

Anatomy of a large Django site

Andy McKay **Mozilla**



Vancouver

mozilla

**Python
Zope and Plone
...now at Mozilla**



mozilla

Using Django

<http://www.djangoproject.com>



1. About the site
2. Performance
3. Localisation
4. Reuse

1. About the site

[Register or Log in](#)[Other Applications ▾](#)

ADD-ONS

2,326,721,130 add-ons downloaded

Categories

[Alerts & Updates](#)[Appearance](#)[Bookmarks](#)[Download Management](#)[Feeds, News & Blogging](#)[Games & Entertainment](#)[Language Support](#)[Photos, Music & Videos](#)[Privacy & Security](#)[Shopping](#)[Social & Communication](#)[Tabs](#)[Web Development](#)[Other](#)[Collections](#)[Dictionaries & Language Packs](#)[Personas](#)[Plugins](#)[Search Tools](#)[Themes](#)

Discover great add-ons for everyone



Echofon for Twitter

Echofon (formerly TwitterFox) notifies you of your friends' tweets on Twitter. ...

[Continue to Download →](#)
Featured

ColorfulTabs

The most beautiful yet the simplest add-on that makes a strong colorful appeal. ...

[+ Add to Firefox](#)
Featured

Flagfox

Displays a country flag depicting the location of the current website's server ...

[+ Add to Firefox](#)
Featured

Like these? Find more add-ons in the [Rock Your Firefox - September 2010](#) collection.

[Introduction](#)[Firefox 4 Beta](#)[Rock Your Firefox](#)[Privacy](#)[Web Development](#)[Travel](#)

Browse Add-ons

[Featured](#)[Popular](#)[Recently Added](#)[Recently Updated](#)

Amplify

by Clipmarks[+ Add to Firefox](#)
Featured

Not available for Firefox 4.0b11

Use Amplify to clip, share & spark conversation about excerpts

Collections

Collections are groups of related add-ons that anyone can create and share.



All code is open:
<https://github.com/jbalogh/zamboni>

All* bugs are open:
<https://bugzilla.mozilla.org>

**Convert from
CakePHP (remora)
to Django (zamboni)**



Changing one URL at a time from CakePHP to Django

General trend to move
away from PHP and do
more Python and Django

How large?

250k+ addons

150 mn views month

500 mn API hits day

Lines of code

PHP 40k

Python 18k

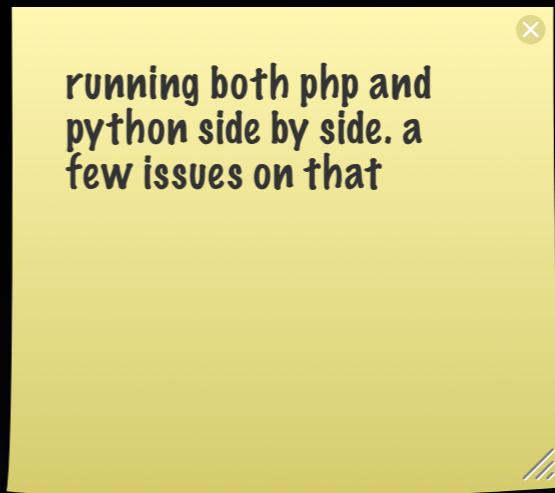
Lines of code

PHP 40k

Python 18k

Unit tests 15k

No pages go out until
they are faster



3 zeus load balancers

24 django (and php)

1 mysql + 4 slaves

3 memcached

3 sphinx

1 rabbitmq + 2 celeryd

1 redis master + 1 slave

2. Performance

As usual, database bottleneck

Cache machine

<http://bit.ly/cache-machine>



```
from django.db import models
import caching.base

class Addon(caching.base.CachingMixin,
            models.Model):
    ...
    status = models.IntegerField()
    objects = caching.base.CachingManager()
```



```
>>> Addon.objects.filter(status=public)
>>> len(connection.queries)
13
```

```
>>> Addon.objects.filter(status=public)
>>> len(connection.queries)
13
```

```
>>> Addon.objects.filter(status=public)
>>> len(connection.queries)
13
```

Invalidation

md5('select... a')

[addon 3615]

`md5('select... a')`

`[addon 3615]`

`addon 3615`

`md5('select... a')`

`md5('select... a')`

[addon 3615]

`md5('select... b')`

[addon 3615, addon 1685]

addon 3615

`md5('select... a')`
`md5('select... b')`

Memcached

```
rules = cache.get(3615)
rules.add('select... ')
cache.set(3615, rules)
```

Redis

redis.SADD(3615, ‘select...’)

Home page
20+ addons
400+ sql queries

add-on

version

version

version

files

files

files

standard answer in
django is select-related

django: select_related()

<http://bit.ly/select-related>



Transformer

<http://bit.ly/queryset-transform>

```
@staticmethod
def transformer(addons):
    addon_dict = dict((a.id, a) for a in addons)
    vs = filter(None, (a.current_version_id for a in addons))
    versions = list(Version.objects.filter(id__in=vs))
    for version in versions:
        addon_dict[version.addon_id].current_version = version
```

slightly outdated
example

big SQL statements... :(

14313 character

Home page
20+ addons
~14 sql queries

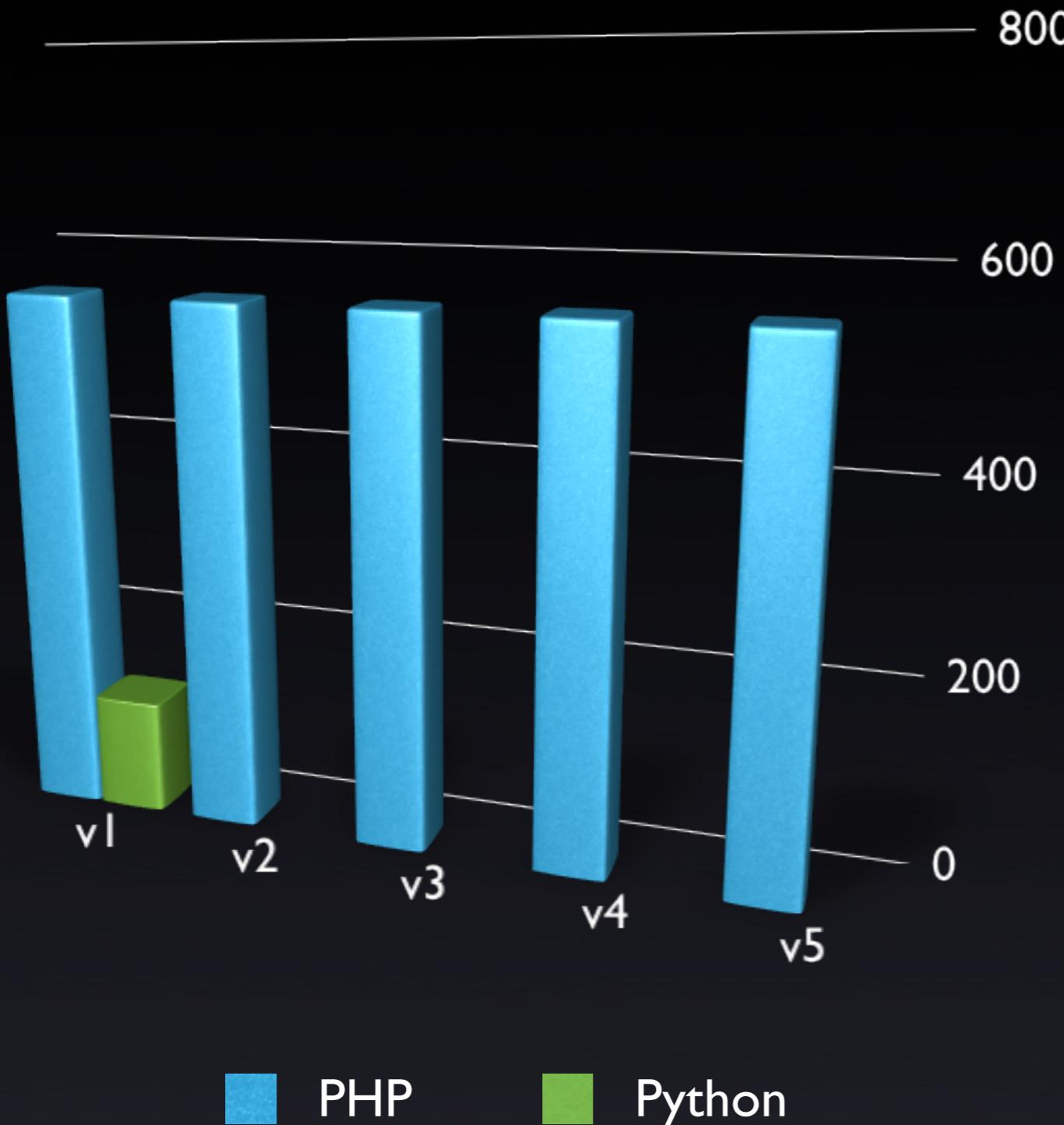
Update
Called on startup
about:config ➡ extensions.update.url

Incoming 8,000 req/sec
Uncached 1,600 req/sec



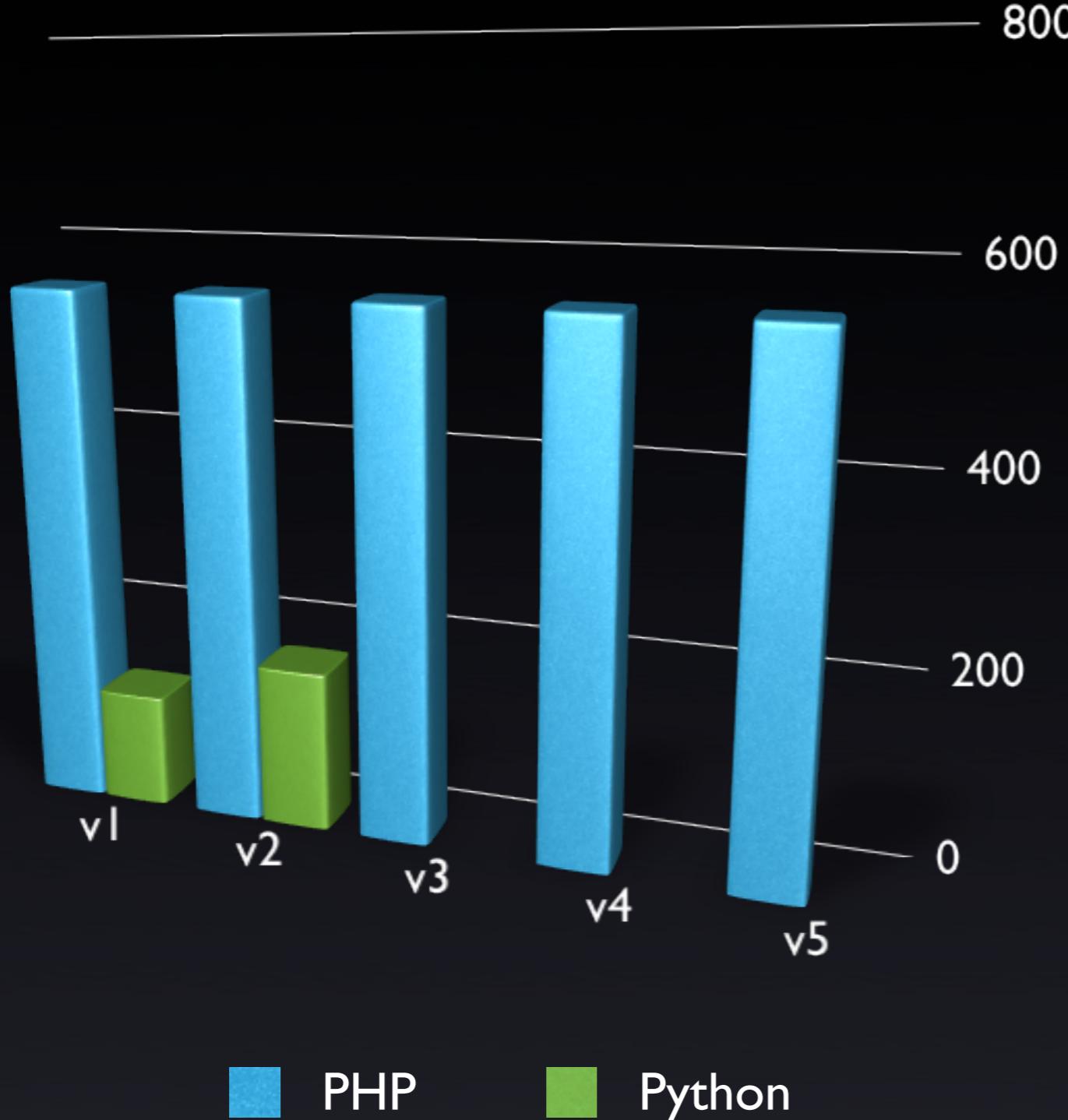
Incoming 8,000 req/sec
Uncached 1,600 req/sec
PHP 550 req/sec

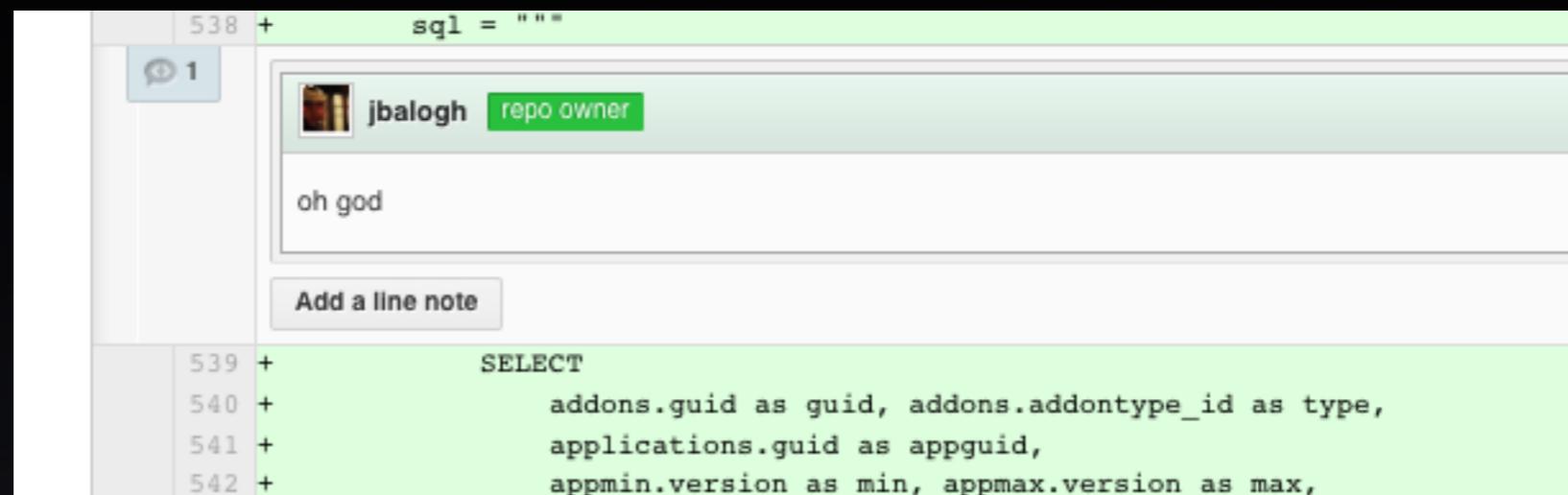




v1
Plain
Django

v2 Min. SQL queries

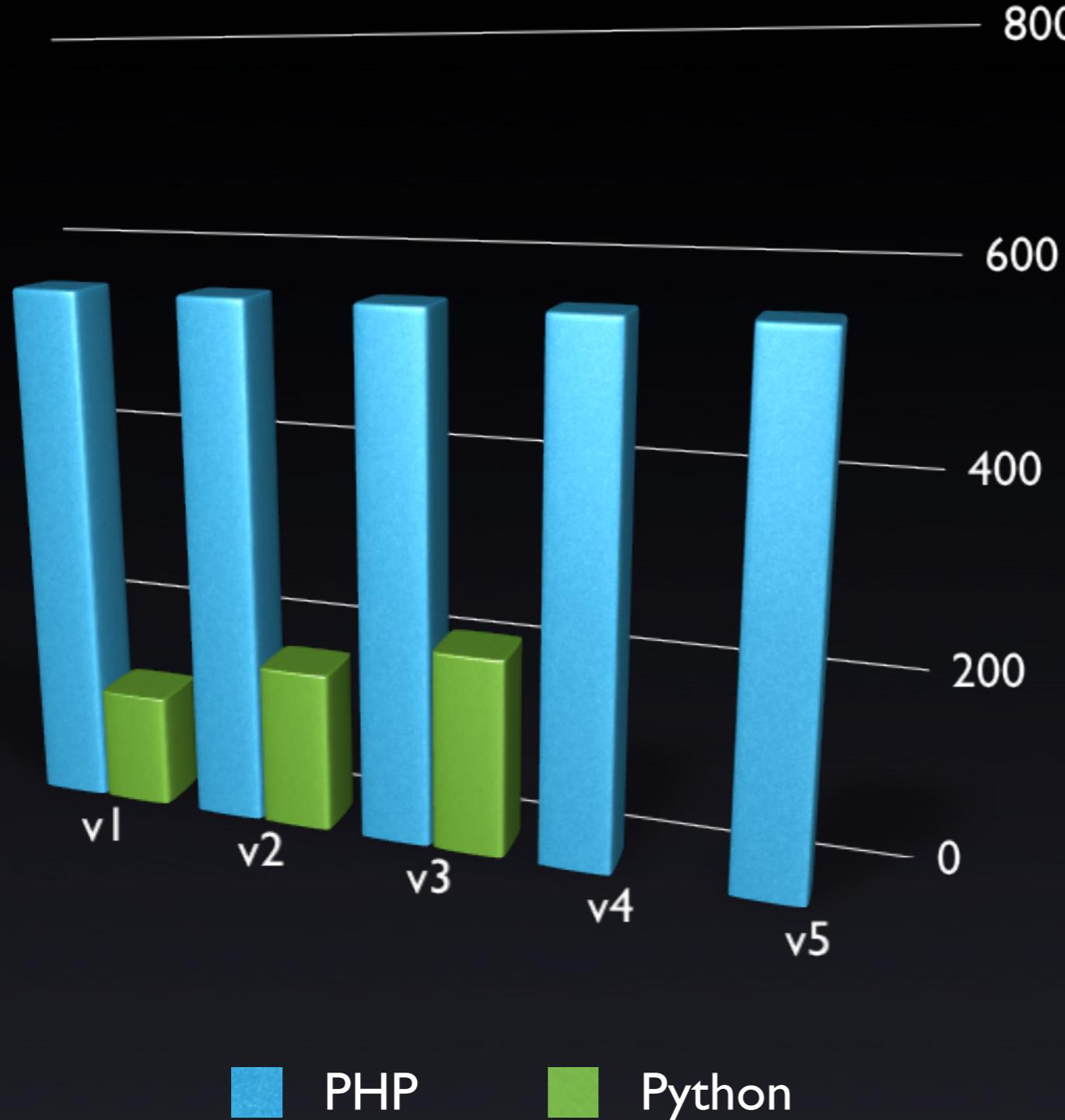




A screenshot of a GitHub pull request interface. At the top, line 538 shows the assignment of a variable `sql` to an empty string. Below this, a note from a user named `jbalogh` (repo owner) is displayed, containing the text "oh god". A button labeled "Add a line note" is visible. The bottom part of the screenshot shows code starting at line 539, which includes a `SELECT` statement with several columns from tables like `addons`, `applications`, and `appmin`.

```
538 +     sql = ""  
1  
jbalogh repo owner  
oh god  
Add a line note  
539 +     SELECT  
540 +         addons.guid as guid, addons.addontype_id as type,  
541 +         applications.guid as appguid,  
542 +         appmin.version as min, appmax.version as max,
```

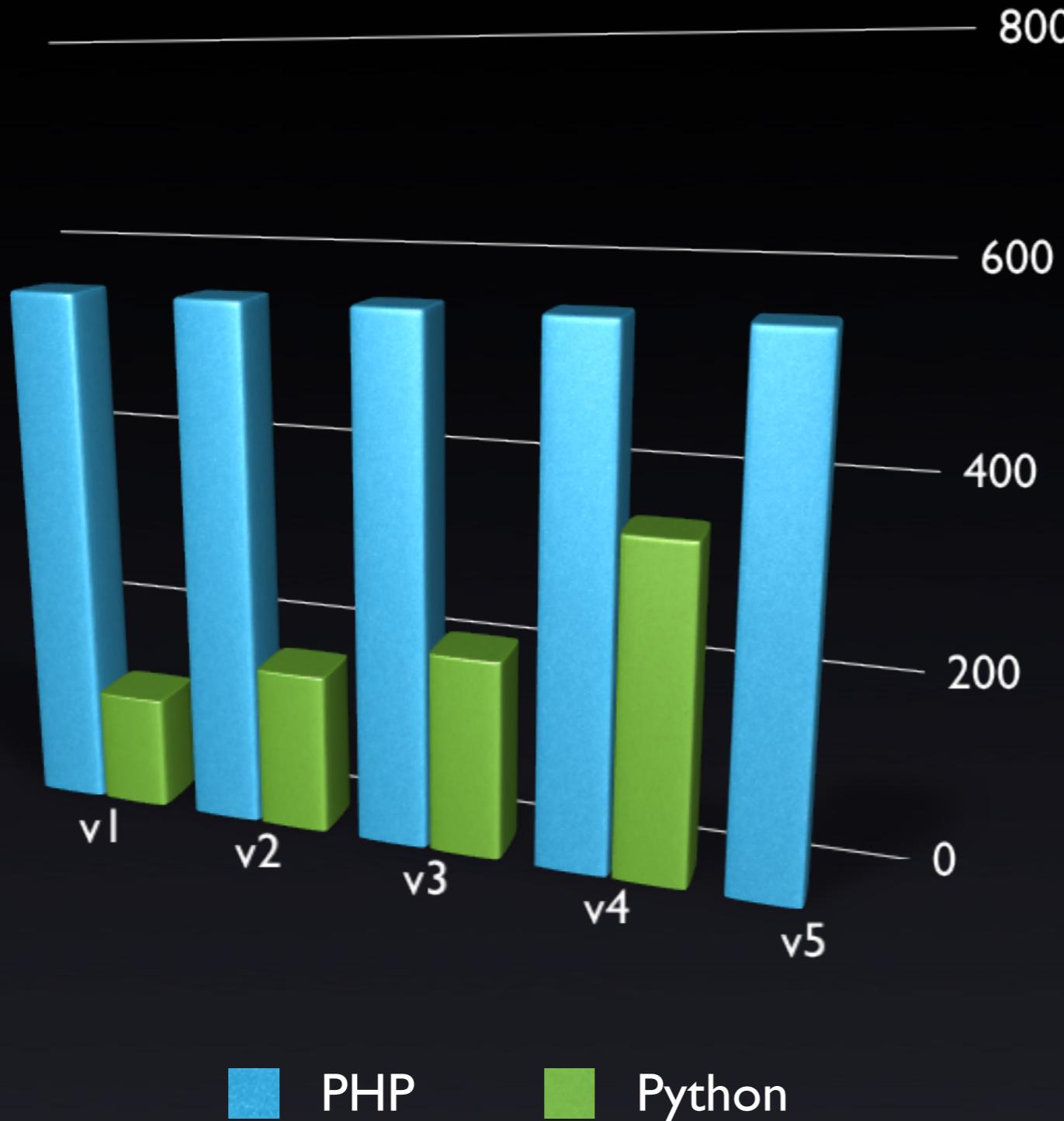
oh god



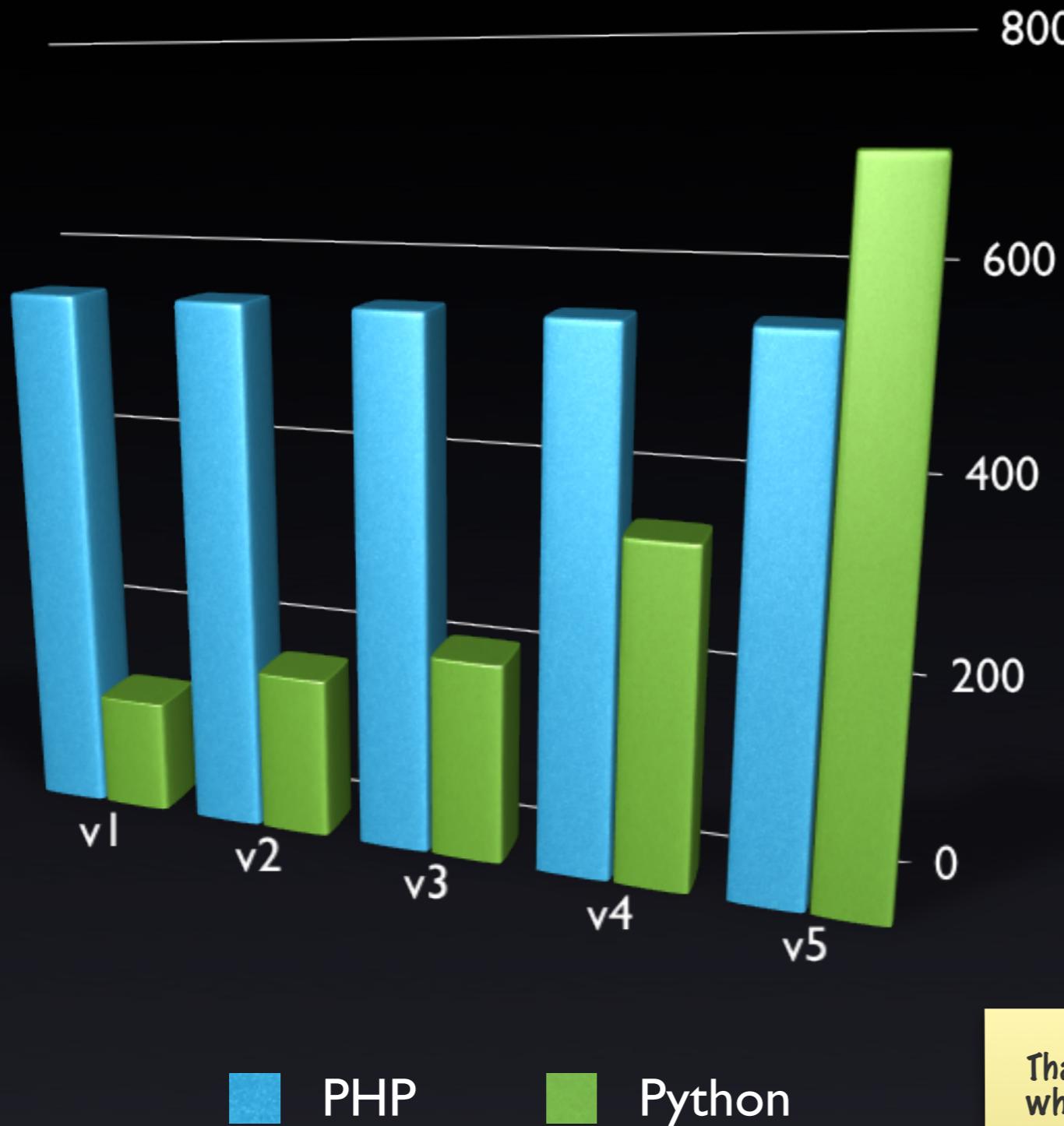
v3 Django and raw SQL

max-requests 200,
actually we hit 210

v4
WSGI
no Django



v5 Pooling, optimised queries



That's 700 req/sec
which translates into

Reducing the SQL queries...
doesn't always help

mySQL query cache is fast

A close-up photograph of fresh green celery stalks and leaves. The stalks are long, thin, and ribbed, with bright green leaves at the top. The background is dark and out of focus.

Celery

<http://celeryproject.org>

Push things async
email
image processing
add-on validation

specifically fixing data
changes between php
and python

```
from celeryutils import task

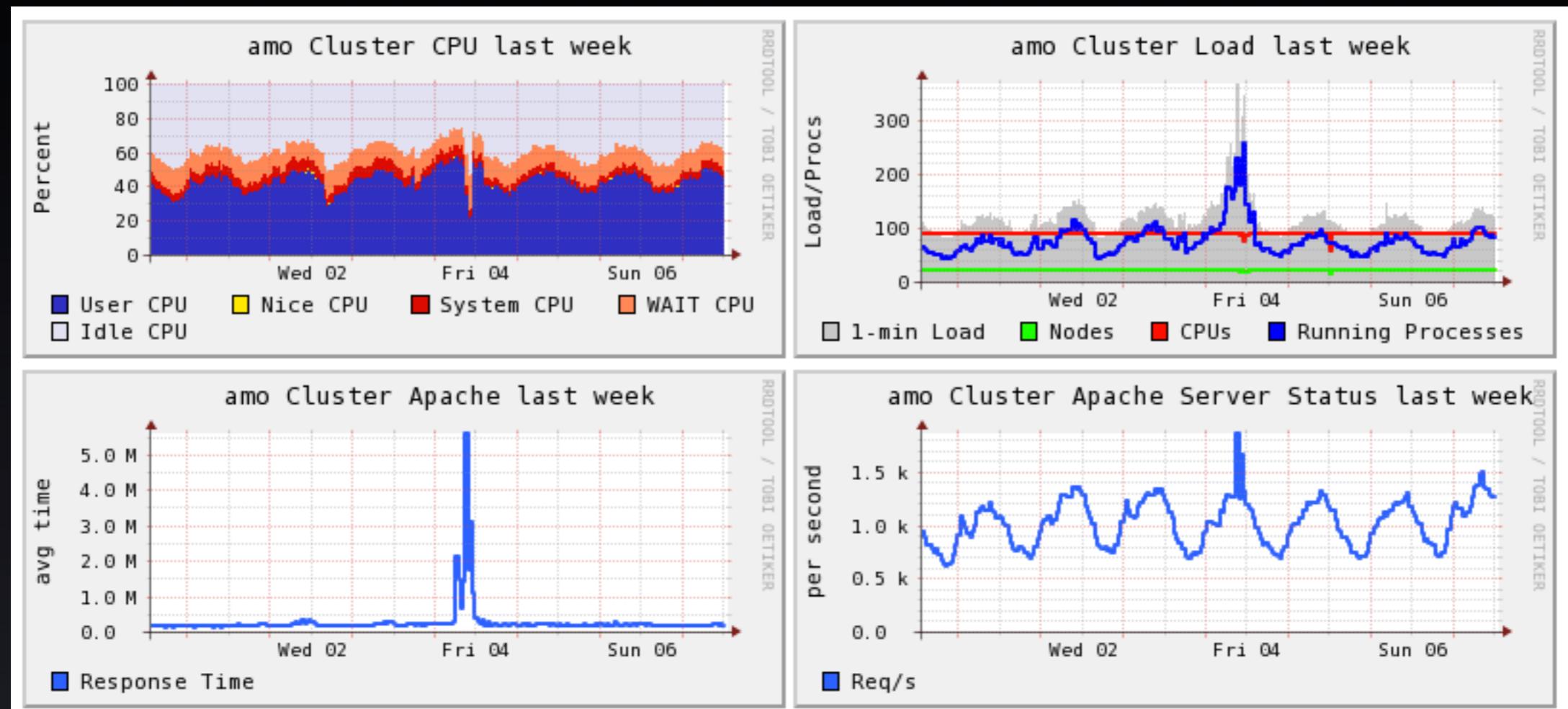
@task
def update_tag(tag, **kw):
    tag.update_stat()
```

```
from celeryutils import task

@task(rate_limit='60/h')
def update_tag(tag, **kw):
    tag.update_stat()
```

```
from tasks import update_tag  
update_tag.delay(tag)
```

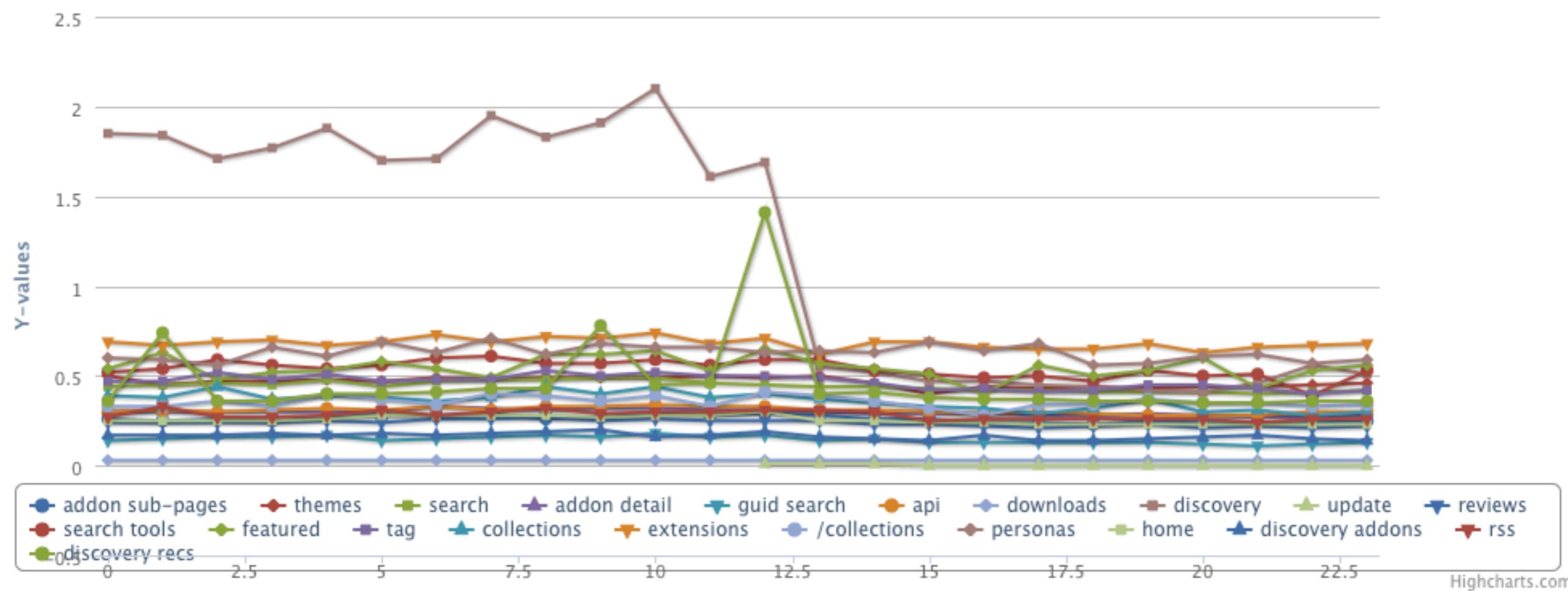
Measurement



Timing Middleware

<http://bit.ly/timing-ware>

Page performance for 02-25



mozilla

3. Localization

40+ languages
including rtl





content translated
and
templates

Database strings

```
class Addon(caching.base.CachingMixin,  
models.Model):  
    ...  
    name = models.ForeignKey(Translation)
```

```
addon.name = 'name'  
addon.save()
```

```
addon.name = 'name'  
addon.save()
```

```
addon.name = {'fr': 'la nomme'}  
addon.save()
```



Templates

Django

```
{% blocktrans with app=request.APP %}  
    Add-ons for {{ app }}  
{% endblocktrans %}
```

Jinja2

<http://jinja.pocoo.org/>

```
{{ _('Add-ons for {0}') } | if(request.APP) }}
```

Python Unicode hell

UnicodeDecodeError: 'ascii' codec can't
decode byte 0xd0 in position 16: ordinal
not in range(128)

4. Reuse

Bleach

<http://pypi.python.org/pypi/>



```
>>> bleach.clean('an  
<script>evil()</script>  
example')  
  
'an &lt;script&gt;evil()&lt;/  
script&gt; example'
```

```
>>> bleach.linkify('an http://  
ex.com url')
```

```
'an <a href="http://ex.com"  
rel="nofollow">http://ex.com</  
a> url'
```

Javascript tests

django-qunit

<http://bit.ly/django-qunit>



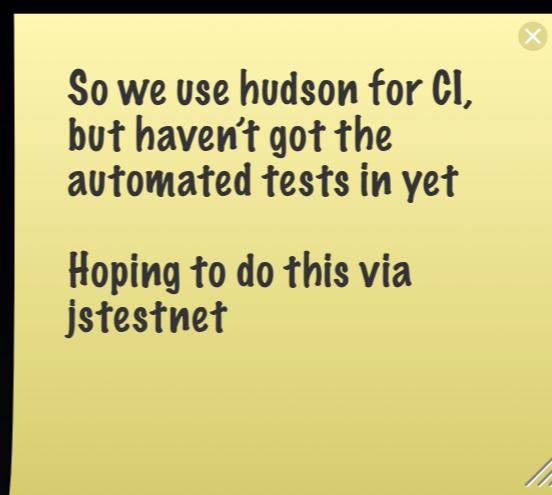
mozilla

```
test('English', function() {
  z.refreshL10n('en-us');
  equals($('textarea:visible', this.sandbox).text().trim(),
    'Firebug integrates with Firefox to put ' +
    'a wealth of development tools...');
});
```

```
test('Japanese', function() {
  z.refreshL10n('ja');
  equals($('textarea:visible', this.sandbox).text().trim(),
    'Firebug は、Web ページを閲覧中にクリック一つ ' +
    ' で使える豊富な開発ツールを Firefox' +
    ' に統合します。あなたはあらゆる');
});
```

Use Hudson Jenkins

<http://bit.ly/jstestnet>





Flake8

<http://bit.ly/flake8>

pep 8
py flakes
MacCabe

mozilla

```
~/sandboxes/zamboni(632719) $ flake8 apps/editors/tasks.py  
apps/editors/tasks.py:1: 'datetime' imported but unused  
apps/editors/tasks.py:3: 'stat' imported but unused
```



Playdoh

<http://bit.ly/mozilla-playdoh>

fred wenzel

Basis for new projects

Celery support
Jinja2 support
Simple migrations

By default:
SHA-512 password hashing
X-Frame-Options: DENY
secure and httponly flags on cookies

fred wenzel

zilla

**Take inspiration from...
but not the best for you**

for example jinja2 which
makes integration with
lots of django addons
possible, but a bit harder

Questions?

@andymckay

andym@mozilla.com

andym on irc.freenode.net, irc.mozilla.org