

Construyendo proyectos escalables con Drupal

Pablo Cerda

Director de desarrollo de **ideup!**
www.ideup.com

Drupalcamp Spain 2010

ideup!

De qué va el tema

Crónicas de un programador que heredó un proyecto (lanetro.com) que no seguía las buenas prácticas de Drupal y petaba por todos lados.

Narraremos cómo reconducimos y refactorizamos un proyecto que parecía ir directo al fracaso y que al final se ha convertido en un caso de éxito de una aplicación bien hecha en Drupal.

Velocidad de evolución

- Los servicios web que utilizamos a diario (Google, Flickr, Facebook, etc.) evolucionan rápidamente.
- El usuario se vuelve más experto.
- Para permanecer líderes, no hay otro camino que reinventarse constantemente.
- Reinventarse significa invertir en un equipo tecnológico que sea capaz de mantener lo que funciona y expandirlo de forma inteligente.

Cuidado con perder el control



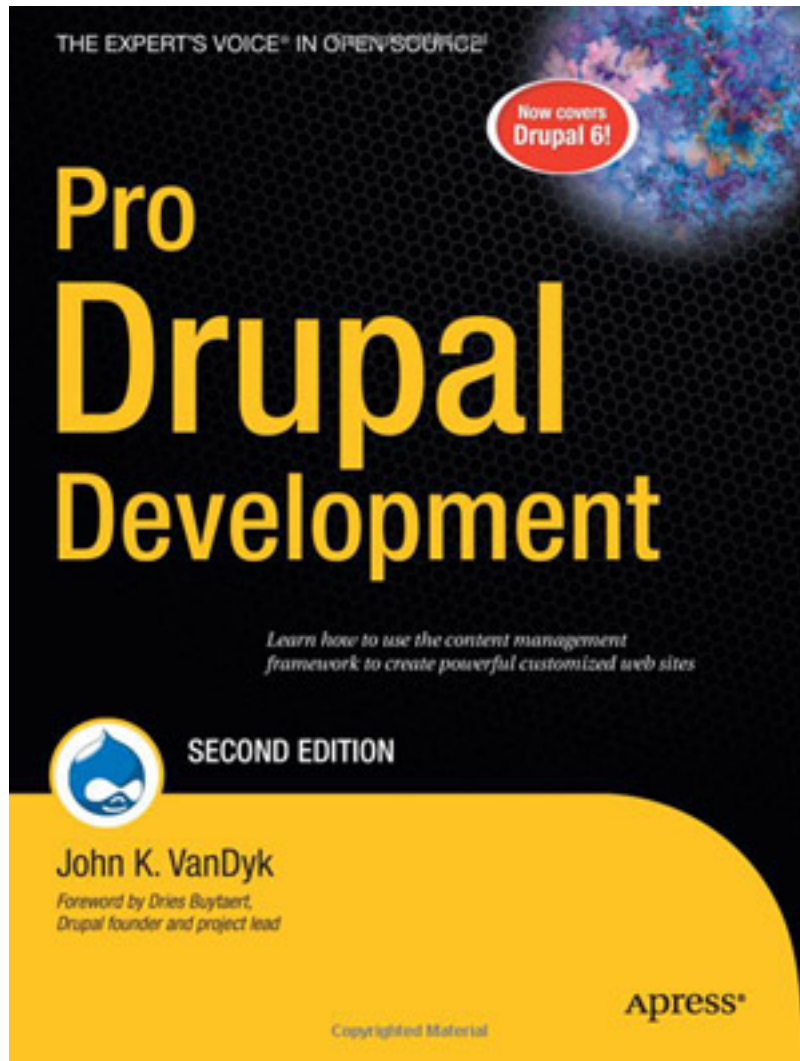
Es difícil ver el punto de ruptura entre una aplicación sostenible o no. El síntoma es que de pronto es muy costoso corregir errores o añadir funcionalidades.

Miles de cabezas piensan más que unas cuantas

- ✓ Muchas de las funcionalidades que quieres desarrollar ya están hechas. ¡Reutilízalas!
- ✓ Muchos de los desarrolladores que colaboran con Drupal tienen décadas de experiencia. ¡Benefíciate!
- ✓ ¿Podrá un equipo de 1 o 2 programadores juniors hacerlo mejor?



Cimenta tus conocimientos de Drupal



- ✓ Este libro es una lectura fundamental para cualquiera que desee hacer cosas serias y bien hechas con Drupal.
- ✓ En Abril sale la edición para Drupal 7. ¡Estad atentos!

Drupalcamp Spain 2010

ideup!

¡Que se convierta en éxito no en pesadilla!



- ✓ Una instalación básica de Drupal está bien para empezar.
- ☹ La dificultad comenzará cuando se rebase el umbral de las 5,000 visitas diarias.
- ✓ ¡Ha llegado el momento de escalar!

¡Drupal puede escalar, y mucho!

Lanetro Estandar Panel

01/01/2010 - 31/01/2010

En comparación con: Sitio



Uso del sitio

2.167.757 Visitas

5.698.853 Páginas vistas

2,63 Páginas/visita

54,34% Porcentaje de rebote

00:02:31 Promedio de tiempo en el sitio

73,04% Porcentaje de visitas nuevas

Visión general de usuarios



Usuarios del sitio web
1.891.521

Gráfico de visitas por ubicación



Definiendo escalabilidad

- ✓ El sistema puede acomodarse al incremento de su uso.
- ✓ El sistema puede acomodarse al incremento de su volumen de datos.
- ✓ El sistema es sostenible.

- Cal Henderson, Building Scalable Websites (O'Reilly)

Definiendo rendimiento

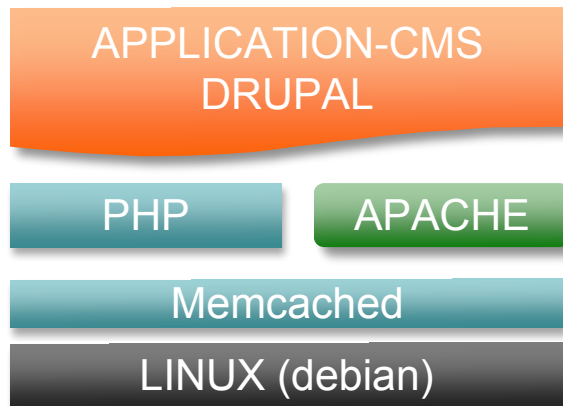
- Tiempo de respuesta - umbral entre que se solicita y se completa una petición.
- Tiempo de carga de página - tiempo de todos los componentes que se descargan para renderizarse en el navegador.
- Sus requerimientos pueden variar dependiendo de la audiencia.

¿Hasta que punto mejorar el rendimiento?

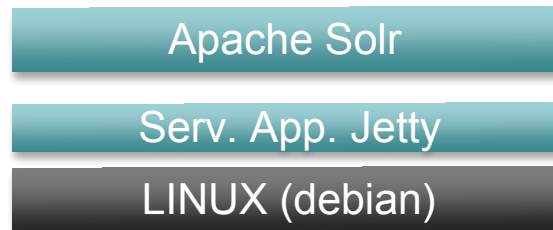
- ✓ Lo suficiente para complacer la impaciencia del usuario y siempre por encima de nuestra competencia.
- ✓ Llegar al punto óptimo es muy costoso, hay un umbral donde sólo se puede mejorar a costa de pasar mucho tiempo reescribiendo código con algoritmos más complejos.
- ✓ Pero... El precio actual de los servidores es menor que el de desarrolladores.

Arquitectura de servidores en lanetro

frontales



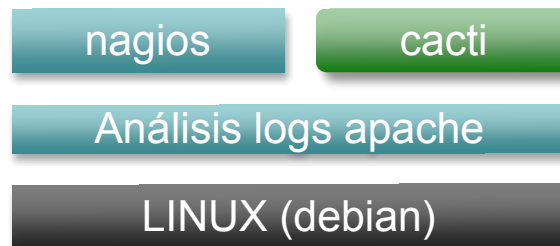
búsqueda



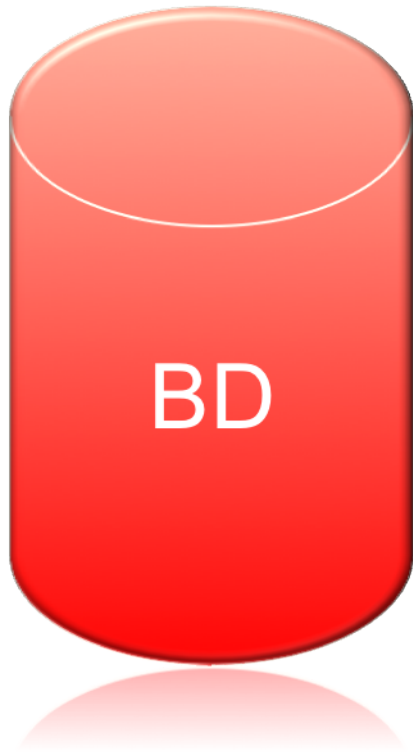
bd



monitor



El punto clave para escalar Drupal es la BD



Por su estructura modular y de sobre escritura de funciones y consultas, Drupal no es muy amigable con los motores de base de datos.

Cada módulo necesita realizar muchas consultas sobre las tablas base de Drupal (node, users, taxonomy, etc.) que resultan en una sobrecarga.

¡Hay que liberarle carga a la BD!

- ✓ Cachear todos los contenidos que no cambian. El motor de almacenamiento por excelencia es Memcached.
- ✓ Configurar la BD para un rendimiento óptimo, no usar configuración por defecto.
- ✓ Prevenir los bloqueos de base de datos. En muchos escenarios no se ha llegado al límite real de la capacidad de la base de datos si no que se ha desviado la relación entre escrituras y lecturas.

Caché del núcleo de Drupal

- Un gran problema de la caché del núcleo de Drupal es que no tiene reglas personalizables y que no es consistente con usuarios autenticados. ¡Lo advierte la cache agresiva!
- Si lo has hecho bien, muchas piezas estarán desarrolladas como bloques y podrás beneficiarte de la caché de bloques.

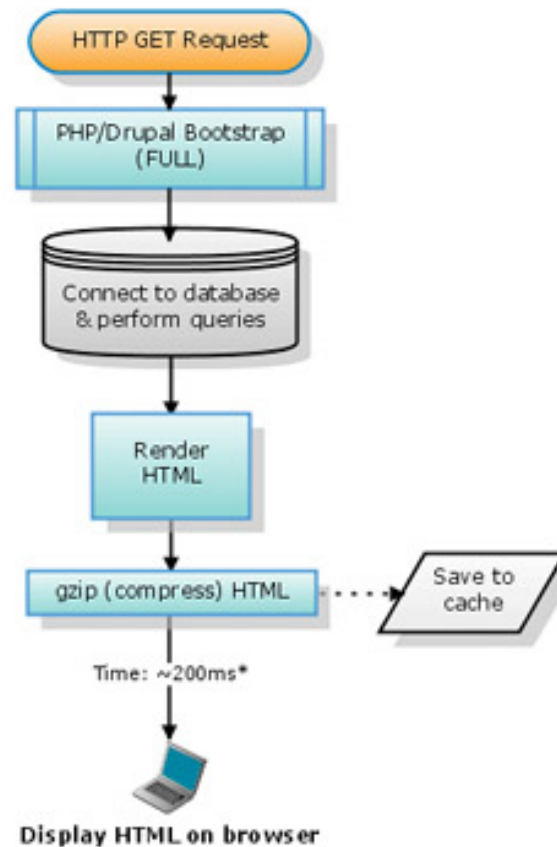
<http://drupal.org/project/blocktools>

Authcache, una cache mejorada

- Ofrece la funcionalidad de cachear para usuarios anónimos o autenticados lo que permite a Drupal servir las peticiones en 1-2 milisegundos y liberar la carga de los servidores.
- No interfiere con módulos de estadística de Drupal pues provee un hook AJAX que dispara las funciones con las que damos seguimiento a ciertos eventos. `hook_exit` por ejemplo.
- Permite crear reglas complejas de cache según expresiones regulares de rutas (paths).

Authenticated User Page Caching (Authcache)

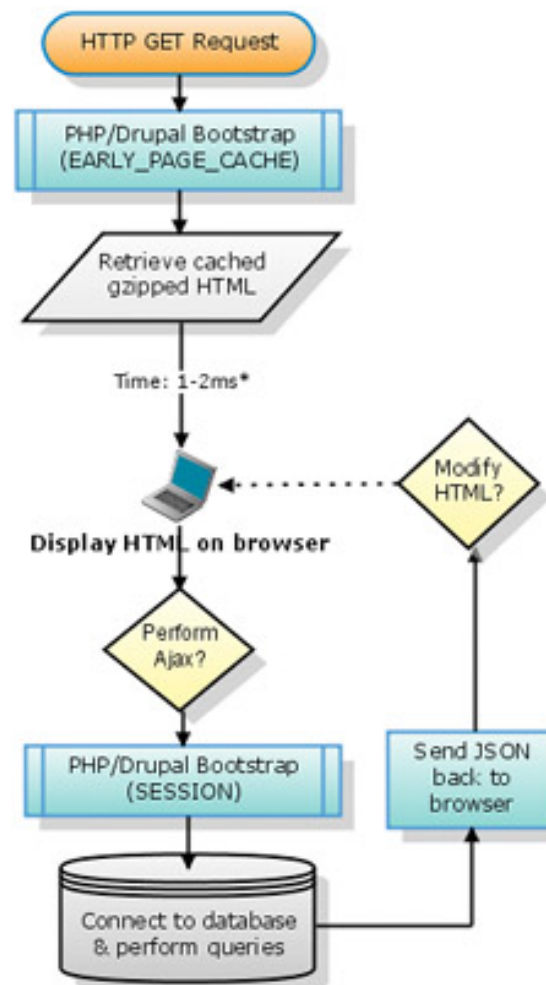
Non-Cached Page Request



* Millisecond times are for demonstrative purposes only. Actual times depend on webserver setup and page complexity

<http://drupal.org/project/authcache>

Cached Page Request

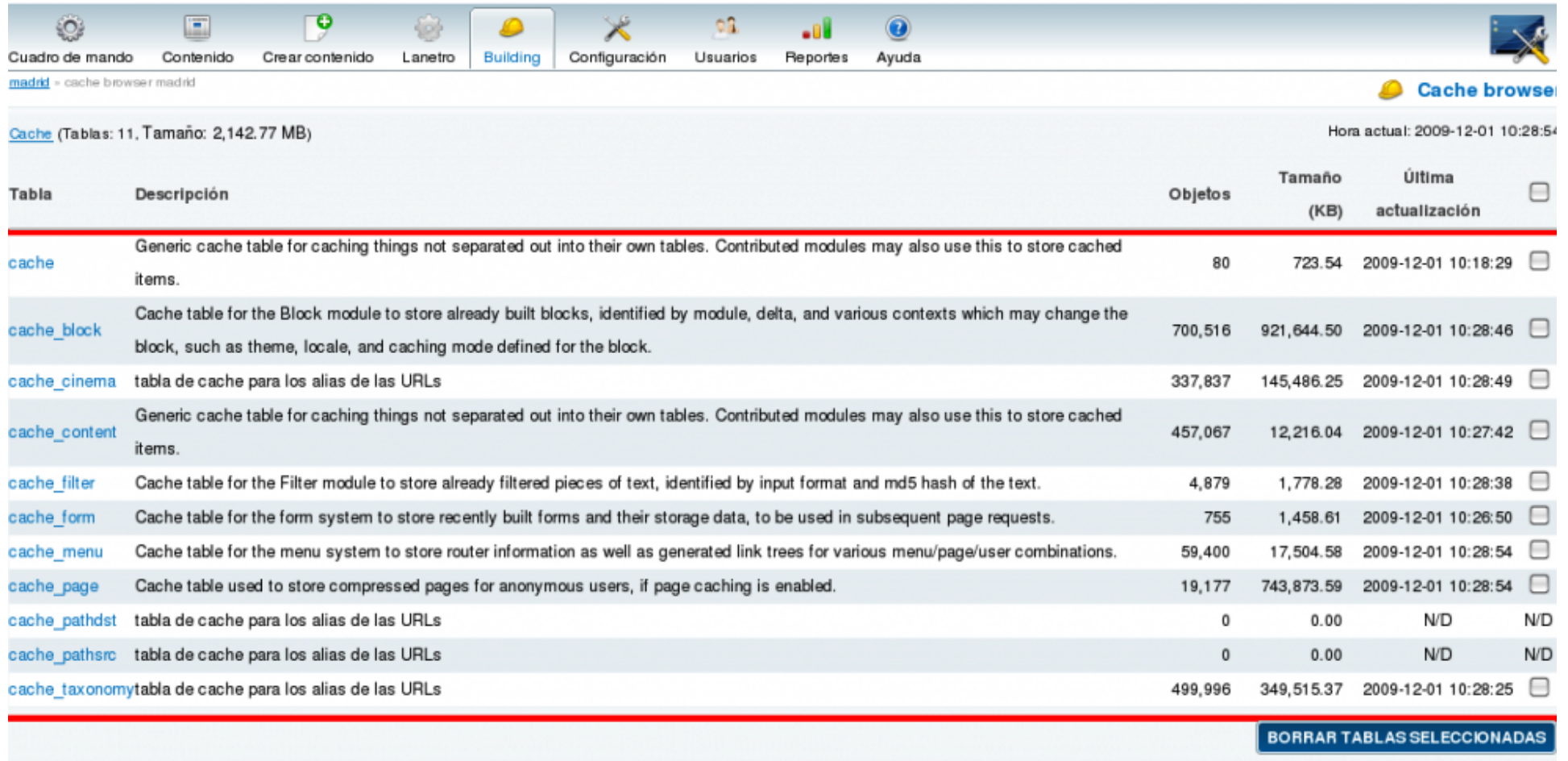


<http://drupal.org/project/authcache>

Invalidación de cache

- La cache es útil siempre y cuando no represente un desfase de la información, si está obsoleta hay que invalidarla.
- Caché browser es un módulo que nos permitirá navegar por un inventario de cache y resetear de forma manual, evitando el brusco y nada recomendable reseteo global.
http://drupal.org/project/cache_browser

Cache Browser, un navegador de cache



Cache browser

Cache (Tablas: 11, Tamaño: 2,142.77 MB) Hora actual: 2009-12-01 10:28:54

Tabla	Descripción	Objetos	Tamaño (KB)	Última actualización	
cache	Generic cache table for caching things not separated out into their own tables. Contributed modules may also use this to store cached items.	80	723.54	2009-12-01 10:18:29	☐
cache_block	Cache table for the Block module to store already built blocks, identified by module, delta, and various contexts which may change the block, such as theme, locale, and caching mode defined for the block.	700,516	921,644.50	2009-12-01 10:28:46	☐
cache_cinema	tabla de cache para los alias de las URLs	337,837	145,486.25	2009-12-01 10:28:49	☐
cache_content	Generic cache table for caching things not separated out into their own tables. Contributed modules may also use this to store cached items.	457,067	12,216.04	2009-12-01 10:27:42	☐
cache_filter	Cache table for the Filter module to store already filtered pieces of text, identified by input format and md5 hash of the text.	4,879	1,778.28	2009-12-01 10:28:38	☐
cache_form	Cache table for the form system to store recently built forms and their storage data, to be used in subsequent page requests.	755	1,458.61	2009-12-01 10:26:50	☐
cache_menu	Cache table for the menu system to store router information as well as generated link trees for various menu/page/user combinations.	59,400	17,504.58	2009-12-01 10:28:54	☐
cache_page	Cache table used to store compressed pages for anonymous users, if page caching is enabled.	19,177	743,873.59	2009-12-01 10:28:54	☐
cache_pathdst	tabla de cache para los alias de las URLs	0	0.00	N/D	N/D
cache_pathsrc	tabla de cache para los alias de las URLs	0	0.00	N/D	N/D
cache_taxonomy	tabla de cache para los alias de las URLs	499,996	349,515.37	2009-12-01 10:28:25	☐

BORRAR TABLAS SELECCIONADAS

ApacheSolr nos salvó la vida

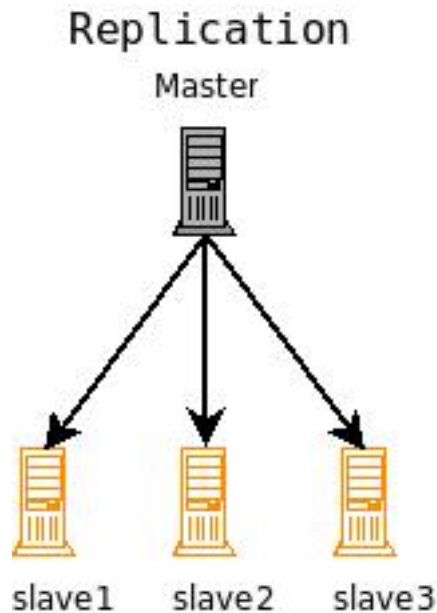
Drupal tiene integrada en el núcleo la función de búsqueda de contenidos pero cuando supera los 50,000 nodos e incrementa su tráfico es una gran carga para la BD.

ApacheSolr es la mejor solución de búsqueda, que además puede servir para generar cualquier tipo de listado o consultas de taxonomías.

ApacheSolr es extensible

- ✓ El módulo Apache Solr Search Integration ha tenido gran aceptación por la comunidad de Drupal y han surgido múltiples complementos.
- ✓ Se puede disponer de filtrado dinámico de resultados (facet search) con Apache Solr AJAX.
http://drupal.org/project/apachesolr_ajax
- ✓ Webservice para realizar búsquedas de los contenidos gestionados con Drupal e integrarlos con otras aplicaciones.
http://drupal.org/project/solr_service

ApacheSolr es escalable



- ✓ Se pueden disponer de múltiples servidores de lectura balanceados, lo que permite crecer de forma simple.

Prevención de bloqueos de BD

- Se ha desviado la relación entre escrituras y lecturas respecto a la media y las configuraciones por defecto de Mysql nos empiezan a degradar el funcionamiento.
- Un ejemplo común es que las tablas Mylsam están pensadas para recibir 95% de lecturas y 5% de escrituras.
- ✓ Hay que buscar métodos de escritura más eficientes como tablas InnoDB o Heap.

Optimización backend

- ✓ Identificar los datos que no cambian y que pueden guardarse en cache para no requerirlos repetidamente a la BD.
- ✓ Optimizar las consultas para asegurar que siempre se resuelven de la forma más rápida.
- ✓ Verificar que no faltan índices.

Una consulta SQL de muerte

```
SELECT qc.enid AS ID_EVENTO, qc.snid AS ID_SITIO,
       ns.title AS NOMBRE_SITIO, ne.title AS TITULO_EVENTO,
       vote.value as RANKING,
       fes.fecha AS FECHA,
       IF(res.horario IS NULL OR TRIM(res.horario)="", '--', res.horario) AS HORA,
       IF(res.precio IS NULL OR TRIM(res.precio)="", '--', res.precio) AS PRECIO,
       IF(res.titular IS NULL OR TRIM(res.titular)="", ne.title, res.titular) AS TEXTO_EVENTO,
       IF(res.resumen_alternativo IS NULL OR TRIM(res.resumen_alternativo)="", ln_e.title, res.resumen_alternativo) AS RESUMEN_EVENTO,
       tde.name AS CATEGORIA_EVENTO
FROM (".$IQueryCentral.$IJoins.$IWheres. ") qc
     INNER JOIN {node} ns ON qc.snid = ns.nid
     INNER JOIN {node} ne ON qc.enid = ne.nid
     INNER JOIN {ln_evento} ln_e ON ne.vid = ln_e.vid
     INNER JOIN {term_node} tne ON ne.vid = tne.vid
     INNER JOIN {term_data} tde ON tne.tid = tde.tid
     INNER JOIN {ln_rel_evento_sitio} res ON qc.relES_id = res.id
     INNER JOIN {ln_fecha_evento_sitio} fes ON qc.relES_id = fes.relid
     LEFT JOIN {votingapi_cache} vote ON vote.content_type='node' AND function=
WHERE fes.fecha >= %d AND fes.fecha <= %d AND tde.vid = %d
```

Una consulta compleja pero razonada

```
SELECT DISTINCT ncs.nid as id
  FROM {node_comment_statistics} ncs WHERE
  EXISTS
  (
    SELECT res.enid FROM {ln_rel_evento_sitio} res,
    {term_node} tne, {leftandright} lre, {term_node} tns,
    {leftandright} lrs
    WHERE ncs.nid=res.enid AND res.enid=tne.nid
    AND res.snid=tns.nid
    AND tne.tid=lre.tid AND tns.tid=lrs.tid
    AND lre.vid = 22 AND lre.lft >= %d AND lre.rgt <= %d
    AND lrs.vid = 11 AND lrs.lft >= %d AND lrs.rgt <= %d
  )
ORDER BY ncs.comment_count DESC LIMIT 0,50
```

Optimización frontend

- Yslow Firefox addon
<http://developer.yahoo.com/performance/rules.html>
- Minimizar las peticiones http (agregación de archivos, js, css, etc.)
- Orden de carga (css al principio, scripts al final)...
- CSS sprites, que los iconos se agrupen en una sola imagen para hacer una sola carga http.
- Muchas otras.

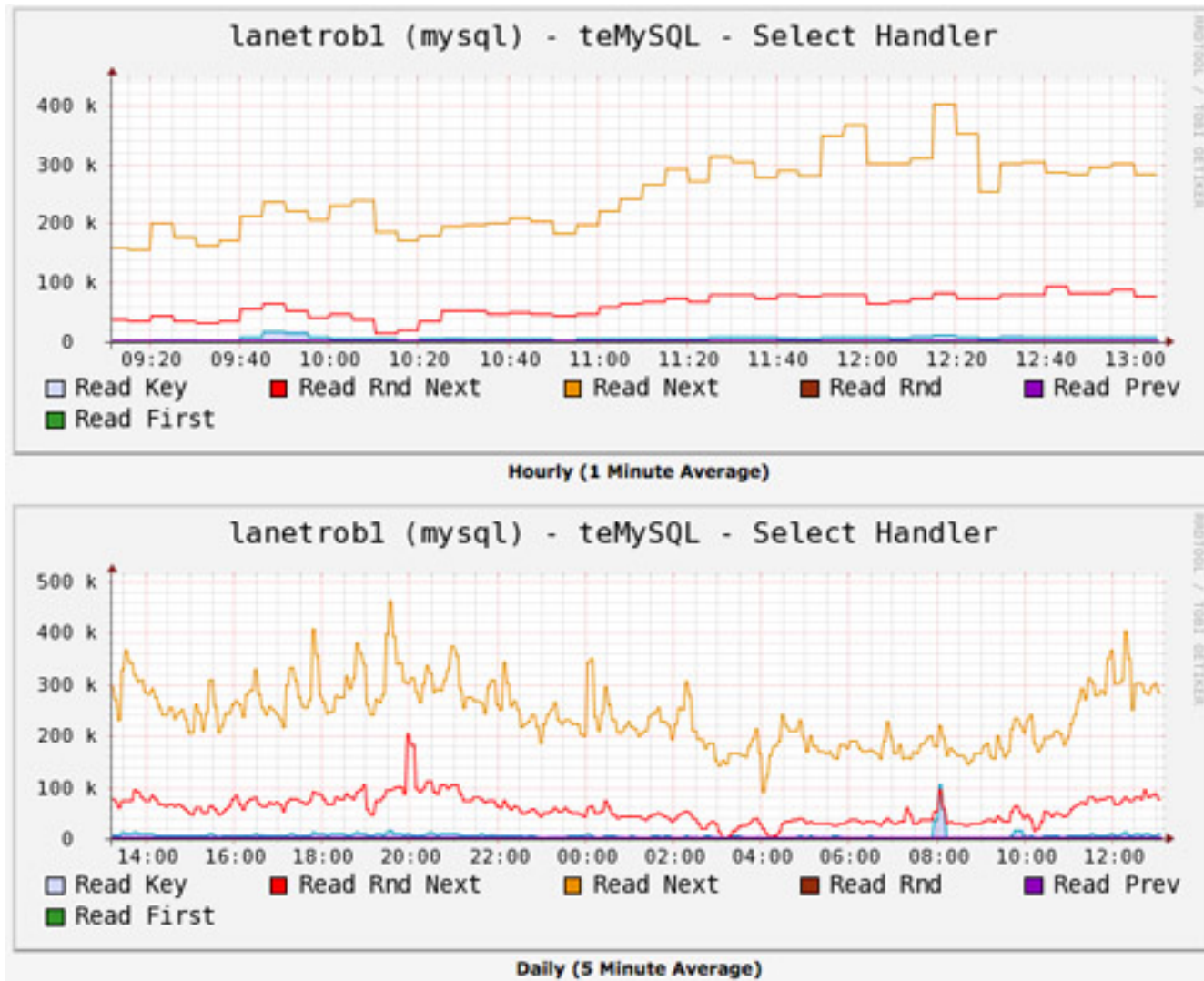
Optimizaciones en servidores frontales

- ✓ Apache mod_gzip - ahorro de transferencia y reducción de tiempo de descarga
- ✓ Cache opcode con APC, eAccelerator, Xcache, etc.
- ✓ Limitar la variable MaxClients para que no sobrepase el uso máximo de memoria de cada frontal.
- ✓ Hacer pruebas de carga con Jmeter, Nutch (simulando robots de buscadores).

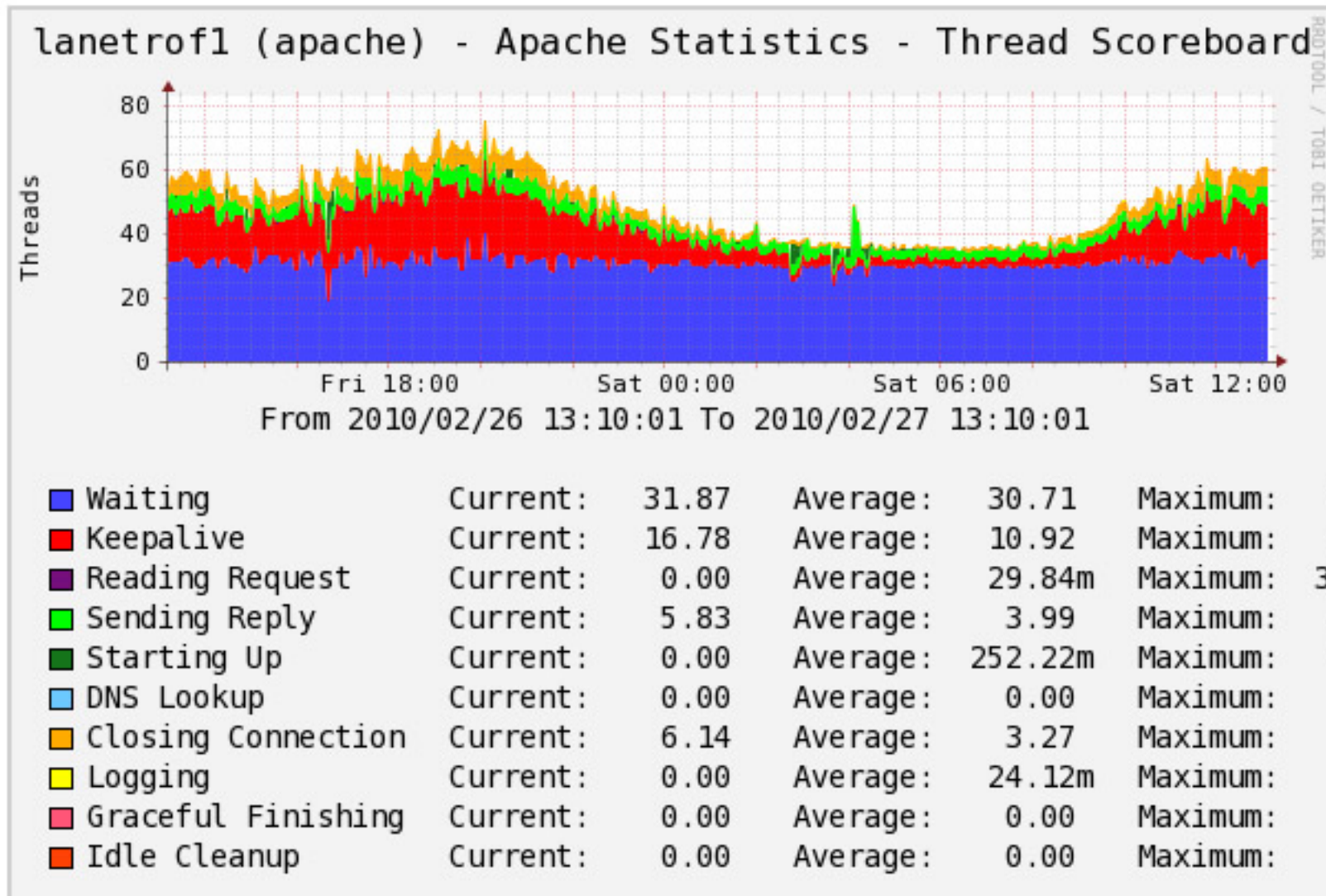
Monitoriza lo que pasa en tu arquitectura de servidores

- ✓ Probar, probar y probar
- ✓ Monitoriza cada capa y punto crítico
- ✓ Identifica los puntos de ruptura de tus servicios
- ✓ Si hay una consulta que se sale del umbral de rendimiento, verifica si hay índices apropiados, simplifica las consultas
- ✓ Cacti, Drupal Devel, EXPLAIN SELECT

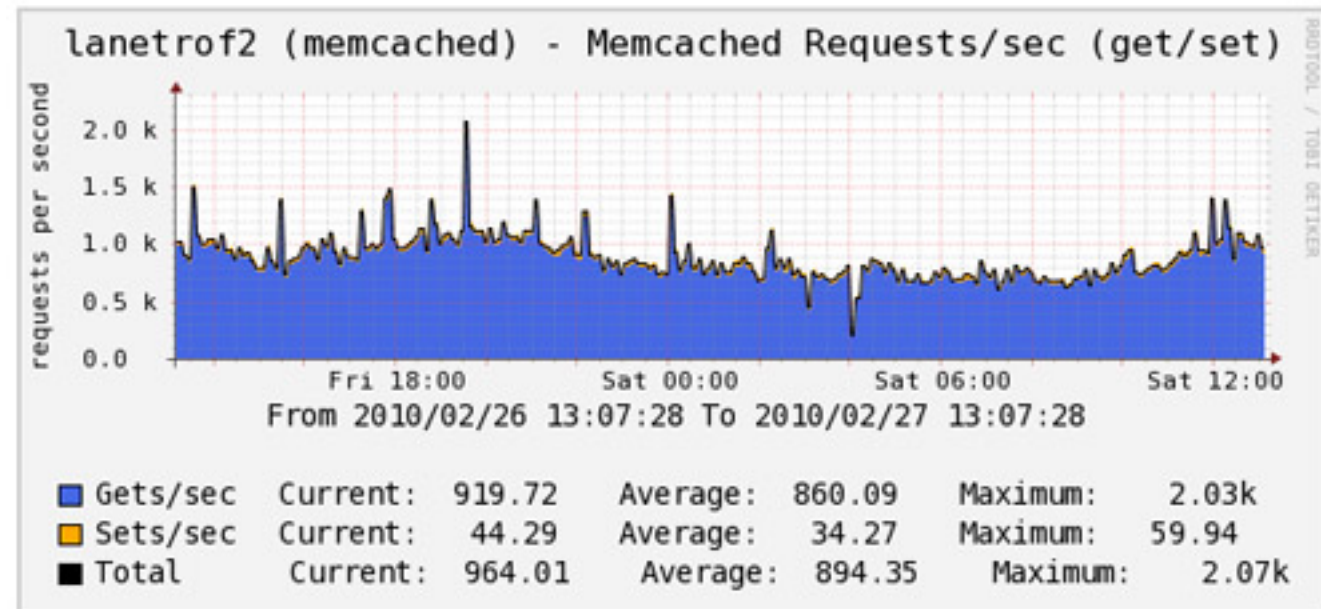
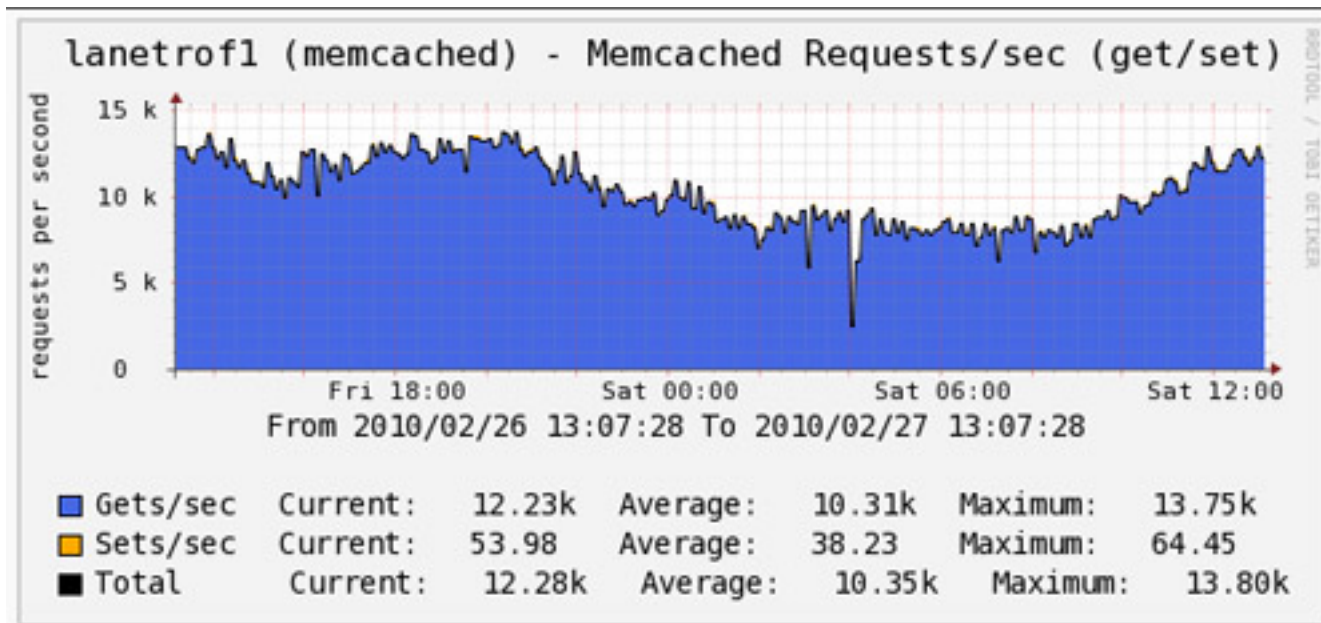
Cacti: monitorizacion Mysql



Cacti: monitorizacion Apache



Cacti: monitorizacion Memcached



Consejos de desarrollo sobre Drupal

- ✓ Apaláncate en el trabajo de toda la comunidad del software libre.
- ✓ Investiga, reutiliza, no reinventes la rueda. Si vas a innovar hazlo en un campo que nadie lo haya hecho.
- ✓ Mantén equipos reducidos con talento, ingenio y conocimiento. Persigue grandes hitos en menos tiempo.

Más consejos ...

- ✓ Involúcrate a todos los niveles del ciclo de desarrollo de las aplicaciones para tener una visión global de la arquitectura.
- ✓ Automatiza todas las tareas repetitivas y utiliza todas las herramientas de desarrollo y de gestión de proyecto disponibles que te ahorren tiempo.
- ✓ Así tu proyecto será más competitivo, innovador y apasionado.

GRACIAS ¿Preguntas?

Artículos que extienden lo que hemos visto:

<http://www.tratonera.com/?q=blog/3>

Pablo Cerda

Director de desarrollo de **ideup!**

www.ideup.com

Drupalcamp Spain 2010

ideup!