

# Web Performance Optimization: Analytics

---

Wim Leers

# Bachelor thesis

---

## “Improving Drupal’s page loading performance”

- Promotor: Prof. dr. Wim Lamotte
- Co-promotor: dr. Peter Quax
- Mentors: Stijn Agten & Maarten Wijnants
- Score: 19/20



# Goal

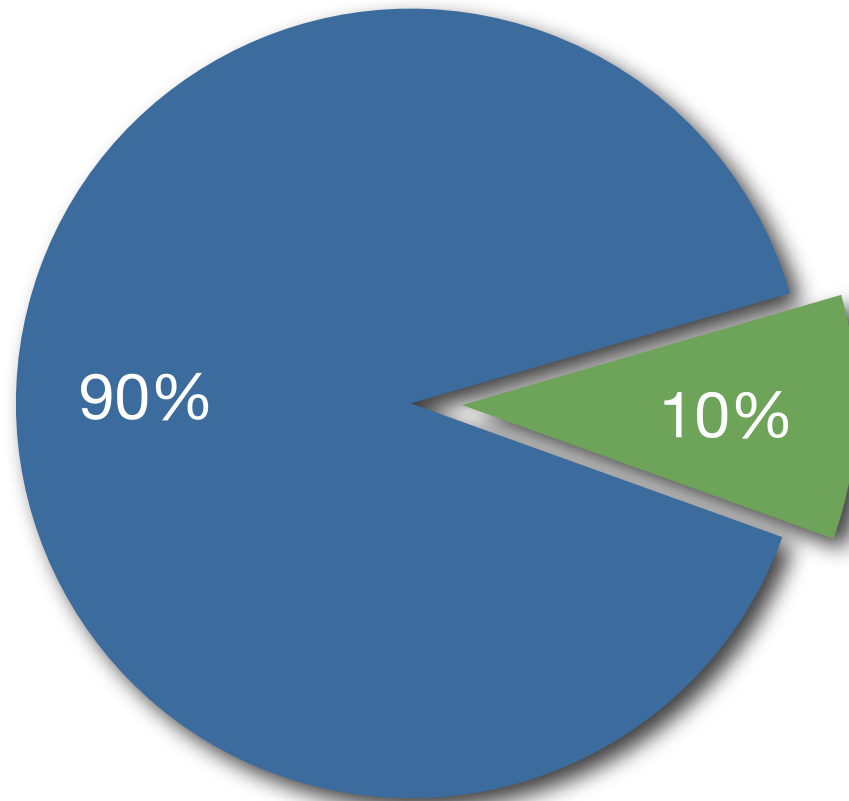
---

- **Faster web sites**
  - Speed = satisfaction = more & happier visitors = more revenue
- Google: +0.5s → -20% searches



# Terminology: page loading performance

---



- CSS, JS, images ...
- HTML

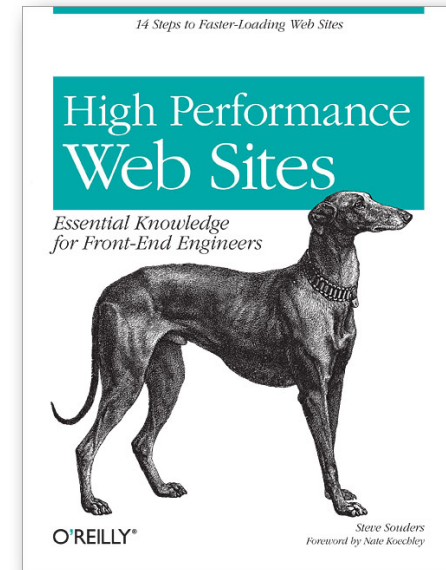
# Drupal's page loading performance

---

- One of the most effective measures:

*“Use a CDN”*

- Drupal: ~~not yet possible!~~ **Possible now! :)**



# Terminology: CDN

---



# Profiling: *Episodes*

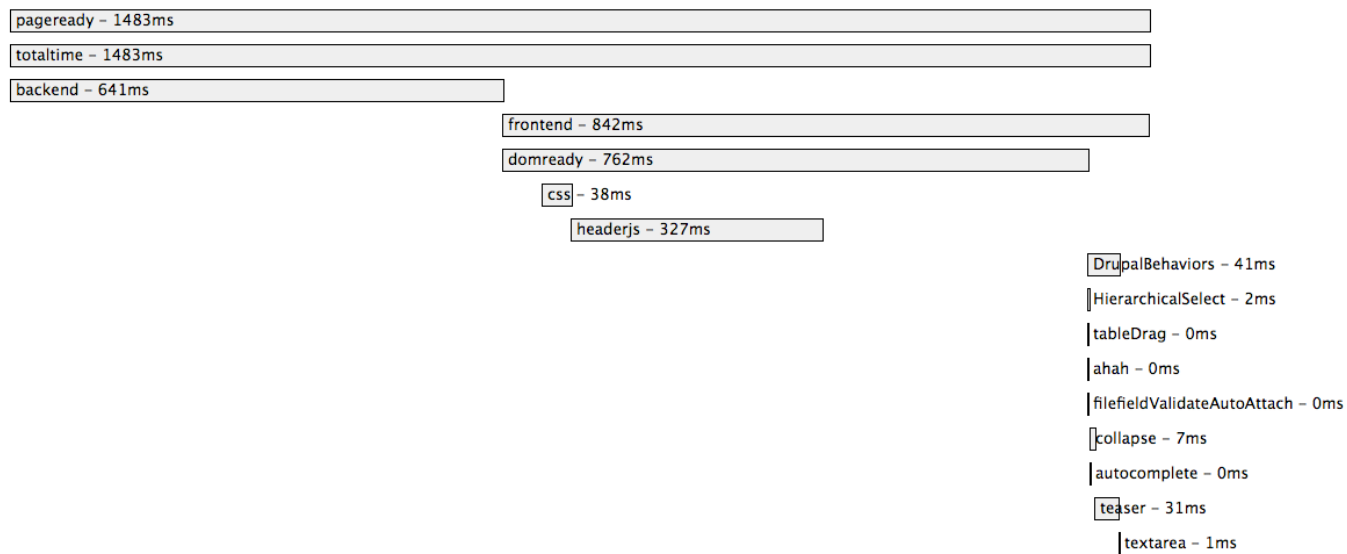
---

- Measures “episodes” during page loading
- **Real measurements:** JS in browser, for *each* visitor
- No simulation!

# Episodes module

---

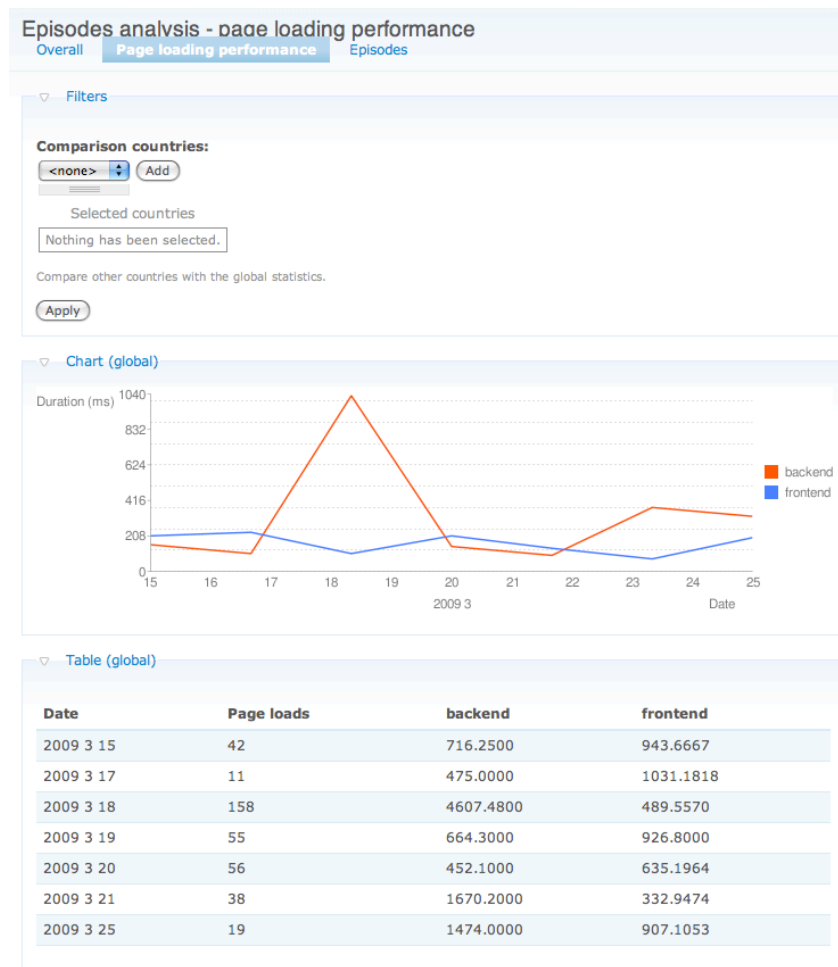
- **Drupal module** that offers Episodes integration





# Episodes Server module

- **Drupal module** that visualizes the collected measurements



# CDN integration module

---

- Drupal core patch for Drupal 6, committed to Drupal 7
- Basic mode: Origin Pull CDN
- Advanced mode: **any CDN** (File Conveyor)

The image displays two screenshots of the Drupal 7 administration interface for the CDN integration module, specifically the 'CDN integration' configuration page. Both screenshots show the breadcrumb trail: Home > Administer > Site configuration > CDN integration.

**Left Screenshot (Basic mode):**

- CDN integration** (Settings | **Basic mode** | Advanced mode)
- CDN URL:**   
The CDN URL prefix that should be used. Only works for CDNs that support the 'Origin Pull' method.  
**WARNING:** do not use subdirectories when you're serving Content from within the CSS files will break because the path to Images and fonts from within the CSS files will break because the path is relative to the domain.
- Allowed extensions:**   
Only files with these extensions will be synced.
- Buttons: Save configuration, Reset to defaults

**Right Screenshot (Advanced mode):**

- CDN integration** (Settings | Basic mode | **Advanced mode** | Other)
- Message:** The synced files database was found and can be opened for reading.
- Synced files database:**   
Enter the full path to the daemon's synced files database file.
- Buttons: Save configuration, Reset to defaults


# Test case

News | DriverPacks.net

# DriverPacks.net

RÉPARER WINDOWS

Prévenir les erreurs de Windows



GRATUITEMENT

www.Quad-Cleaner.com

News

Downloads

Forum

Bugtracker

Well-known users

About

Donate

## What are the "DriverPacks"?

Each DriverPack is a package of drivers that fit a certain category. There are for example DriverPacks for Sound devices and for Graphics devices. For each of these DriverPacks we include *every* device driver of that category. On top of that, we solve tricky problems — such as driver conflicts — for you.

They are intended to be used while creating a Windows 2000, XP, 2003 or Vista (and soon Windows 7) installation disc/image. We also provide you with the software to make that as simple as possible.

We've built this to **simplify the lives of network administrators and employees of pc shops**, but also enthusiast pc users.

Sounds cool? Then explore our [downloads](#) and come and say hi in our [forum](#)!

### Latest DriverPacks

XP

Vista

7

- [Chipset 9.02](#)
- [Mass Storage 9.01](#)
- [Graphics A 8.12.1](#)
- [Graphics B 8.12.1](#)
- [Graphics C 8.12.1](#)

Welcome from *Belgium*!

For the fastest experience, you are downloading the images, CSS and JS from our static file server in Belgium.

35 online

save on Delicious

tags: windows drivers software unattended xp installation tools hardware slipstream driver

saved by 548 others

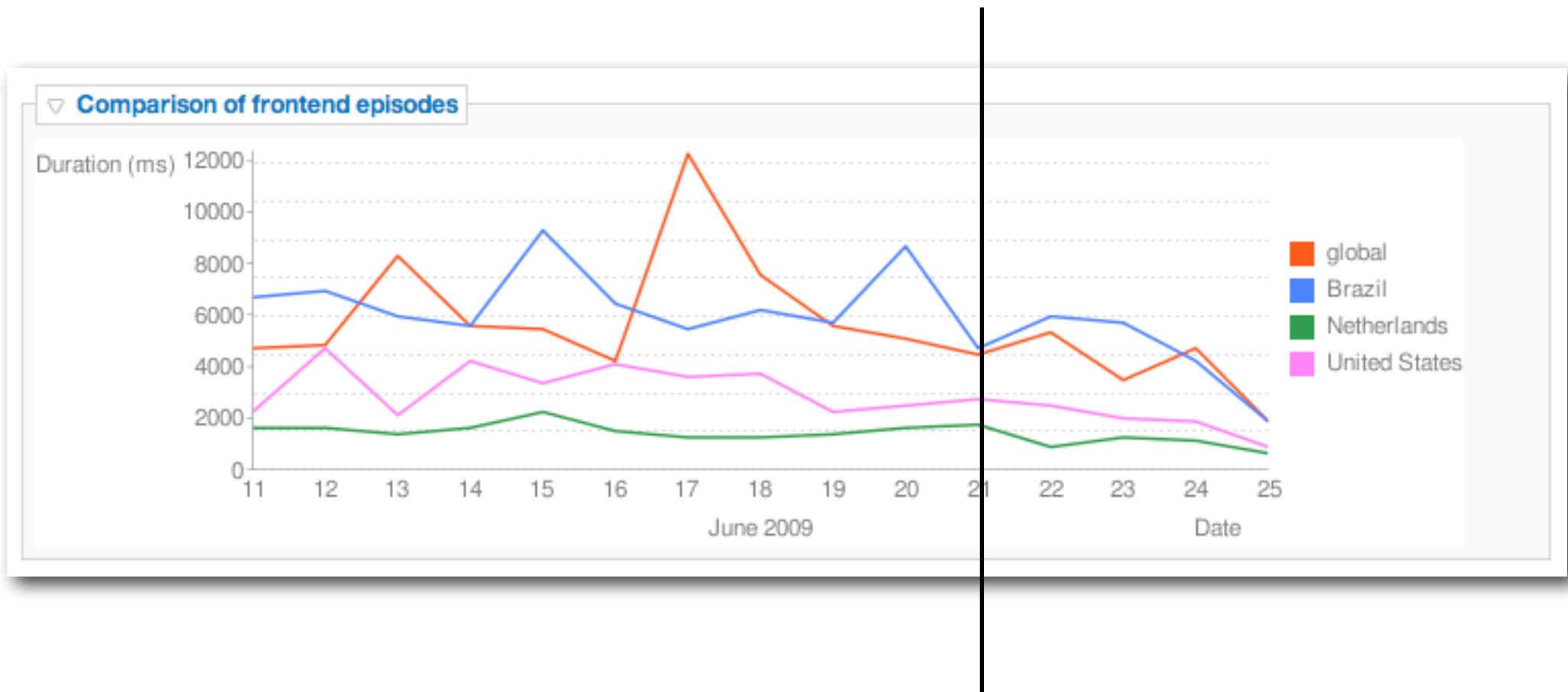
## New site live!

Posted on Thu, 06/11/2009 - 03:14 by [Wim Leers](#)

Finally the time has come! The new web site went live today!

# Test case

---



# Master thesis

---

## **“Web Performance Optimization: Analytics”**

- Promotor: Prof. dr. Jan van den Bussche



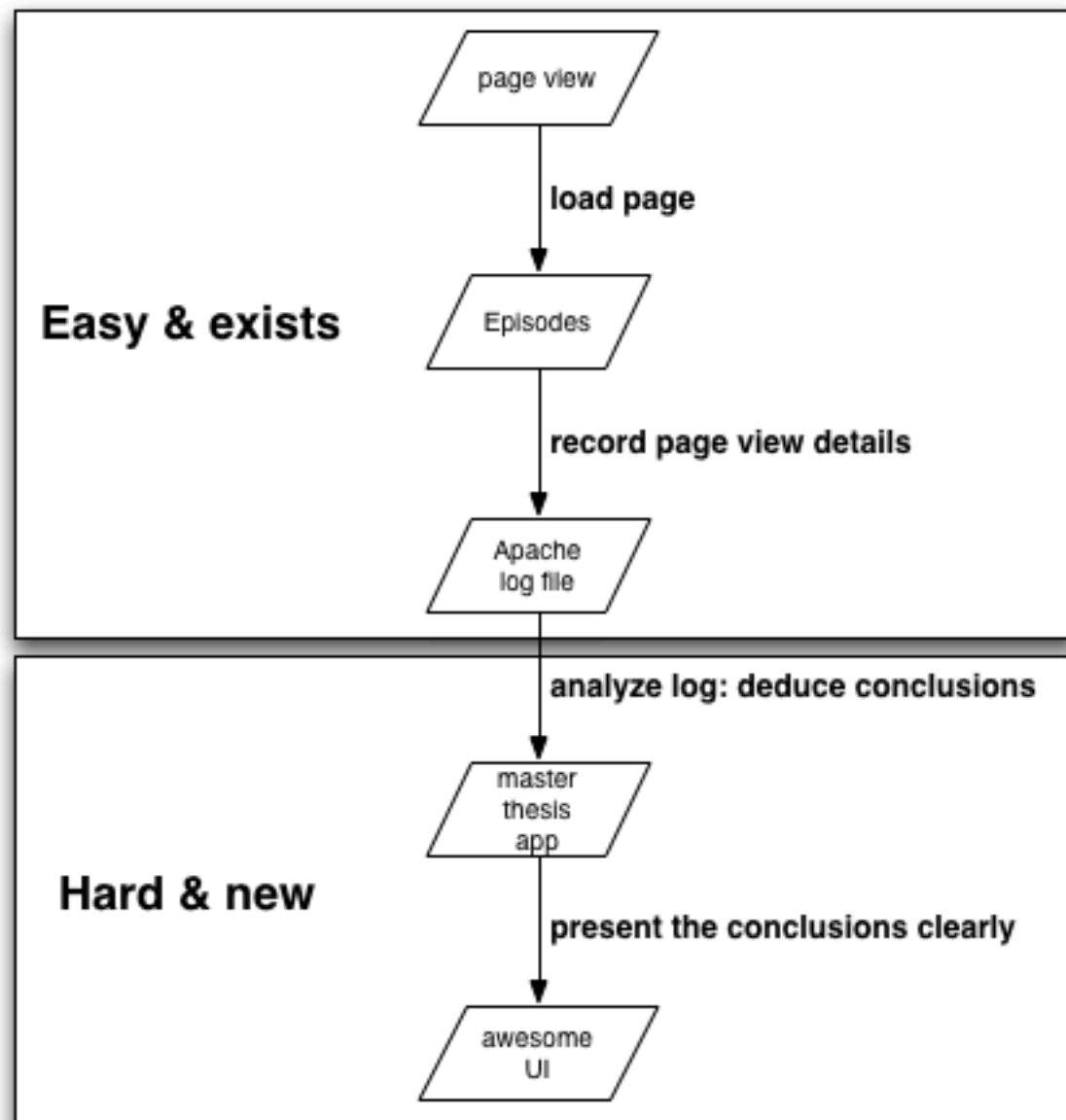
# Master thesis

---

- An application that can automatically extract **conclusions** out of Episodes logs and **visualize** them.  
i.e. **automatically pinpoint causes of slow page loads.**
- e.g.:
  - “<http://d.o/> is slow in Belgium, for users of the ISP Telenet”
  - “<http://d.o/project/esi> has slowly loading CSS”
  - “<http://d.o/project/cdn> has slowly loading JS for visitors that use the browser Internet Explorer 6 or 7”
- Challenges: complex data mining, distributed computing, storage, performance

# Concept

---



# Apache log file

---

211.138.37.206 [Sunday, 21-Jun-2009 06:23:37 +0200] "?ets=css:63,headerjs:4453,footerjs:16,domready:7359,tabs:31,ToThePointShowHideChangelog:0,gaTrackerAttach:16,DrupalBehaviors:47,frontend:8015" 200 "<http://driverpacks.net/applications>" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; (R1 1.6); .NET CLR 2.0.50727)" "driverpacks.net"

211.138.37.206 [Sunday, 21-Jun-2009 06:24:04 +0200] "?ets=css:0,headerjs:2313,footerjs:0,tabs:31,ToThePointShowHideChangelog:0,gaTrackerAttach:15,DrupalBehaviors:46,domready:5671,frontend:6031" 200 "<http://driverpacks.net/driverpacks/latest>" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; (R1 1.6); .NET CLR 2.0.50727)" "driverpacks.net"

114.217.216.157 [Sunday, 21-Jun-2009 06:24:15 +0200] "?ets=backend:3750,css:16,headerjs:7125,footerjs:16,domready:15125,tabs:31,ToThePointShowHideChangelog:16,gaTrackerAttach:0,DrupalBehaviors:47,pageready:19422,frontend:15672,totaltime:19422" 200 "<http://driverpacks.net/>" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1) ; InfoPath.2; .NET CLR 2.0.50727)" "driverpacks.net"

121.73.5.7 [Sunday, 21-Jun-2009 06:24:17 +0200] "?ets=css:2156,headerjs:1328,footerjs:484,tabs:63,ToThePointShowHideChangelog:0,gaTrackerAttach:0,DrupalBehaviors:63,domready:7625,frontend:7672" 200 "<http://driverpacks.net/driverpacks/latest>" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1; .NET CLR 1.1.4322; .NET CLR 2.0.50727)" "driverpacks.net"



# Sample conclusions explained

---

- *“<http://d.o/> is slow in Belgium, for users of the ISP Telenet”*
  - associate IP ranges with specific IPs and relate this to slow page loads (i.e. a slow totaltime episode)
- *“<http://d.o/project/esi> has slowly loading CSS”*
  - associate a slow css episode with a specific URL
- *“<http://d.o/project/cdn> has slowly loading JS for visitors that use the browser Internet Explorer 6 or 7”*
  - associate a slow js episode with a specific URL and user-agents

# Required techniques

---

- conclusions: fairly advanced (or at least elaborate) data mining
  - IP → location, IP → ISP
- scalability: distributed computing
- performance: near-real time results (e.g. 5 minute lag) required
- storage: custom storage necessary: >400 bytes/pageview is not acceptable
  - granularity/resolution: 5/15/30/60 minutes?
  - take data visualization into account: granularity of data limits the possibilities!

## News feed

---

At <http://wimleers.com/tags/master-thesis>