Drupal CDN integration: easier, more flexible and faster!

Wim Leers ~ http://wimleers.com/

Drupal.org, IRC, Twitter, LinkedIn: wimleers
Overview
Overview

- Page loading performance
  - Why does it matter?
  - What is it?
  - Why is it important to Drupal?
Overview

- Page loading performance
  - Why does it matter?
  - What is it?
  - Why is it important to Drupal?

- Key Properties of a CDN
Overview

• Page loading performance
  • Why does it matter?
  • What is it?
  • Why is it important to Drupal?

• Key Properties of a CDN

• The CDN Effect
  • YSlow
  • Episodes
Overview

• Page loading performance
  • Why does it matter?
  • What is it?
  • Why is it important to Drupal?

• Key Properties of a CDN

• The CDN Effect
  • YSlow
  • Episodes

• Drupal + CDN (where we are)
  • Basic Drupal core patch
  • Simple CDN module
  • CDN integration module
Overview

• Page loading performance
  • Why does it matter?
  • What is it?
  • Why is it important to Drupal?

• Key Properties of a CDN

• The CDN Effect
  • YSlow
  • Episodes

• Drupal + CDN (where we are)
  • Basic Drupal core patch
  • Simple CDN module
  • CDN integration module

• Bachelor thesis
  • What exactly?
  • Daemon requirements
  • Collaborating companies
Overview

• Page loading performance
  • Why does it matter?
  • What is it?
  • Why is it important to Drupal?

• Key Properties of a CDN

• The CDN Effect
  • YSlow
  • Episodes

• Drupal + CDN (where we are)
  • Basic Drupal core patch
  • Simple CDN module
  • CDN integration module

• Bachelor thesis
  • What exactly?
  • Daemon requirements
  • Collaborating companies

• The Future (where we will be in 3-4 months)
  • Goals
  • Sample scenario 1
  • Sample scenario 2
  • Sample scenario 3
  • Other use cases
  • What else?
Overview

• **Page loading performance**
  • Why does it matter?
  • What is it?
  • Why is it important to Drupal?

• **Key Properties of a CDN**

• **The CDN Effect**
  • YSlow
  • Episodes

• **Drupal + CDN (where we are)**
  • Basic Drupal core patch
  • Simple CDN module
  • CDN integration module

• **Bachelor thesis**
  • What exactly?
  • Daemon requirements
  • Collaborating companies

• **The Future (where we will be in 3-4 months)**
  • Goals
  • Sample scenario 1
  • Sample scenario 2
  • Sample scenario 3
  • Other use cases
  • What else?

• **Questions?**
Terminology: page loading performance
Terminology: page loading performance

**page rendering performance**
the time the server needs to *render*
a web page
Terminology: page loading performance

**page rendering performance**
the time the server needs to *render* a web page
Terminology: page loading performance

**Page rendering performance**
the time the server needs to *render* a web page

is included in

**Page loading performance**
the time it takes to *load* a web page
*and* all its components
Terminology: page loading performance

**Page rendering performance**
The time the server needs to *render* a web page

is included in

**Page loading performance**
The time it takes to *load* a web page and *all* its components
Why does it matter?

source: http://www.slideshare.net/stubbornella/designing-fast-websites-presentation
Why does it matter?

• Users care about performance!

  • Amazon: 100 ms of extra load time caused a 1% drop in sales
  • Google: 500 ms of extra load time caused 20% fewer searches

source: http://www.slideshare.net/stubbornella/designing-fast-websites-presentation
Why does it matter?

• Users care about performance!
  • Amazon: 100 ms of extra load time caused a 1% drop in sales
  • Google: 500 ms of extra load time caused 20% fewer searches

• **Fast** web sites are **rewarded**, **slow** web sites are **punished**

source: [http://www.slideshare.net/stubbornella/designing-fast-websites-presentation](http://www.slideshare.net/stubbornella/designing-fast-websites-presentation)
Why is it important to Drupal?
Why is it important to Drupal?

- The Drupal Experience: happier users (and developers)
Why is it important to Drupal?

• The Drupal Experience: happier users (and developers)

• Drupal’s numbers
  
  • Big, high-traffic web sites

• Popular (>125,000 sites): optimization affects many sites
Why is it important to Drupal?

• The Drupal Experience: happier users (and developers)

• Drupal’s numbers
  
  • Big, high-traffic web sites
  
  • Popular (>125,000 sites): optimization affects many sites

• Drupal is international (thanks to i18n/L10n)
  
  • International audiences: high network latencies
  
  • Developing countries: low-speed connections
What is it?

- End user response time
  - 10-20%: the HTML document (mix of back-end + front-end)
  - 80-90%: the components in the page (front-end only)

source: http://stevesouders.com/hpws/
What is it?

• End user response time

  • 10-20%: the HTML document (mix of back-end + front-end)

  • 80-90%: the components in the page (front-end only)

• More effective to focus on front-end performance!

source: [http://stevesouders.com/hpws/](http://stevesouders.com/hpws/)
What is it?

- End user response time
  - 10-20%: the HTML document (mix of back-end + front-end)
  - 80-90%: the components in the page (front-end only)
- More effective to focus on front-end performance!
  - CDNs have major impact

source: [http://stevesouders.com/hpws/](http://stevesouders.com/hpws/)
Terminology: CDN

A content delivery network (CDN) is a collection of web servers distributed across multiple locations to deliver content more efficiently to users. The server selected for delivering content to a specific user is typically based on a measure of network proximity.

source: http://developer.yahoo.com/performance/rules.html#cdn
Key Properties of a CDN
Key Properties of a CDN

- Geographical spread (PoPs)
Key Properties of a CDN

- Geographical spread (PoPs)

- Populating: Push versus Pull

<table>
<thead>
<tr>
<th>Transfer Protocol</th>
<th>Pull</th>
</tr>
</thead>
<tbody>
<tr>
<td>automatically</td>
<td>virtually no setup</td>
</tr>
<tr>
<td>no flexibility</td>
<td>no redundant traffic</td>
</tr>
</tbody>
</table>
Key Properties of a CDN

- Geographical spread (PoPs)

- Populating: Push versus Pull

<table>
<thead>
<tr>
<th></th>
<th>Pull</th>
<th>Push</th>
</tr>
</thead>
<tbody>
<tr>
<td>transfer protocol</td>
<td>automatically</td>
<td>FTP, SFTP, rsync, WebDAV, Amazon S3 ...</td>
</tr>
<tr>
<td>+</td>
<td>virtually no setup</td>
<td>flexibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no redundant traffic</td>
</tr>
<tr>
<td>-</td>
<td></td>
<td>setup</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no flexibility redundant traffic</td>
</tr>
</tbody>
</table>
Key Properties of a CDN

- Geographical spread (PoPs)
- Populating: Push versus Pull
- Lock-in

<table>
<thead>
<tr>
<th></th>
<th>Pull</th>
<th>Push</th>
</tr>
</thead>
<tbody>
<tr>
<td>transfer protocol</td>
<td>automatically</td>
<td>FTP, SFTP, rsync, WebDAV, Amazon S3</td>
</tr>
<tr>
<td>virtual no setup</td>
<td>no flexibility</td>
<td>flexibility</td>
</tr>
<tr>
<td>redundant traffic</td>
<td>redundant traffic</td>
<td>setup</td>
</tr>
<tr>
<td>no flexibility no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>redundant traffic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The CDN Effect: YSlow
The CDN Effect: YSlow

- Make YSlow think Drupal.org is using a CDN:
The CDN Effect: YSlow

- Make YSlow think Drupal.org is using a CDN:
The CDN Effect: YSlow

- Make YSlow think Drupal.org is using a CDN:
The CDN Effect: YSlow

• Make YSlow think Drupal.org is using a CDN:

    ![Performance Grade: F (59)]

    1. Make fewer HTTP requests
    2. Use a CDN
    3. Add an Expires header
    4. Gzip components
    5. Put CSS at the top
    6. Put JS at the bottom
    7. Avoid CSS expressions
    8. Make JS and CSS external
    9. Reduce DNS lookups
   10. Minify JS
   11. Avoid redirects
   12. Remove duplicate scripts
   13. Configure ETags

    ![Performance Grade: C (70)]

    1. Make fewer HTTP requests
    2. Use a CDN
    3. Add an Expires header
    4. Gzip components
    5. Put CSS at the top
    6. Put JS at the bottom
    7. Avoid CSS expressions
    8. Make JS and CSS external
    9. Reduce DNS lookups
   10. Minify JS
   11. Avoid redirects
   12. Remove duplicate scripts
   13. Configure ETags

• YSlow does **not** reflect **real-world** page loading performance!
The CDN Effect: Episodes
The CDN Effect: Episodes

- JS code that measures episodes of a page’s loading sequence
The CDN Effect: Episodes

- JS code that measures episodes of a page’s loading sequence
The CDN Effect: Episodes

- JS code that measures episodes of a page’s loading sequence

- Reflects **real-world** page loading performance!
The CDN Effect: Episodes

- JS code that measures episodes of a page’s loading sequence
- Reflects **real-world** page loading performance!
- Goal: **industry standard**!
The CDN Effect: Episodes

• JS code that measures episodes of a page’s loading sequence

• Reflects real-world page loading performance!

• Goal: industry standard!

• Drupal module: available next week!
Drupal + CDN: Basic Drupal core patch

(http://www.slayerment.com/blog/drupal-cdn-file-server-amazon-s3-way)
Drupal + CDN: Basic Drupal core patch

(http://www.slayerment.com/blog/drupal-cdn-file-server-amazon-s3-way)

• Change all code that calls base_path():
  • common.inc
  • file.inc
  • theme.inc
  • page.tpl.php
Drupal + CDN: Basic Drupal core patch

(http://www.slayerment.com/blog/drupal-cdn-file-server-amazon-s3-way)

- Change all code that calls `base_path()`:
  - `common.inc`
  - `file.inc`
  - `theme.inc`
  - `page.tpl.php`

- Cons
  - Very limited
Drupal + CDN: Simple CDN module

http://drupal.org/project/simplecdn
Drupal + CDN: Simple CDN module

http://drupal.org/project/simplecdn

• simplecdn_rewrite_url($url, 'modulename')
Drupal + CDN: Simple CDN module

http://drupal.org/project/simplecdn

- `simplecdn_rewrite_url($url, 'modulename')`

- Cons
  
  - **Very limited**
  
  - Not Drupal core & **more dependencies**
  
  - Immature (released Feb 21, 2009), coding standards, no proper releases
Drupal + CDN: CDN integration module

http://drupal.org/project/cdn (by me)
Drupal + CDN: CDN integration module

http://drupal.org/project/cdn (by me)

- Also includes a **Drupal core patch**
Drupal + CDN: CDN integration module

http://drupal.org/project/cdn (by me)

• Also includes a **Drupal core patch**

• Automatic synchronization via **FTP**
Drupal + CDN: CDN integration module

http://drupal.org/project/cdn (by me)

- Also includes a **Drupal core patch**

- Automatic synchronization via **FTP**

- **Cons**
  - **Still too limited** (rsync, Amazon S3, image optimization ...)
  - Drupal 5 only
  - **Proof-of-concept**: not scalable (cron and ... serialized arrays!)
Bachelor thesis

“Improving Drupal’s page loading performance”

Hasselt University has allowed me to work in an open source fashion and all of my work will therefore be available for everybody. Thanks!

- Promotor: Prof. dr. Wim Lamotte
- Co-promotor: dr. Peter Quax
- Mentors: Stijn Agten & Maarten Wijnants
Bachelor thesis: What exactly?

• **Episodes.module**
  
  • Basic reports (e.g. average page load time per browser)

• **CDN integration**
  
  • Drupal core patch: unify file URL generation/alteration (sufficient for Pull CDNs)
  
  • Drupal module: administration & basic reports

  • **Daemon** for synchronization (Push CDNs)

• **JS at the bottom**
Bachelor thesis: Daemon requirements

• Python

• Full **unit testing** coverage

• Very **extensible**
  
  • Plug-ins for transfer protocols
  
  • Plug-ins for processors

• Very **robust & scalable**
  
  • Multi-threaded, must be able to handle *any* amount of files
Bachelor thesis: Collaborating companies

The following companies will be evaluating the improvements to Drupal or the daemon.

The following CDN companies will be providing their services for free for testing purposes.

A big thanks to all of them!
The Future: Goals
The Future: Goals

- **Transparency** (transfer protocol irrelevant)
The Future: Goals

- **Transparency** (transfer protocol irrelevant)

- **Mixing** CDNs and static file servers
The Future: Goals

• **Transparency** (transfer protocol irrelevant)

• **Mixing** CDNs and static file servers

• **Processing** before sync: image optimization, transcoding …
The Future: Goals

• **Transparency** (transfer protocol irrelevant)

• **Mixing** CDNs and static file servers

• **Processing** before sync: image optimization, transcoding …

• Detect (& sync) new files **instantly** (inotify/FSEvents/WMI)
The Future: Goals

- **Transparency** (transfer protocol irrelevant)

- **Mixing** CDNs and static file servers

- **Processing** before sync: image optimization, transcoding ...

- Detect (& sync) new files **instantly** (inotify/FSEvents/WMI)

... thanks to the daemon I’ll write.
The Future: Sample scenario 1

Company wants to switch from CDN provider A to provider B. A only supports FTP, B only supports SFTP.
The Future: Sample scenario 1

Company wants to switch from CDN provider A to provider B. A only supports FTP, B only supports SFTP.

- Setup:
  - CDN A → CDN B
  - Daemon
The Future: Sample scenario 1

Company wants to switch from CDN provider A to provider B. A only supports FTP, B only supports SFTP.

• Setup:

  • CDN A → CDN B

• Daemon

• Alternative: write a lot of code
The Future: Sample scenario 1

Company wants to switch from CDN provider A to provider B. A only supports FTP, B only supports SFTP.

- Setup:
  - CDN A $\rightarrow$ CDN B
  - Daemon

- Alternative: write a lot of code

- Result: no switching cost $\rightarrow$ no lock-in!
The Future: Sample scenario 2

U.S. company expands to South-Korea
The Future: Sample scenario 2

U.S. company expands to South-Korea

- Setup:
  - North-American CDN
  - Static file servers in South-Korea
  - Daemon + GeoIP API/Domain GeoLocalization module
The Future: Sample scenario 2

U.S. company expands to South-Korea

• Setup:
  • North-American CDN
  • Static file servers in South-Korea
  • Daemon + GeoIP API/Domain GeoLocalization module
• Alternative: global CDN
The Future: Sample scenario 2

U.S. company expands to South-Korea

• Setup:
  
  • North-American CDN
  
  • Static file servers in South-Korea
  
  • Daemon + GeoIP API/Domain GeoLocalization module

• Alternative: global CDN

• Result: faster web site, more flexible & cheaper setup, same setup cost
The Future: Sample scenario 3

U.S. company hosts an event in Washington D.C.
The Future: Sample scenario 3

U.S. company hosts an event in Washington D.C.

• Setup:
  
  • North-American CDN
  
  • Static file servers in Washington D.C.
  
  • Daemon + GeoIP API/Domain GeoLocalization module
The Future: Sample scenario 3

U.S. company hosts an event in Washington D.C.

• Setup:
  • North-American CDN
  • Static file servers in Washington D.C.
  • Daemon + GeoIP API/Domain GeoLocalization module
• Alternative: a slower web site
The Future: Sample scenario 3

U.S. company hosts an event in Washington D.C.

- Setup:
  - North-American CDN
  - Static file servers in Washington D.C.
  - Daemon + GeoIP API/Domain GeoLocalization module
- Alternative: a slower web site
- Result: faster web site, same setup cost
The Future: Other use cases

• Massive back-up tool

• Transcoding server (cfr. MediaMover module)

• Key component in creating your own CDN

• Key component for a Dropbox-like tool
The Future: What else?

What else do you need?

More use cases?

More ideas?

…

Join the BoF in the <location> room at <time>!
Questions?

Relevant links:

• This presentation – along with relevant links – can be found online at http://wimleers.com/talk/drupalcon-dc-2009

• You can track my bachelor thesis’ progress at http://wimleers.com/tags/bachelor-thesis

• Questions & feedback are welcome! Contact me via http://wimleers.com/contact

• Modules: http://drupal.org/project/cdn & http://drupal.org/project/episodes