

Monitor de Estadísticas de IDECanarias

Deepak P. Daswani¹, J. J. Rodrigo¹ y J. Rosales²

¹Depto. de Ingeniería
GRAFCAN. Cartográfica de Canarias, S.A
C/ Panamá 34, Naves 8 y 9
Santa Cruz de Tenerife
{ddaswani, jrodrigo}@grafcan.com

²Dtor. Técnico
GRAFCAN. Cartográfica de Canarias, S.A
C/ Panamá 34, Naves 8 y 9
Santa Cruz de Tenerife
jrosales@grafcan.com

Resumen

En este documento se recoge toda la información correspondiente al Monitor de Estadísticas de IDECanarias. Se expone detalladamente una descripción del aplicativo y sus funcionalidades, analizando con precisión toda la información que proporciona acerca de una IDE e incidiendo en los aspectos técnicos intrínsecos a la implementación del mismo.

Palabras clave: JIDEE2008, Monitor de Estadísticas, WMS, IDECanarias,.

1 Introducción

Un sistema de información, de una magnitud y dimensión considerables, debe llevar asociado una herramienta que permita monitorizar y auditar con un nivel preciso de detalle, todos aquellos aspectos que derivan de la explotación del mismo. Esta afirmación cobra aún mayor trascendencia cuando el sistema a auditar

en cuestión es una Infraestructura de Datos Espaciales, tal y como representa IDECanarias [1].

La necesidad de registrar, analizar y monitorizar el consumo de los distintos servicios WMS [2] ofrecidos por IDECanarias, en aras de garantizar su correcto uso y disponibilidad, así como poder descubrir y depurar posibles errores o situaciones anómalas y mantener con ello la calidad de los mismos, motiva la aparición del Monitor de Estadísticas de IDECanarias.

2 Descripción y funcionalidades

El resultado es una aplicación web, accesible desde Internet, que permite obtener diariamente para cada uno de los servicios publicados en IDECanarias, no sólo estadísticas de uso general, que incluyen entre otras el número de accesos registrados, tráfico generado en cada servidor, número de visitantes distintos, distribución de usuarios por franja horaria, tiempos de respuesta máximos y mínimos del servidor... así como también estadísticas de uso particular para el caso de una IDE analizadas desde el punto de vista geográfico del servicio WMS, como por ejemplo clasificación de acceso por sistemas de referencia, tipo de peticiones WMS, acceso por capas, tamaño de las imágenes solicitadas, formato de las imágenes, zonas del territorio accedidas clasificadas por islas... que provienen del análisis y explotación de los diferentes parámetros que se incluyen en la cadena de cada petición WMS solicitada al servidor.

Como cabe esperar, todas estas estadísticas que la aplicación presenta al usuario están estructuradas por servicios y pueden consultarse de manera global para analizar la actividad de IDECanarias en conjunto o por cada servicio en particular, pudiendo consultar las mismas por períodos determinados de días, meses o años. Asimismo, existe una versión de administración orientada a una consulta más técnica, que además permite consultar las estadísticas de cada servicio para un servidor de mapas determinado. De este modo puede monitorizarse el estado y actividad de cada servicio en un servidor concreto y obtener medidas de rendimiento que de otro modo serían difíciles de percibir.

El aplicativo utiliza como fuente primera de datos los ficheros de registro de los servidores web donde se hallan alojados los distintos servidores de mapas de IDECanarias. En dichos ficheros se almacena toda la información de acceso generada como consecuencia de la navegación de los usuarios del exterior, pues

cada petición de servicio WMS se corresponde con una petición HTTP [3] en la que los distintos parámetros van incluidos en la cadena URL que se solicita al servidor. La información de acceso es recogida siguiendo la estructura característica de ficheros de registro del servidor web en cuestión, en este caso Apache [4]. Dichos ficheros de registro son sometidos a un tratamiento específico implementado mediante un algoritmo a medida para poder extraer correctamente no sólo cada uno de los campos que vienen recogidos en el fichero original (como pueden ser fecha y hora de la petición, código de respuesta HTTP, agente, bytes recibidos...) sino también analizar y separar cada uno de los parámetros de la cadena WMS incluidos en la petición solicitada al servidor. La información generada es cargada diariamente en un motor de bases de datos y procesada posteriormente en diferentes pasos para generar los datos que permiten explotar los diferentes tipos de estadísticas presentados en el aplicativo web final.

3 Estadísticas de acceso a Servicios WMS

He aquí una pequeña muestra del aspecto del Monitor de Estadísticas IDECanarias.

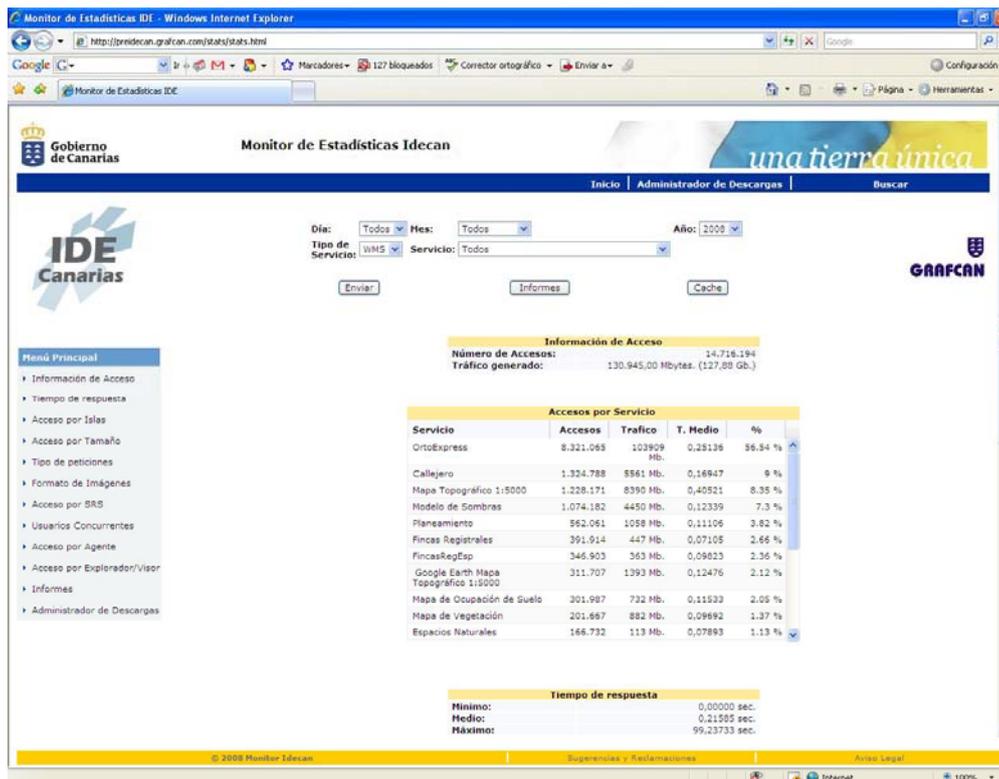


Figura 1. Aspecto visual del Monitor de Estadísticas IDECanarias

En la pantalla inicial, al arrancar el aplicativo en el navegador, se cargan por defecto las estadísticas globales que muestran la actividad correspondiente a todos los servicios de IDECanarias contabilizadas desde el día su arranque hasta la fecha actual.

Podemos ver como primer dato la información del número de accesos y el tráfico generado en los servidores como consecuencia de la navegación de los usuarios. En esta opción general, además de la información que se muestra para cada servicio en particular y que detallaremos a continuación, se incluye un cuadro resumen que clasifica y ordena la actividad global por servicios, accesos, tráfico, tiempo medio de respuesta, y porcentaje que representa el servicio en la actividad de IDECanarias. Esto es muy significativo pues permite obtener rápidamente una visión general de la actividad registrada.

Accesos por Servicio				
Servicio	Accesos	Trafico	T. Medio	%
OrtoExpress	8.321.065	103909 Mb.	0,25136	56.54 %
Callejero	1.324.788	5561 Mb.	0,16947	9 %
Mapa Topográfico 1:5000	1.228.171	8390 Mb.	0,40521	8.35 %
Modelo de Sombras	1.074.182	4450 Mb.	0,12339	7.3 %
Planeamiento	562.061	1058 Mb.	0,11106	3.82 %
Fincas Registrales	391.914	447 Mb.	0,07105	2.66 %
FincasRegEsp	346.903	363 Mb.	0,09823	2.36 %
Google Earth Mapa Topográfico 1:5000	311.707	1393 Mb.	0,12476	2.12 %
Mapa de Ocupación de Suelo	301.987	732 Mb.	0,11533	2.05 %
Mapa de Vegetación	201.667	882 Mb.	0,09692	1.37 %
Espacios Naturales	166.732	113 Mb.	0,07893	1.13 %

Figura 2. Cuadro resumen general de acceso a servicios WMS

El resto de información que el Monitor de Estadísticas de IDECanarias presenta para cada servicio o para el conjunto de todos los servicios en general se describe en los apartados que vienen a continuación.

3.1 Tiempo de Respuesta

Se presenta una tabla con los cálculos de tiempos de respuesta mínimos, medio y máximo ofrecidos por los servidores para el servicio en cuestión, y se obtiene directamente de los ficheros de registro.

Tiempo de respuesta	
Mínimo:	0,00000 sec.
Medio:	0,21585 sec.
Máximo:	99,23733 sec.

Figura 3. Estadísticas de Tiempo de Respuesta

3.2 Tamaño de Imágenes

En este apartado se muestra una tabla con los tamaños de imágenes mínimos, medios y máximo solicitados en las peticiones *getmap* para el servicio consultado. Esta información se obtiene analizando los parámetros WIDTH y HEIGHT que se especifican en la petición WMS, disponible en la cadena de parámetros de la URL registrada.

Tamaño de Imágenes		
Tamaño	Width	Height
Minimo:	0	0
Medio:	269	265
Máximo:	21.151	13.739

Figura 4. Estadísticas de Tamaño de Imágenes

3.3 Acceso por Islas

Este apartado se obtiene a partir de la explotación del parámetro BBOX (*Bounding Box*) especificado en la petición WMS, y que determina la región geográfica a consultar. Cruzando la información de las coordenadas geográficas solicitadas a los servidores de mapas con los límites en coordenadas geográficas de las islas del territorio canario, podemos obtener el número de accesos que se registran para cada isla. Podemos ver un ejemplo en la imagen que se muestra a continuación:

Acceso por Islas	
El Hierro	226.125
Fuerteventura	825.528
Gran Canaria	2.822.248
La Gomera	331.156
La Palma	903.375
Lanzarote	802.581
Tenerife	4.580.290

Figura 5. Estadísticas de acceso a servicios WMS por islas

3.4 Tipo de peticiones:

Esta opción muestra una tabla con el número de accesos que se han registrado clasificados por tipo de peticiones WMS, que vienen determinadas por el parámetro REQUEST, pudiendo ser éstas no sólo *getmap* (la mayor parte que son las peticiones de imágenes), sino también *getcapabilities*, *getfeatureinfo* o *getlegendgraphics*.

Tipo de peticiones	
getcapabilities	15.766
getfeatureinfo	20.648
getlegendgraphic	14.988
getmap	13.099.024

Figura 6. Estadísticas de acceso por tipo de peticiones REQUEST

3.5 Formato de imágenes

Una clasificación que muestra el número de peticiones *getmap* para el servicio consultado clasificadas por formato solicitado. Este dato se extrae analizando el parámetro FORMAT de la cadena de parámetros WMS en la URL.

Formato de Imágenes	
GIF	776.393
JPEG	16.394.962
PNG	3.996.978

Figura 7. Formato de imágenes consultadas

3.6 Acceso por Sistema de Referencia

En este apartado se muestra una clasificación de accesos por Sistemas de Referencia solicitado, que se obtiene de extraer el parámetro SRS de la cadena de parámetros WMS.

Acceso por Sistema de Referencia	
4326	12.707.114
32627	424
32628	408.128

Figura 8. Estadísticas de acceso por sistema de referencia SRS

3.7 Peticiones Concurrentes

Este punto muestra una clasificación de los 10 instantes de tiempo (con precisión de segundo) con mayor número de peticiones registradas en los servidores para el servicio consultado. Este dato se obtiene agrupando el número de peticiones registradas por la unidad mínima de tiempo que nos permite obtener el servidor, que en este caso es el segundo.

Peticiones Concurrentes	
Nº Peticiones	Fecha
281	2008-07-10 09:16:59
219	2008-06-27 07:52:04
196	2008-06-27 07:52:39
186	2008-07-10 09:16:49
171	2008-05-28 15:36:58
149	2008-05-28 09:56:48
140	2008-06-26 08:25:05
137	2008-06-26 08:26:35
135	2008-07-10 09:16:55
132	2008-06-10 14:30:05

Figura 9. Peticiones concurrentes de acceso a servicios WMS

3.8 Acceso por agente

Cada petición HTTP lleva asociada consigo un agente que permite obtener información de su procedencia, y que queda registrado en los ficheros de registro

del servidor. Generalmente los sitios web incluyen la información de su dominio en el agente, con lo que podemos establecer una clasificación de los sitios web que más utilizan nuestros servicios a partir de este campo.

Acceso por Agente	
http://visor.grafcan.com/	7.342.260
Indefinido	3.482.952
http://visor.grafcan.es/	2.826.582
http://www.goolzoom.com/	481.366
http://213.172.37.189:8080/	98.347
http://localhost/	68.437
http://www.cenit-consulting.com/	44.343
http://www.clubuma.net/	28.589
http://www.ideo.es/	13.909
http://wms-sites.com/	12.345
http://localhost:8080/	6.843
http://goolzoom.com/	6.600
http://192.168.0.101/	2.070
http://www.localdev.com/	253
http://161.67.130.131/	201
http://192.168.0.224:8080/	189
http://www.google.es/	186
http://idecan3.grafcan.es/	129
http://161.67.130.130/	125
http://localhost:1344/	89
http://ideo.unizar.es/	86
http://213.96.167.3/	73
http://mail.google.com/	40
http://www.wmsviewer.com/	30
http://visorpre.grafcan.es/	28
http://g5ccm/	22
http://www.cienladrillos.com/	18
http://www.google.com/	17
http://es.search.yahoo.com/	17
http://www.ogc-services.net/	13

Figura 10. Acceso por agente a servicios WMS

3.9 Acceso por visor / explorador

Al igual que el para el caso anterior, en cada petición HTTP existe un campo *referer*, en el que se registra información sobre el navegador que ejecuta la petición. Analizando exhaustivamente se pueden apreciar ciertos patrones introducidos por los navegadores web y también visores WMS, con lo que podemos clasificar los accesos a los servicios según el explorador / visor desde el cual son consumidos.

Acceso por Explorador/Visor	
Internet Explorer 7.0	5.223.579
Internet Explorer 6.0	4.252.842
Firefox 2.0	1.374.230
GoogleEarth	673.280
Firefox 3.0	407.172
Global Mapper	131.067
ArcMap Service Layers	26.277

Figura 11. Estadísticas de acceso por explorador

4 Monitorización de la Caché

Además de estas estadísticas, se ha desarrollado para el Monitor de IDECanarias una sección especial para analizar y auditar el comportamiento de la caché bajo demanda creada para algunos servicios, ya que es de vital importancia conocer el número de aciertos (*hits*), fallos (*miss*) y redirecciones por servicio, así como los tiempos de respuesta medios obtenidos al servir las imágenes cacheadas en cada caso. Con estos datos podemos calcular el porcentaje de aciertos en la caché, que debe ir aumentando conforme avanzan los días desde la publicación de la información.

En este apartado de monitorización de la caché se incluye además la opción de filtrar por sede, para auditar de forma individual los servidores de caché y obtener una medida de rendimiento más ajustada para cada caso.

He aquí un ejemplo de la información mostrada en la sección de monitorización de la caché.

Información de Acceso	
Número de Accesos:	9.333.606
Hits en Caché:	8.315.634
Miss en Caché:	1.677.316
Redirecciones en Caché:	61.174
Tiempo Medio de Hits:	0,08640 sg.
Tiempo Medio de Miss:	0,55320 sg.
Aciertos en Caché:	82.7086874998 %
Tráfico generado:	84.955,61 Mbytes. (82,96 Gb.)

Figura 12. Estadísticas generales de Acceso a la Caché

Al igual que para la sección de estadísticas principal, se muestra un cuadro resumen con la información de todos los servicios cacheados en la opción general que consulta todos los servicios de forma global.

Accesos por Servicio					
Servicio	Accesos	Hits	Miss	Redirect	%
OrtoExpress	6.137.415	5.131.550	983.558	22.101	83.61 %
Callejero	796.046	574.521	215.069	6.450	72.17 %
Modelo de Sombras	793.350	630.176	162.932	174	79.44 %
Mapa Topográfico 1:5000	587.134	383.434	176.583	27.037	65.31 %
Google Earth Mapa Topográfico 1:5000	367.586	367.322	0	264	99.93 %
Google Earth Callejero	194.041	193.966	0	75	99.96 %
Mapa de Ocupación de Suelo	149.305	91.351	56.915	1.033	61.19 %

Figura 13. Cuadro resumen general de Acceso a la Caché por Servicios WMS

Otra funcionalidad especial que se ha añadido a la sección de monitorización de la caché es la posibilidad de visualizar un árbol de carpetas al estilo explorador de Windows que permite recorrer la estructura de directorios de caché y observar las ramificaciones y carpetas que se han ido creando en la misma, así como el número de ficheros generados y el tamaño ocupado por los mismos. Conforme va avanzando el tiempo, se hace más difícil analizar y monitorizar la estructura de directorios de la caché de forma manual, pues estos crecen de manera exponencial.

El árbol se genera de forma dinámica, cargando la información de un nodo una vez que se pincha sobre el mismo, pues en el caso de cargar toda la información de una sola vez, la página podría tardar infinidad de tiempo en cargarse, teniendo en cuenta que el tamaño de la estructura de directorios crece con el paso del tiempo.

Asimismo, se añade la opción de poder consultar el árbol de cada sede de forma individual, y de este modo controlar si se han producido errores o divergencias en la generación de la estructura en los distintos servidores.

5 Sección de Informes

Esta sección se ha contemplado para registrar y auditar estadísticas particulares de distintos servicios / informes que puedan irse añadiendo a IDECanarias con el paso del tiempo, y que no entren dentro de la monitorización general descrita en apartados anteriores.

Por el momento, en esta sección se puede auditar el número de accesos a informes de consulta urbanística generado, para poder así comprobar el grado de seguimiento y aceptación que el servicio tiene por parte de los usuarios.

6 Aspectos Técnicos

Como se describe de forma somera en el apartado de descripción, el Monitor de Estadísticas de IDECanarias se nutre inicialmente de los ficheros de registro de los servidores web de mapas, que luego se transforman mediante un algoritmo desarrollado a medida para este caso particular.

Los servidores de IDECanarias están distribuidos en dos sedes, una en la isla de Tenerife y otra en la de Gran Canaria. Esta arquitectura de sistemas supone tener 8 servidores por sede, con lo cual hablamos de un total de 16 servidores a monitorizar.

Para llevar a cabo esta tarea es preciso coordinar correctamente todos los pasos del proceso y sincronizar la hora de ejecución de los mismos, pues es importante respetar la precedencia para la obtención final de las estadísticas.

Los ficheros de registro de los servidores web son utilizados como entrada para un algoritmo codificado en Perl que parsea y extrae tanto los campos que vienen en dichos fichero como los parámetros WMS de la cadena de parámetros de la URL. Además el algoritmo añade el campo de servidor cogiendo automáticamente la dirección IP de la máquina en la que se ejecuta el programa. De este modo se puede tener correctamente identificados los registros de cada servidor, y el añadir o quitar un servidor no supone una tarea crítica para la coordinación del proceso. Para correr al algoritmo se utiliza un script de shell en los servidores Linux y un ejecutable BAT en los servidores Windows, que se programa para su ejecución diaria. Este script se encarga básicamente de las siguientes tareas:

- Comprobar si existen usuarios en el servidor web
- Esperar a que las peticiones de los usuarios finalicen
- Parar el servidor Web
- Copiar el fichero de registro en un histórico para mantener una copia de seguridad bruta en un lugar seguro.
- Rearrancar el servidor Web.
- Ejecutar el algoritmo en Perl sobre el fichero de registro y guardar la salida en una carpeta repositorio.
- En el caso de los servidores Linux copiar la salida del algoritmo a una carpeta repositorio en un servidor Windows.

Una vez la salida del algoritmo se encuentra en la carpeta repositorio de los servidores correspondientes, un servicio de integración del motor de bases de datos SQL Server 2005 procede a utilizar ese fichero para cargar la información correspondiente en una tabla. Existe un servidor SQL Server 2005 [5] en cada sede, pues de lo contrario el trasvase de los ficheros de salida podría saturar las líneas de comunicación entre islas. Recordemos que cada fichero de salida tiene una línea por petición, y diariamente pueden cargarse en la tabla de la base de datos más de 300.000 registros.

Después de haber cargado los datos en la tabla, se realiza un proceso de replicación entre los datos de los servidores SQL de ambas islas.

Posteriormente, se ejecutan sobre el servidor SQL Server en cuestión una serie de procedimientos almacenados que se encargan de precalcular las estadísticas mostradas finalmente en la aplicación web. Las estadísticas han de ser pre-generadas pues realizar todas las consultas que se requieren para mostrar las estadísticas en tiempo de ejecución sobre una tabla con más de 15 millones de registros actualmente no es viable para el usuario final que se sienta ante la

aplicación. Y más aún teniendo en cuenta que las consultas requieren de agrupaciones y filtrado de diversos campos, combinaciones, y en más de algún caso, búsquedas de subcadenas, lo cual incrementa más aún el tiempo de procesamiento de cada consulta.

Para precalcular las estadísticas se utilizan diversas tablas de parámetros con los valores discretos que se van a almacenar para sistemas de referencia, explorador, formato de imágenes, tamaños de imágenes, agente.... Se guardan las estadísticas de cada día clasificadas por día, mes y año y estos valores precalculados luego se combinan para construir las consultas finales en los períodos de tiempo seleccionados por el usuario en la aplicación final.

Además, por otra parte, en los servidores que alojan la caché se lanza un proceso en Pitón que devuelve la estructura de directorios de caché en un fichero XML que es utilizado posteriormente por la aplicación para construir el árbol de nodos que vimos en el apartado anterior.

Todos estos pasos descritos en los párrafos superiores se ejecutan diariamente y han de sincronizados debidamente en el tiempo para dar lugar a la generación de las estadísticas finales.

Por último, la aplicación web final del Monitor de Estadísticas hace uso de las tablas finales de estadísticas precalculadas para realizar las consultas pertinentes , que ya son directas y devuelven los resultados en un tiempo más que razonable para el usuario. Esta aplicación está desarrollada utilizando HTML, CSS para la presentación de estilos, PHP para la lógica de consultas, y Javascript con Ajax para los cuadros resumen y el árbol de directorios.

Referencias

- [1] IDECanarias, <http://www.idecan.grafcan.es/idecan>
- [2] Web Map Service (WMS), <http://www.opengeospatial.org/standards/wms>
- [3] HTTP Protocol, <http://www.w3.org/Protocols/>
- [4] The Apache Software Foundation, <http://www.apache.org>
- [5] SQL Server 2005, <http://www.microsoft.com/spain/sql/default.msp>