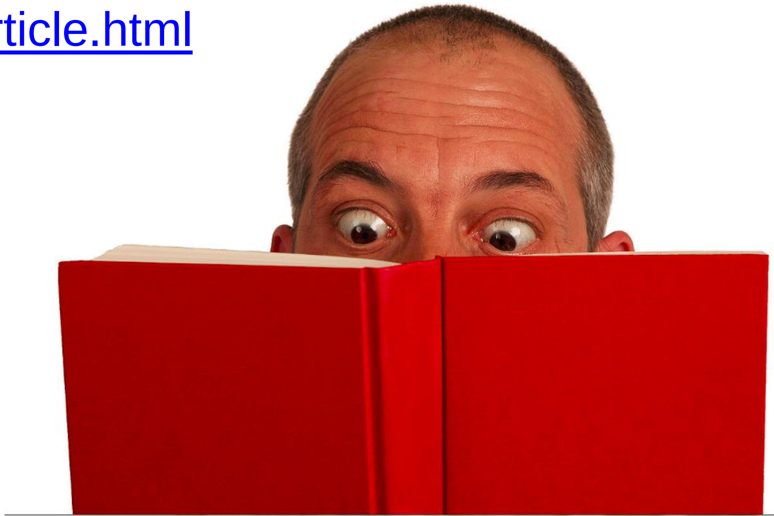
Hands-on Tutorial

The following tutorials give an intro to GAE/J using servlets and JSP‘s

http://www.vogella.de/articles/GoogleAppEngineJava/article.html

http://code.google.com/intl/de-DE/appengine/docs/java/gettingstarted/

http://www.vogella.de/articles/GoogleAppEngine/article.html



License & Acknowledgements

Google Cloud for businesses

99.9 % SLA

pay 8 Euros per user up to a maximum of 1000 Euro (==125 Users)

Hosted SQL

SSL (Secure Sockets Layer) for the domain

Channel Service

Server:

Send messages to the participants (similar to websockets)

Bi-directional

Build on Gmail Chat client (Google Talk)

Sends String data

Client

JavaScript API

Task Queue Server

Perform work in the background asynchronously

Up to 50 requests per seconds

AppStats

Easy profiling tool for Google App Engine

Install a servlet filter and browser all requests and see their performance

Tracks the HTTP request and reponse and the stack trace

http://code.google.com/appengine/docs/java/tools/appstats.html

User Management

```
UserServiceFactory.getUserService();
userService.getCurrentUser();
```

Batch Processing & Data Upload

Task Queue

Can do work outsite the user request

Still 30 seconds limit

Task chaining to overcome the limit -> As soon as your time is up, start a new task with the rest of the data.

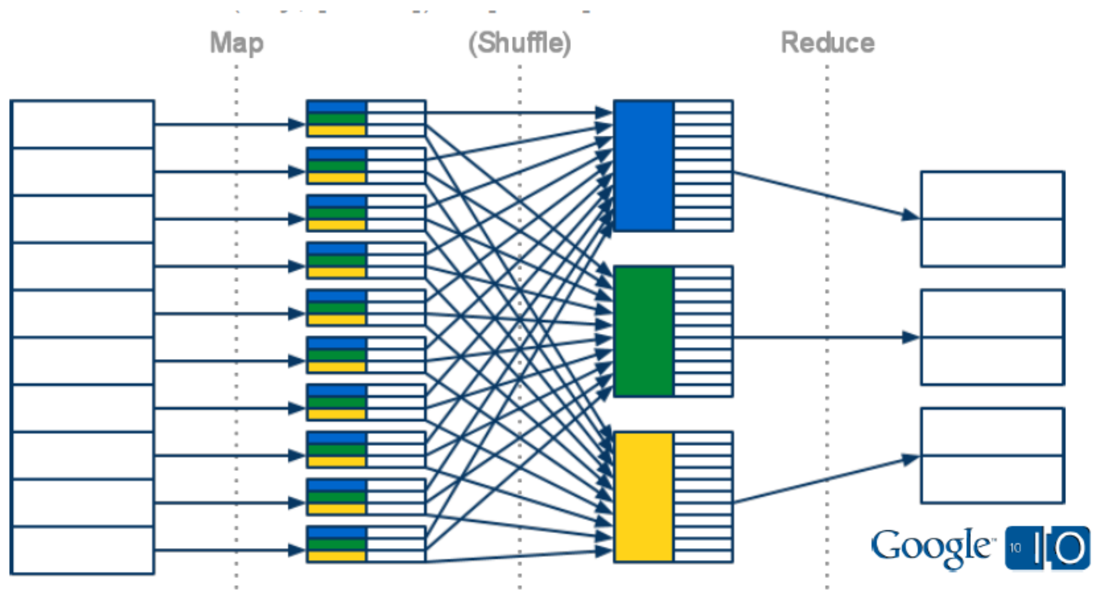
If task fail the Task Queue will retry the task

MapReduce

Defines two functions:

Map(entity) -> [(key, value)]
reduce (key, [value]) -> [value]

Sounds silly, but allows lots of things, e.g. DB schema migration, report generation



Batch Processing in App Engine - Map Reduce

Iterative over Blob data

Implemented using task queue chaining

Uses datastore for state storing and communication

Based on idempotence, the same input should create the same output to be able to re-cover from errors

Map Reduce

Performing processing of mass data on App Engine can be hard due to the time (30 seconds) and volume limits.

Before MapReduce you had to split your data and process them by toaching an URL

Allow to perform mass calculation on data.

Rate limiting to support overall performance and protect user from extreme costs -> rebuilding the MapReduce functionality from Google

MapReduce is build on TaskQueue

Currently only Map available -> Open Source

Iterative over Blob data

Bulk Data Upload

Bulk loader tool can upload and download data from and to your application datastore.

Bulk loader tool ist Python based (appcfg.py)

Requires a request handler registered on the server side (for Java com.google.apphosting.utils.remoteapi.RemoteApiServlet)

http://code.google.com/appengine/docs/python/tools/uploadingdata.html

Alternatively you can upload to the blobstore and use a Mapper to put it into the datastore.

Command Line Tool

AppCfg for

- upload application
- update database configuration
- Download the app logs data to analyse it

Deployment

AppCfg for

- upload application
- update database configuration
- Download the app logs data to analyse it