

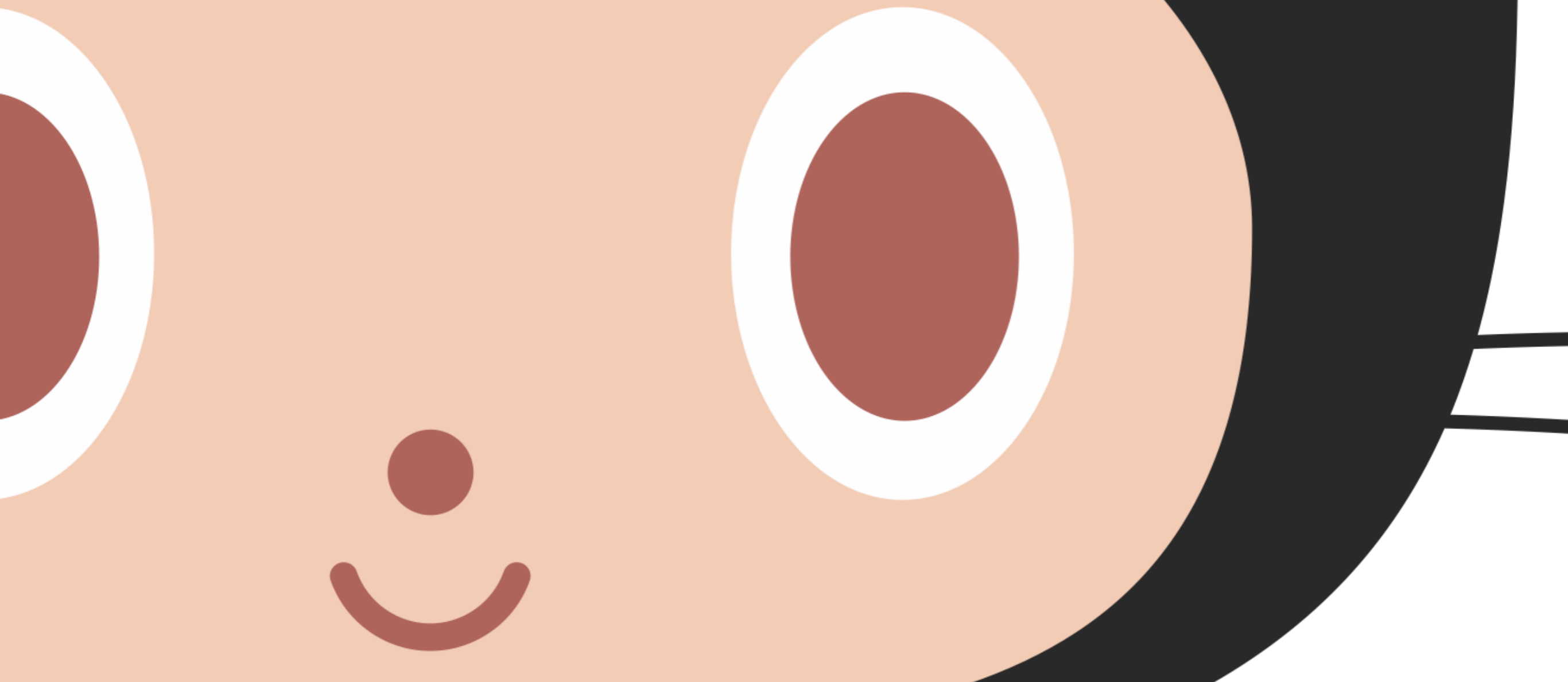
github:pages  
on

riak &

webmachine

[github.com/jnewland](https://github.com/jnewland)

@jnewland



github:pages  
on

riak &

webmachine

**what**

**is**

**pages?**

simple  
static  
file  
hosting\*

jeKyll

transform your text into a monster

[github.com/user/repo/gh-pages](https://github.com/user/repo/gh-pages)



<http://user.github.com/repo>

# *Puppet-lint*

Check that your Puppet manifest conform to the style guide

[</> fork Puppet-lint on GitHub](#)

 Download

 Download

## Usage

```
1 $ gem install puppet-lint
2 $ puppet-lint path/to/your/manifest.pp
```

## Things it tests

### Spacing, Indentation & Whitespace

- Must use two-space soft tabs.
- Must not use literal tab characters.



**github.com/user/user.github.com**



**http://user.github.com**

## Tom Preston-Werner [home](#)

### Blog Posts

- 22 Nov 2011 » [Open Source \(Almost\) Everything](#)
- 03 May 2011 » [Rejected Bio from The Setup](#)
- 29 Mar 2011 » [Ten Lessons from GitHub's First Year](#)
- 11 Dec 2010 » [Designer, Architect, Developer](#)
- 18 Oct 2010 » [Optimize for Happiness](#)
- 23 Aug 2010 » [Readme Driven Development](#)
- 11 May 2010 » [TomDoc - Reasonable Ruby Documentation](#)
- 19 May 2009 » [The Git Parable](#)
- 17 Nov 2008 » [Blogging Like a Hacker](#)
- 03 Nov 2008 » [How to Meet Your Next Cofounder](#)
- 27 Oct 2008 » [Looking back on Selling Gravatar to Automattic](#)
- 18 Oct 2008 » [How I Turned Down \\$300,000 from Microsoft to go Full-Time on GitHub](#)

The screenshot shows the GitHub repository page for 'mojombo /mojombo.github.com'. At the top, there are navigation links for 'Admin', 'Unwatch', 'Your Fork', '321' (watchers), and '143' (forks). Below this is a tabbed interface with 'Code' selected. The repository description reads: 'Jekyll source for my personal blog. — Read more <http://tom.preston-werner.com>'. There are buttons for 'Clone in Mac', 'ZIP', 'HTTP', and 'Git Read-Only' with the URL 'https://github.com/mojombo/mojombo.github.com.g'. A 'Read-Only access' button is also present. Below the repository information, there are tabs for 'Files', 'Commits', 'Branches (1)', 'Tags', and 'Downloads'. A search bar for source code is on the right. The main content area shows the latest commit to the 'master' branch: 'Make PGP file txt format.' by 'mojombo' 18 days ago, with commit ID '1ce6175042' and 1 comment. Below the commit, there is a table of files:

name	age	message	history
_layouts	2 years ago	add clicktale [mojombo]	
_posts	a month ago	Designer, Architect, Developer. [mojombo]	
css	2 years ago	RDD post. [mojombo]	

twitter / [twitter.github.com](https://twitter.github.com) Admin Watch Fork 60 29

Code Network Pull Requests 1 Issues 2 Wiki 2 Graphs

A listing of open source efforts at Twitter on GitHub  
<http://twitter.github.com/>

Clone in Mac ZIP HTTP Git Read-Only <https://github.com/twitter/twitter.github.com.g> Read-Only access

branch: master Files Commits Branches 1 Tags Downloads Search source code...

Latest commit to the master branch

Use simpler clearfix. Fixes #2

mjackson authored 2 months ago commit 03dae5772d

[twitter.github.com /](#)

name	age	message	history
assets	2 months ago	Use simpler clearfix. Fixes #2 [mjackson]	
.gitignore	2 months ago	Initial commit [mjackson]	
index.html	2 months ago	Updated some descriptions on github.com [mjackson]	

**open source**

**Twitter is built on open source software.**

Want to help? [Join the Flock](#)  
Visit [dev.twitter.com](http://dev.twitter.com)  
Logos and other goodies

**Statistics**

68 public repos  
133 members  
opensource@twitter.com

**Recently updated**

[sbt-scrooge](#) Mar 27, 2012 · 385 watchers · 56 forks  
[twitter-text-js](#) Mar 27, 2012 · 385 watchers · 56 forks  
[standard-project](#) Mar 27, 2012 · 89 watchers · 17 forks

**bootstrap**

JavaScript

An HTML, CSS, and JS toolkit designed to kickstart development of webapps and sites

**hogan.js**

JavaScript

A compiler for the Mustache templating language

**bootstrap-server**

JavaScript

The node server that powers the bootstrap customize page

**sbt-package-dist**

Scala

sbt 11 plugin codifying best practices for building, packaging, and publishing

**effectivescala**

Shell

Twitter's Effective Scala Guide

**scalding**

Scala

A Scala API for Cascading

**cassovary**

Scala

Cassovary is a simple big graph processing library for the JVM from Twitter

# custom domain support

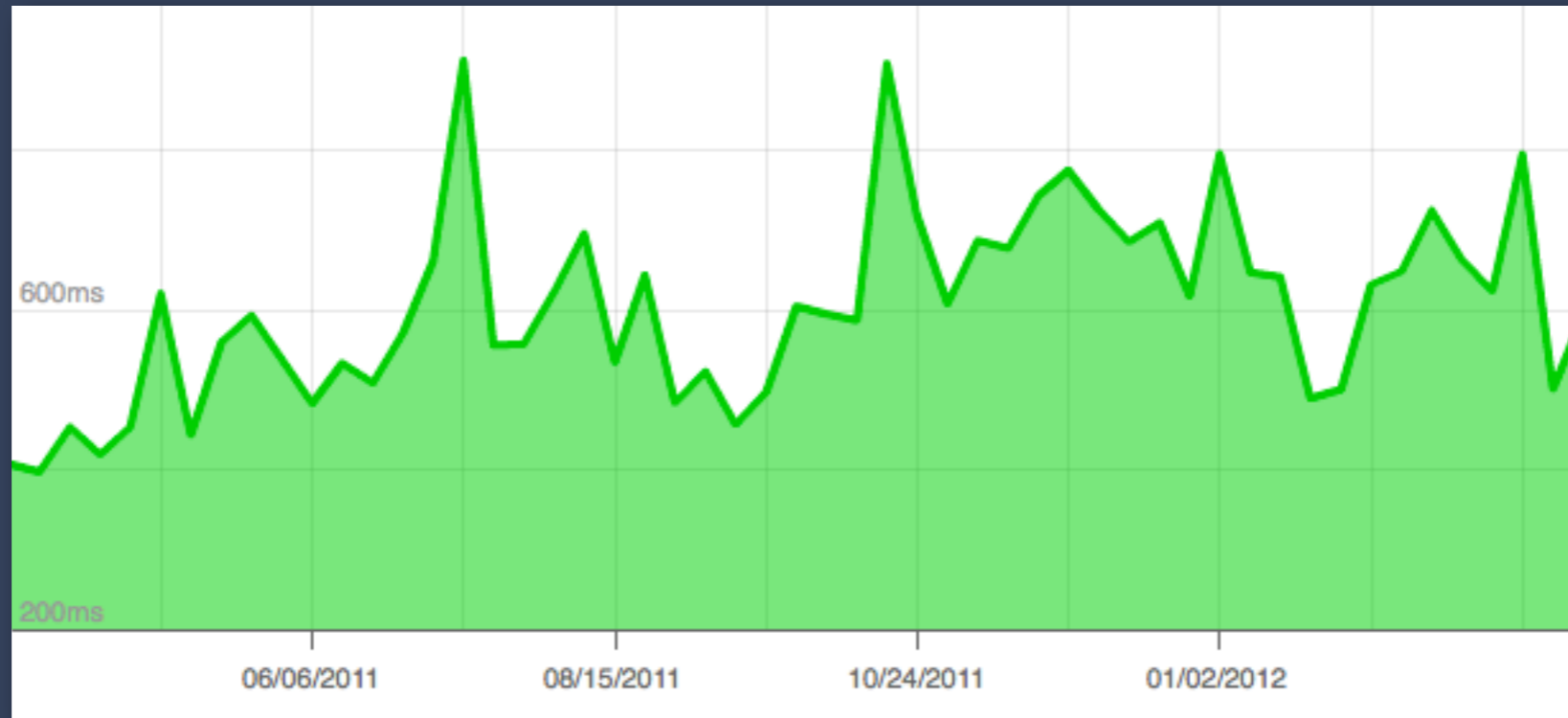
```
$ cat CNAME  
example.com
```

**what**

**is wrong with**

**pages?**

# SLOW



**1 node**

**ext3**

**gigs and gigs of HTML**

10 bound



# fscking downtime



**github** GitHub

Pages is down. We're working on it now, should be back up shortly!

November 8, 2011 26 retweets #



**obfuscurity** Jason Dixon

@github Any ETA on resolution?

November 8, 2011 #



**jnewland** Jesse Newland

@obfuscurity fsck's at 89.1%

November 8, 2011 #



**obfuscurity** Jason Dixon

@jnewland Thanks dude.

November 8, 2011 #



**jnewland** Jesse Newland

@obfuscurity fscking fsck

November 8, 2011 #

**not HA**

**LOTS**

**THROW**

**TO**

**IT ALL**

**FIX**

**OUT**

**what**

**can power**

**pages 2.0?**

1. grab content from git
2. run through jekyll
3. write somewhere
4. serve over HTTP

**what  
currently powers  
pages?**

**1. ruby**

**2. ruby**

**3. ext3**

**4. nginx**

```
$ wc -l pages_map.conf  
57615 pages_map.conf
```



1. ruby

2. ruby

3. riak\_kv

4. riak\_kv\*

 **almost**

**read-only**

**transactional builds**

**CNAMEs & redirects**

**index.html fallback**

**custom 404.html per-repo**

**1. ruby**

**2. ruby**

**3. riak\_kv**

**4. webmachine resource**

**why  
not  
x?**

**squid**  
**+**  
**nginx/apache**  
**+**  
**filesystem**

**should work fine**

**what happens when you need  $N > 1$ ?**

**just shard 'em**

**how do you populate new partitions?**

**just rsync stuff around**

**building a  
distributed  
system**

**AASS**

**FIRST**

**remember, I do ops**

**low maintenance, resilient**

**systems make lazy**

**sysadmins love you**





riak

schema

design

2 buckets

hosts pages

# hosts

**key:** HTTP Host Header

**value:** redirect or repo/sha map

**use:** data key prefix lookup

**index:** user\_id

# hosts

jnewland.github.com

```
{  
  "repos": {  
    "jnewland.github.com": "deadbeef",  
    "pages": "beadfeed"  
  }  
}
```

# pages

**key:** sha/URI

**value:** HTML / other data

**use:** data storage

**index:** repo\_id

# pages

deadbeef/index.html

```
<!DOCTYPE html>
<html xmlns="http://www.w3.org/1999/
xhtml" xml:lang="en" lang="en-us">
<head>
  <meta http-equiv="content-type"
content="text/html; charset=utf-8" />
  <title>Jesse Newland</title>
  <meta name="author" content="Jesse
```

**how does  
data get in?**



**riak-ruby-client**

**read files from disk**

**write to riak**

**update sites object**

**gc old builds**

**how does  
data get out?**

**curl jnewland.github.com/raptor.gif**

**GET /riak/hosts/jnewland.github.com**

**GET /riak/pages/deadbeef/raptor.gif**

**curl jnewland.github.com/pages/flow.png**

**GET /riak/hosts/jnewland.github.com**

**GET /riak/pages/beadfeed/flow.png**

**what**

**about**

**CNAMEs?**

# hosts

example.github.com

```
{  
  "redirect": "example.com"  
}
```

# hosts

example.com

```
{  
  "repos": {  
    "example.com": "deadbeef"  
  }  
}
```

**curl example.github.com/raptor.gif**

**GET /riak/hosts/example.github.com**

**301 example.com/raptor.gif**

**GET /riak/hosts/example.com**

**GET /riak/pages/deadbeef/raptor.gif**



**pages\_wm\_resource.erl**

**webmachine**

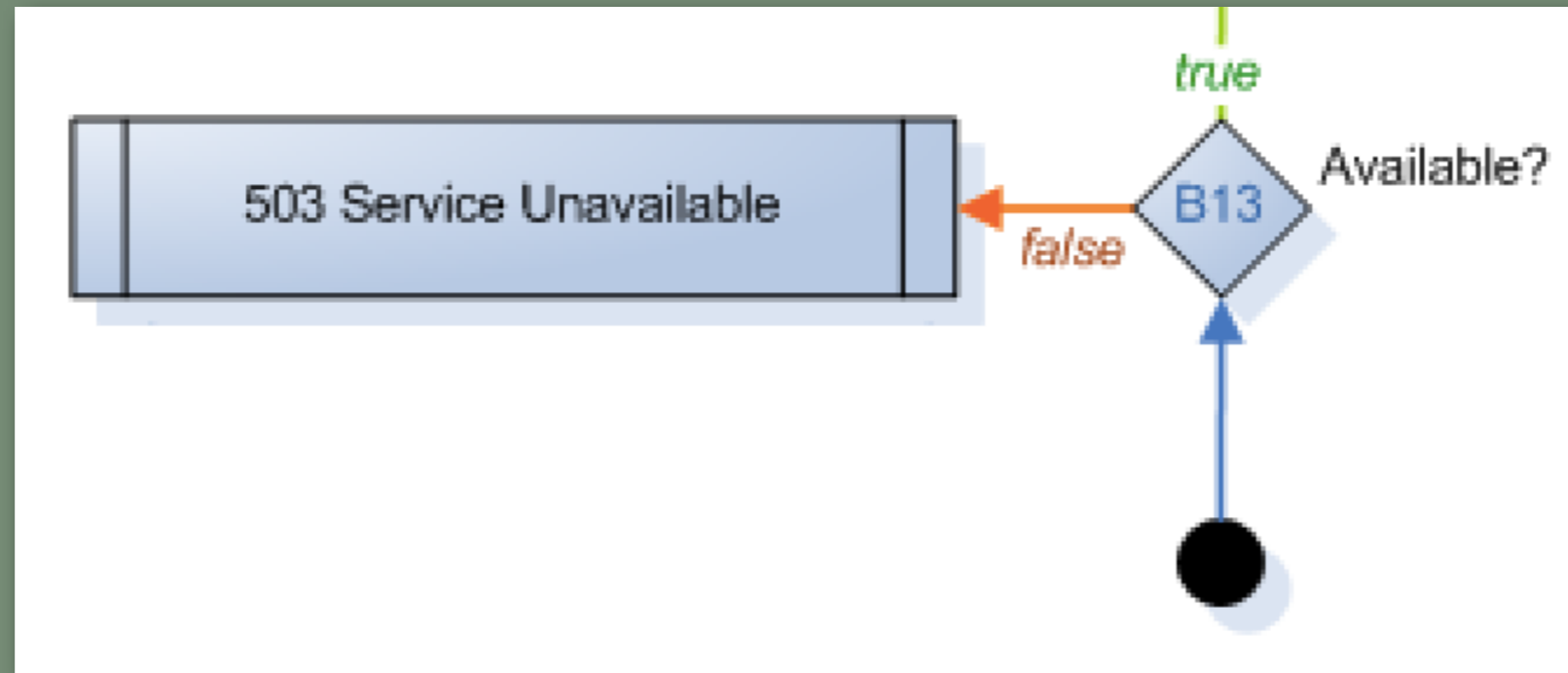
**is**

**SO DAMN COOL**



```
%% webmachine resource exports  
-export([  
    init/1,  
    service_available/2,  
    malformed_request/2,  
    content_types_provided/2,  
    resource_exists/2,  
    previously_existed/2,  
    moved_permanently/2,  
    last_modified/2,  
    generate_etag/2,  
    produce_doc_body/2  
]).
```

# service\_available/2



grab local riak client  
check app config var

```

service_available(RD, Ctx=#ctx{riak=RiakProps, req_id=ReqId}) ->
  IdRD = wrq:set_resp_header("X-Request-Id", ReqId, RD),
  BrandedRD = wrq:set_resp_header(
    "X-GitHub-Pages-Version",
    release_handler_util:app_version(pages),
    IdRD),
case application:get_env(pages, disabled) of
  {ok, true} ->
    {false, BrandedRD, Ctx};
  _ ->
    case riak_kv_wm_utils:get_riak_client(
      RiakProps,
      riak_kv_wm_utils:get_client_id(RD)) of
      {ok, C} ->
        {true, BrandedRD, Ctx#ctx{client=C}};
      _Error ->
        {false, BrandedRD, Ctx}
    end
end.

```

# malformed\_request/2

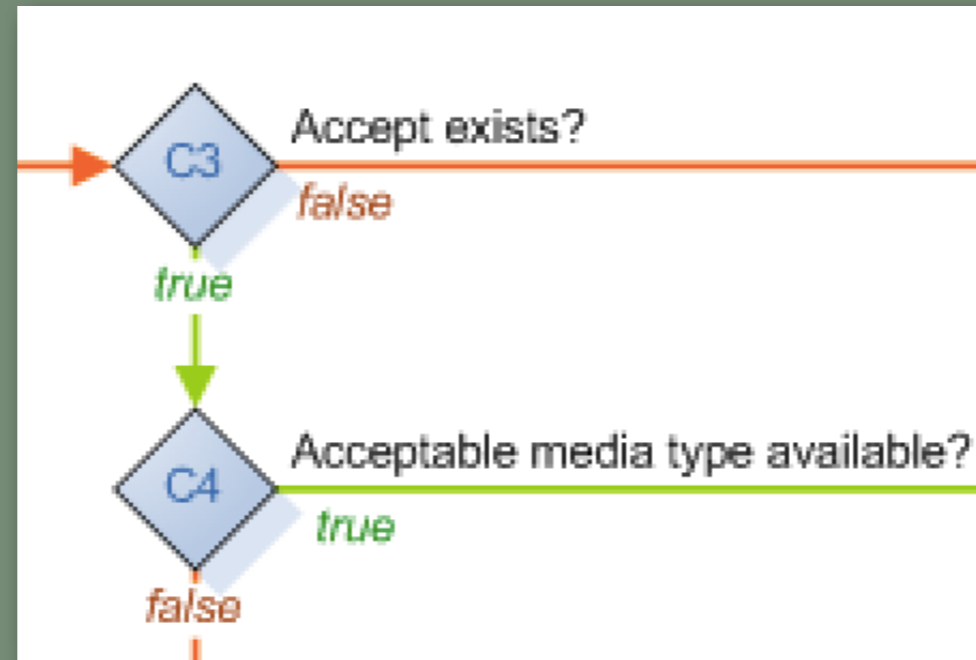


parse Host header, URI

```
malformed_request(RD, Ctx) ->
  try
    Host = wrq:get_req_header("Host", RD),
    HostWithoutPort = re:replace(
      Host,
      "\:.*",
      "",
      [{return,list}]),
    Tokens = [
      riak_kv_wm_utils:maybe_decode_uri(RD, X) ||
      X <- wrq:path_tokens(RD)],
    ParsedCtx = Ctx#ctx{tokens=Tokens,host=HostWithoutPort},
    {false, RD, ParsedCtx}
  catch
    Exception:Reason ->
      log_error({exception, Exception, Reason}, RD, Ctx),
      {true, RD, Ctx}
  end.
```



# content\_types\_provided/2

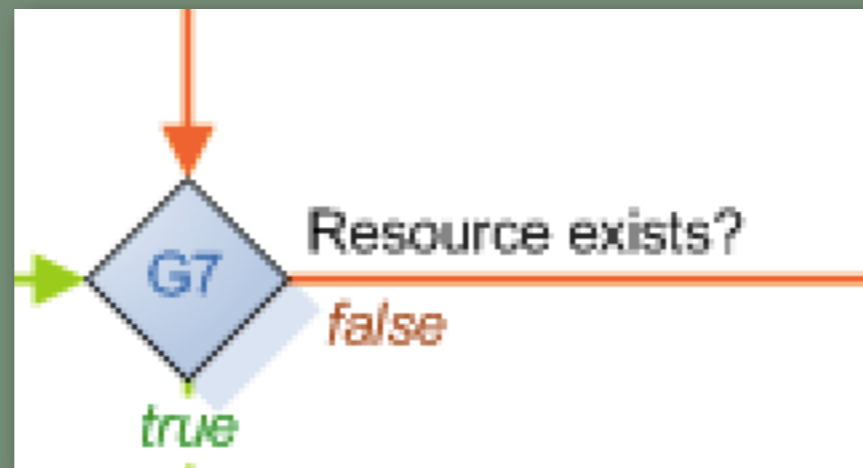


guess mime type from path  
assume text/html for / URIs

```
content_types_provided(RD, Ctx) ->
  Filename = lists:last(Ctx#ctx.tokens),
  Extension = filename:extension(Filename),
  case mochiweb_mime:from_extension(Extension) of
    undefined ->
      [{{"text/html", produce_doc_body}], RD, Ctx};
  Mime ->
    [{Mime, produce_doc_body}], RD, Ctx}

end.
```

# resource\_exists/2

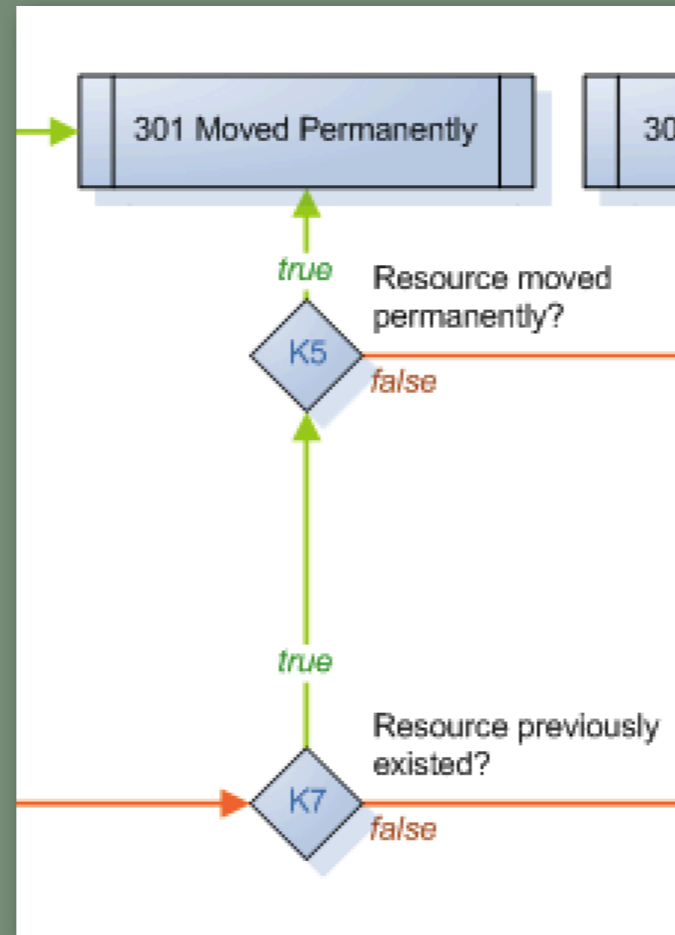


hit hosts bucket ( $r=1$ )

stash redirect or sha

404 if no hosts data

```
resource_exists(RD, Ctx) ->
  RedirectOrSha = redirect_or_sha(Ctx),
case RedirectOrSha of
  {redirect, Redirect} ->
    {true, RD, Ctx#ctx{redirect={redirect, Redirect}}};
  {sha, Sha} ->
    page_data_exists(RD, Ctx#ctx{sha={sha, Sha}});
  _ ->
    {false, RD, Ctx}
end.
```



previously\_existed/2

moved\_permanently/2

```
previously_existed(RD, Ctx) ->  
  case Ctx#ctx.redirect of  
    {redirect, _} ->  
      {true, RD, Ctx};  
    _ ->  
      {false, RD, Ctx}  
  end.
```

```
moved_permanently(RD, Ctx) ->  
  case Ctx#ctx.redirect of  
    {redirect, RedirectHost} ->  
      MovedURI = list_join(lists:append(  
        [RedirectHost],  
        Ctx#ctx.tokens),  
        "/" ),  
      {{true}, MovedURI, RD, Ctx};  
    _ ->  
      {false, RD, Ctx}  
  end.
```

**page\_data\_exists/2**

**hit pages bucket (r=1)**

**fallback for index.html**

**fallback for 404.html**

**curl foo.github.com/**

**GET /riak/hosts/foo.github.com**

**GET /riak/pages/fofofofo/**

**GET /riak/pages/fofofofo/index.html**

**GET /riak/pages/fofofofo/index.htm**

**GET /riak/pages/fofofofo/index.xhtml**

**GET /riak/pages/fofofofo/index.xml**

**GET /riak/pages/fofofofo/404.html**



**simple**

**< 300 lines of erlang**

# nginx proxies to / /pages

```
{webmachine, [{
  dispatch_list, [{
    %% riak_kv stuff

    [{"pages", '*'], pages_wm_resource, []},
    [{"pages"}, pages_wm_resource, []] ]}]
}]}
```

**remember, I do ops**

**one system service**

**data store and api**

**predictable performance**

**busy ops best friend**

**what's  
next**

**metrics with folsom**

**graphite / gaug.es**

**logging with lager**

**HTTP caching**

**???**

**private beta soon**

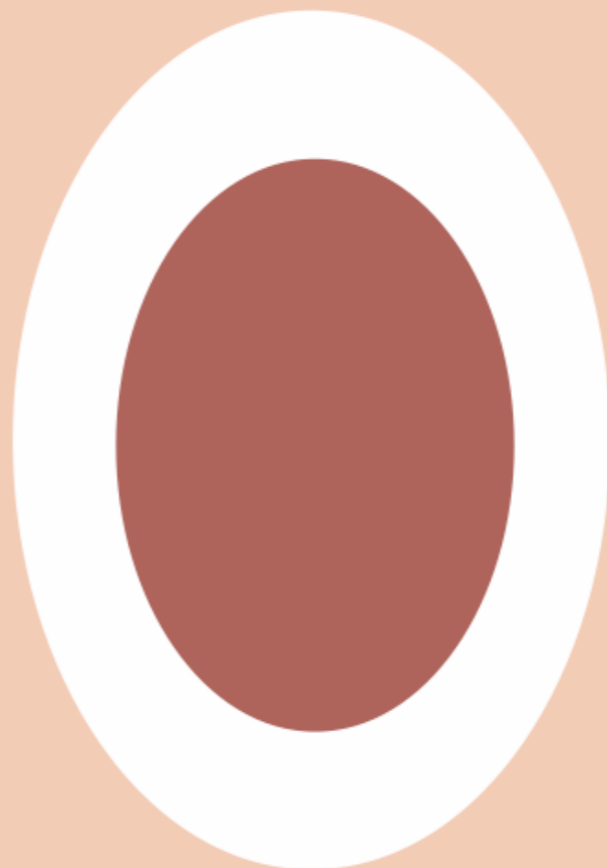
**turn on / off with DNS**

**repo access**

**@jnewland**

**work with me**

**erlang  
ruby  
ops  
c**



thanks

