La Nueva Arquitectura de la IDE Andalucía.

Actualización tecnológica al servicio de un proyecto Integrador

GUTIERREZ José Angel; MARTIN, Abel; MOLINA Antonio; PARDO Emilio; VILLAR Agustín.

Desde sus orígenes en la año 2002 la IDE Andalucía, ha venido prestando servicios de información geográfica, en el marco del desarrollo promovido por la Directiva Inspire y en coordinación con el resto de iniciativas estatales, impulsadas desde el Consejo Superior Geográfico.

Tecnológicamente, el proyecto ha ido evolucionado desde un contexto donde predominaban las máquinas físicas a otro en la que la evolución tecnología ha permitido el despliegue en entornos virtualizados.

Por su parte en el ámbito del software (servidores de aplicaciones, servidores de mapas, web Mapping....) ha habido una decidida apuesta por el software libre que ha aportado un importe know how al conjunto del Sistema Estadístico Cartográfico de Andalucía en esta materia.

Los resultados arrojan cifras elocuentes del impacto social que ha ido tomando el proyecto; de las 10.856 vistas del año 2009, se ha pasado a las 596.254 visitas del año 2014, haciendo de la IDE Andalucía una de las infraestructuras más activas del conjunto del Estado. Sin duda, junto a datos y servicios, la existencia de una infraestructura física que garantiza los niveles de accesibilidad y disponibilidad que establecen los reglamentos que desarrollan la Directiva Inspire en esta materia, han tenido mucho que ver.

El nuevo marco institucional, pionero en España, de integración de la Estadística y la Cartografía en el nuevo Instituto de Estadística y Cartografía, ha requerido el diseño y despliegue de una nueva arquitectura, que sobre la base de lo ya consolidado, fuera capaz de proporcionar servicios de calidad a nuevos proyectos como el Callejero Digital de Andalucía Unificado y a los nuevos paradigmas de servicios de información en la red que proporciona esta nueva situación institucional.

Aunque las arquitecturas virtualizadas, no son ninguna novedad, si lo son determinadas configuraciones aplicadas en el despliegue abordado en el IECA. Este artículo presenta el diseño de esta nueva arquitectura, y su relación con datos y servicios, que ha entrado en funcionamiento a finales de septiembre de 2014

PALABRAS CLAVE

Arquitectura tecnológica, virtualización, software libre, servidores de mapas, web mapping, Infraestructura de Datos Espaciales, cluster,

INTRODUCCIÓN:

EL pasado 29 de septiembre ha entrado en servicio, la nueva infraestructura física de la IDE Andalucía. Hasta esa fecha los servicios se venían prestando desde SANDETEL, empresa pública de la Junta de Andalucía que tiene un papel destacado en el desarrollo de las TIC en el marco de la política tecnológica de la Junta de Andalucía. El nuevo escenario institucional en lo que a la producción y difusión de datos espaciales y estadísticos se refiere con la fusión de los Institutos de Estadística y Cartografía, así como el impulso decisivo que el Plan estadístico y Cartográfico, 2013-2017 ha dado al desarrollo de la Infraestructuras de Datos de Andalucía, ha desencadenado el desarrollo del proyecto de diseño y despliegue de una nueva arquitectura que, manteniendo los servicios actuales, permita dar cobertura a las previsiones de crecimiento de la IDE Andalucía en los próximos años

1

1. ANTECEDENTES

1.1. LOS ORÍGENES DE LA IDE ANDALUCÍA

Los primeros servicios de Información Geográfica en la Red proporcionados por la Junta de Andalucía se remontan al año 2002, una época en la que la incertidumbre por la ausencia de estándares en la publicación de servicios condicionaba el desarrollo de estos trabajos. Atlas Web de Andalucía, solicitud de Mapas y Fotos Aéreas, Andalucía desde el cielo...., son algunos ejemplos de esta época.

Con los primeros borradores de la Directiva INSPIRE y los estudios realizados por el grupo de trabajo de la Infraestructura de Datos Espaciales de España cobran impulso aspectos como el despliegue de una Infraestructura de datos acorde a las norma y estándares internacionales, y que en el plano institucional tenía su correlato en el Decreto 141/2006 de 18 de julio, que incorporaba al cuerpo normativo de la Junta de Andalucía la necesidad de desarrollar una Infraestructura de Datos Espaciales, para 'mejorar la calidad de la Información geográfica e impulsar los medios de difusión'. En 2006, salen a la red los primeros servicios interoperables de de la IDE Andalucía: servicios de visualización de 3 series de ortofotos, de 2 series de cartografía topográfica, y un servicio CSW con un primer catálogo de metadatos con 12.000 registros.

El desarrollo de esta primera infraestructura, estaba basado en software propietario, ya que no se tenían experiencias previas con las opciones de software libre. Respecto a la arquitectura física, su diseño era muy simple para evitar mantenimientos complejos del servicio. Finalmente la gestión del servicio se encomendó a la empresa Pública SADESI (Hoy SANDETEL). Con estas premisas IDE Andalucía se abrió al público con dos servidores, un frront-end con ArcIMS 9.0 de ESRI e INDICIO de la compañía Galdos, desplegado sobre un SO Windows Server 2003; y un back-end con una base de datos Oracle 9i desplegada sobre Linux Debian para el almacenamiento de Metadatos.

Esta primera infraestructura se ve complementada durante 2007, con el desarrollo, en software libre, de un servicio de mapas vectorial (MTA 10:000) con MapServer (v4.10), sobre una BD Postgresql y la sustitución del visor de ArcIMS por MapBender, lo que supuso una modificación de la Arquitectura física con la incorporación de dos nuevos servidores para dar cabida al despliegue Mapserver/Mapbender y a la base de Datos Postgresgl. Se apuntaba así las nuevas tendencias que la Arquitectura de la IDE iba a tener en los próximos años.

1.2. LA IDE ANDALUCÍA V.2

La nueva "IDE Andalucía V2" empieza a cobrar cuerpo a finales de 2007 cuando se empieza a tomar conciencia, de que las decisiones sobre la arquitectura requieren un análisis más ajustado a los requerimientos de los estándares que se iban consolidando, estándares que por la vía de la Directiva INSPIRE, terminarían siendo de obligado cumplimiento. Así se identificaban dificultades para prestar servicios como recuperación y edición de objetos geográficos vectoriales (WFS, WFS-G, WFS-T) y raster (WCS) o servicios de procesado (WPS). Junto a ello el crecimiento de las peticiones de servicios amenazaba con limitar la calidad en la respuesta de los servidores. En positivo, se contaba en la Junta de Andalucía con una respuesta institucional acorde con el Decreto 141/2006, poniéndose en marcha, en colaboración con la Consejería de Innovación, el proyecto SIG-Corporativo con un enfoque de solución integral para la gestión de toda la información espacial de la Junta de Andalucía, y que contemplaba un conjunto de herramientas necesarias para el desarrollo implantación y gestión de la IDE Andalucía.

El análisis de requisitos de software y hardware que se hizo, dio como resultado una planificación de una nueva arquitectura lógica y física que se muestran en las figuras 1 y 2

Aquella arquitectura, contaba ya con 12 máquinas, 8 en el front-end y 4 en el back-end, cuyo objetivo a grandes rasgos era mejorar la estabilidad y el rendimiento del sistema, realizándose una división Lógica de los equipos en función de los servicios que se tenía previsto prestar:

Backend: Constituido por los equipos para servicios de acceso y almacenamiento de datos

Frontend: equipos que sirven las aplicaciones de cliente y procesamiento de datos, Mapserver, Servidor de metadatos, el Geoportal de la IDE Andalucía y el gestor de contenidos V Jornadas Ibéricas de Infra-estruturas de Dados Espaciais

para la web.

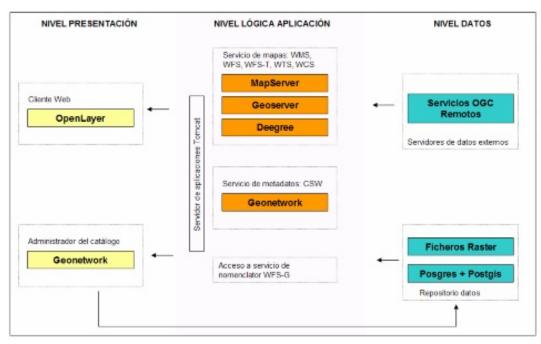


Figura 1 Arquitectura Lógica de la IDE Andalucía en 2008

IDEAndalucía - Arquitectura noviembre 2008

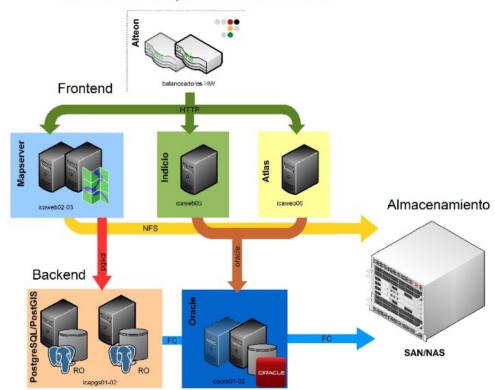


Figura 2 Arquitectura de la IDE Andalucía en noviembre de 2008

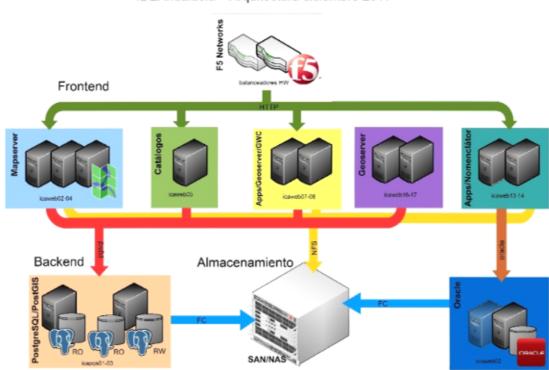
Esta división física permitió una mejor separación entre datos y aplicaciones y dentro de ellas según su función, cuyo efecto más importante fue independizar las distintas subsecciones de la V Jornadas Ibéricas de Infra-estruturas de Dados Espaciais

arquitectura, permitiendo una evolución más flexible con la sustitución de partes sin afecciones al resto de la arquitectura, siendo ésta más independiente de soluciones tecnológicas o proveedores concretos. Además permitía el análisis de rendimiento, la localización de errores, y la escalabilidad del sistema, implantando mejores políticas de gestión de copias de seguridad. Se tuvo en cuenta también la réplica de los equipos dedicados a servicios más críticos, lo cual mejoraba por un lado, la estabilidad del sistema y por otro mejoraba el rendimiento de los servicios que consumían más recursos o requería más ancho de banda.

La creación de subsecciones Ilevó consigo la agrupación de aplicaciones por tipos de servicios prestados, definiéndose las subcategorias de servicios de visualización (mapserver), servicios de descarga (geoserver), visualizadores (IDEAvisor, Visor 3D), herramientas (Creomapas, Mapshup, configurador de Mapshup), aplicaciones (Nomenclator, Line@, Situate, Cartografía Histórica), catálogos (Catálogo de Datos y Catálogo de Metadatos) y Geoportal.

A finales de 2011, el crecimiento de la demanda que había sufrido la IDE Andalucía, unido a cierta obsolescencia que empezaba a tener la infraestructura de servidores, hizo necesaria la adquisición de nuevo equipamiento hardware. En paralelo se realizaron las primeras operaciones de virtualización de servidores, que abarcó solo el front-end de la arquitectura.

Este proceso no contempló la actualización del software ni la alta disponibilidad de todos los servicios. De esta forma a finales 2011 la arquitectura de la IDE quedo configurada como se refleja en la figura 3.



IDEAndalucía – Arquitectura diciembre 2011

Figura 3 Arquitectura de la IDE Andalucía en diciembre de 2011

1.4. EVOLUCIÓN DE LA DEMANDA.

Desde su entrada en servicio en Junio de 2006, la demanda de servicio de la IDEA ha tenido un crecimiento sostenido. Hemos analizado el registro de visitantes totales, contabilizando cualquier sesión que inicia un usuario y que culmina cuando lo abandona, o está inactivo más de 30 minutos. Como se aprecia en la figura el crecimiento ha sido exponencial en estos nueve años multiplicándose

por 14 los usuarios de 2006 pasando de los 41.485 visitantes de entonces, hasta los 596.254 de 2014.

Durante los tres primeros años de existencia de la infraestructura, el crecimiento fue moderado, en coherencia con la discreta oferta de servicios que se prestaban. Pero a partir de 2009, coincidiendo con el despliegue de la Infraestructura de la IDE Andalucía V2, la demanda sufrió un proceso de aceleración significativo, particularmente intenso en el periodo 2009-2010, coincidiendo con el despliegue del grueso de la oferta de servicios de la IDE Andalucía.

Así, si a principios de 2008 se podía acceder a 7 servicios de visualización de mapas -4 mapas y 3 ortofotos- en octubre de 2009 son ya 103 los servicios disponibles, agrupados en las siguientes

Categorías:

- 9 mapas topográficos
- 15 ortofotos y ortoimágenes de cobertura regional
- 27 mapas ambientales

Desde 2010 hasta la fecha, los incrementos anuales de la demanda han sido también muy notables, aunque su justificación, más que con algún hito concreto vinculado a la publicación de nuevos servicios, está relacionada con la progresiva incorporación de usuarios, que se incorporan al consumo de esta tecnología.

Sin embargo los nuevos retos que afronta el Instituto de Estadística y Cartografía de Andalucía hace más que previsible, nuevos crecimientos significativos de la demanda. De mantenerse este crecimiento y unido a los nuevos retos que punta el Plan Estadístico y Cartográfico es previsible que los próximos 4 años (2018) se alcance la cifra del millón de visitas anuales.

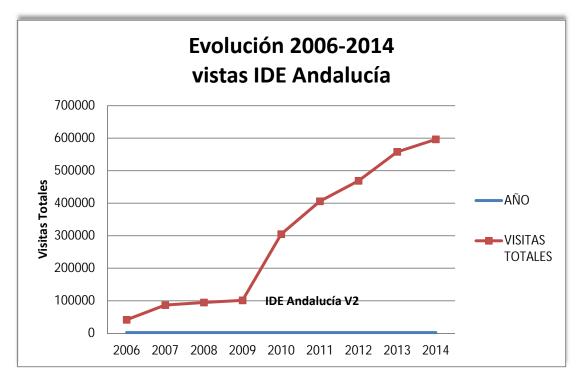


Figura 4 Evolución 2006-2014 visitas totales IDE Andalucía

2. LOS NUEVOS RETOS DE LA INFRAESTRUCTURA DE DATOS ESPACIALES DE ANDALUCÍA

La creación del Instituto como resultado de la fusión del Instituto de Estadística de Andalucía y el Instituto de Cartografía de Andalucía, no es sólo una apuesta por mejorar la eficiencia de la administración pública andaluza, sino que sitúa a nuestra Comunidad Autónoma en una posición avanzada, a nivel europeo y español, en la integración de la información estadística y la información geográfica, lo que conlleva, además de un proceso innovador de confluencia tecnológica, una mejora de los servicios de información que se prestan a la sociedad andaluza.

En ese contexto se ha formulado la ley 3/2013, de 24 de julio, por la que se aprueba el Plan estadístico y Cartográfico de Andalucía 2013-2017 y en el que el aprovechamiento de la Infraestructuras de Información se convierte en una estrategia para el desarrollo del Plan. En base a ésta, el Plan establece como ámbitos da actuación prioritaria:

- a) Respecto del territorio, la Red Andaluza de Posicionamiento, el Callejero Digital de Andalucía Unificado y el Sistema de Gestión de Entidades Territoriales de Andalucía para direcciones postales y los límites político-administrativos.
- b) Respecto de la población, la Base de Datos Longitudinal de Población de Andalucía.
- c) Respecto de La actividad económica, el **Directorio de Empresas y Establecimientos con Actividad Económica en Andalucía**, que integrará las fuentes sectoriales, catastrales, fiscales y de la Seguridad Social.
- d) Respecto a los servicios públicos, el **inventario de los equipamientos y servicios** públicos tales como los de carácter sanitario, educativo, cultural, deportivo, de empleo, social, medioambiental o local.

Su consideración por el Plan como infraestructuras básicas del Sistema Estadístico y Cartográfico, plantean nuevos retos en la prestación de servicios a ofrecer por la IDE Andalucía, y que justifican sobradamente, la planificación, diseño y despliegue de una nueva arquitectura que garantice la eficiencia, escalabilidad y disponibilidad a largo plazo.

Hay que tener también presente que en los ocho años trascurridos desde el despliegue de la primera versión del infraestructuras, todo el equipo técnico vinculado a las tareas de gestión y administración de la IDE, han interiorizado el hnow how de una tecnología entonces desconocida. Junto a ello la gestión de la infraestructura en el nuevo Instituto de Estadística y Cartografía, que tiene la consideración de agencia administrativa con presupuesto propio y capacidad de gestión autónoma, facilita el control de una actividad tan especializada como esta.

La nueva Arquitectura deberá estar dimensionada para soportar:

A nivel de servicios:

- cluster de mapserver 65 servicios (WMS)
- cluster de geoserver 24 servicios (WMS y WFS)
- Dos catálogos de metadatos uno de datos y otro de servicios (CSW)

A nivel de datos

- 13 bases de datos espaciales con un volumen total de 80 GB de información para dar soporte a los servicios vectoriales
- Repositorio de ficheros raster de 630 GB

El resumen de Los servicios a prestar es

- 11 mapas topográficos
- 15 ortofotos y ortoimágenes de cobertura regional

- 12 series de cartografía derivada (Espacios Naturales, Carreteras, fisiográficos...)
- 3 Modelos Digitales del Terreno.
- 4 series de Cartografía Urbana.
- Calleiero Digital de Andalucía.
- 3 series de cartografía temática
- Nomenclátor Geográfico de Andalucía

3. LA NUEVA ARQUITECTURA DE LA IDE

La nueva arquitectura garantiza el cumplimento de lo establecido en el REGLAMENTO (CE) Nº 976/2009 DE LA COMISIÓN de 19 de octubre de 2009 por el que se ejecuta la Directiva 2007/2/CE del Parlamento Europeo y del Consejo en lo que se refiere a los servicios de red, así como al Rendimiento, Capacidad y disponibilidad de los servicios de red.

3.1 CARACTERISTICAS Y BENEFICIOS.

La definición de la nueva arquitectura ha tenido como objetivo la optimización de los recursos, basándose en los criterios de simplicidad, eficiencia, escalabilidad y disponibilidad. Con la nueva arquitectura se han organizado mejor los servicios que suministra IDE Andalucía, permitiendo asignar más recursos a los servicios más demandados por la ciudadanía, reduciéndolos para los menos demandados, lo que repercute en la eficiencia del sistema.

Las características y beneficios más notables de la presente arquitectura frente a las anteriormente utilizadas son:

3.1.1. VIRTUALIZACIÓN SOBRE EL HARDWARE

Aunque la virtualización está ya sobradamente extendida, no podemos omitir las ventajas que se han obtenido utilizando esta técnica frente a la empleada anteriormente.

Son de destacar una más fácil gestión de la operación, mantenimiento y despliegue de nuevas versiones, así como un mejor uso de los recursos (memoria, procesador).

En efecto, la nueva arquitectura permite su ampliación o escalabilidad de una forma mucho más eficaz, al tratarse de maquinas virtuales, en momentos de necesidad pueden desplegar las máquinas necesarias de forma rápida, o incluso asignar más recursos a las existentes. Asimismo aporta escalabilidad el hecho de que los sistemas se encuentren todos en cluster: para aumentar el rendimiento bastaría con desplegar un nodo adicional y añadirlo al cluster, ya que los repositorios de datos son comunes.

La migración de una arquitectura física a una arquitectura virtualizada permite la optimización de recursos, desde el punto de vista económico se ahorra espacio en los CPD y se reducen los costes eléctricos y de refrigeración. A su vez, su flexibilidad permite redimensionar los recursos asignados según la demanda, de esta manera se elimina la infrautilización de recursos que en algunos casos se producía en las máquinas físicas.

Desde el punto de vista de la gestión o administración, ésta es más fácil, ya que se puede redefinir según necesidades la disposición de los servicios sin tener que adquirir nuevo hardware. Los técnicos pueden controlar en todo momento a través de los hipervisores el estado de las máquinas para asignar nuevos recursos o incluso desplegar de forma rápida clones de máquinas existentes para escalar el sistema según demanda.

3.1.2. ALTA DISPONIBILIDAD.

Se ha obtenido una mejora importante en la duplicidad de los servicios.

Se ha sustituido, con importante ahorro de costes, el balanceador hardware por balanceo software.

Además se ha completado la alta disponibilidad activo/activo en todos los niveles (la capa de presentación visores web, capa de servicios de mapas y capa de datos).

Esto hace que ante fallos en alguna parte del sistema solo se pueda producir cierta degradación temporalmente, pero nunca una caída total.

3.1.3. SISTEMA DE FICHEROS EN BLOQUES GFS2.

El empleo de un sistema de fichero en bloques frente al tradicional sistema de ficheros en red (NFS) ha mejorado sensiblemente el rendimiento en escritura.

GFS2 tiene como ventaja frente a otras soluciones un mejor rendimiento al gestionar los bloqueos en escritura en sistema de ficheros (algo fundamental por ejemplo en el cluster de servicios como Geowebcache). Se partía originalmente de una arquitectura en la que se usaba NFS para compartir ficheros entre servidores, lo que era ineficiente en los accesos de lectura/escritura.

3.1.4. DESPLIEGUE DE ÚLTIMAS VERSIONES DEL SOFTWARE.

Esta operación ha permitido la consiguiente mejora de la compatibilidad y funcionalidades, permitiendo hacer uso de herramientas o software de gestión o mantenimiento que antes no era posible debido a la antigüedad de las versiones existentes en la antigua infraestructura.

3.1.5. UBICACIÓN EN EL CPD DEL INSTITUTO DE ESTADÍSITICA Y CARTOGRAFÍA.

El despliegue de a nueva arquitectura ha coincido con la mudanza al CPD del IECA, abandonando su ubicación en el CPD de SANDETEL, que hasta la fecha tenía encomendada la gestión y administración de la IDE. Este hecho, unido a las mejoras en la gestión introducidas por la virtualización, permite una administración más ágil y tiempos de despliegue más reducidos.

Los sistemas de comunicaciones, almacenamiento y backup se comparten con otros sistemas del IECA, lo que, sin disminuir en absoluto su disponibilidad, constituye un ahorro importante.

Por otro lado, la administración se efectúa por el propio personal que presta sus servicios en el IECA, con un notable ahorro de costes.

Antes de la puesta en producción de la nueva arquitectura se diseñó una prueba que validase el correcto funcionamiento de los servicios para situaciones de carga. Para su caracterización se realizaron peticiones de teselas de 256x256 distribuidas aleatoriamente sobre la superficie de Andalucía a escala 1:10000 para diversos servicios en formato PNG de 24 bits para una selección de los servicios raster y vectoriales más representativos.

- Ortofoto 2009
- Ortofoto 2010-2011
- MTA 400r 2008
- MTA 10r 2001
- MDT 2009
- MTA 10v 2007
- MTA 10v 2007n
- DEA 100
- MTA 100v 2005

Aunque la norma habla de imágenes de 800x600 de unos 470 KB, indica que se deben realizar las peticiones en una *situación normal*, es decir, donde no se producen picos de carga. En el caso que nos ocupa las pruebas han estado caracterizadas por los siguientes parámetros (hasta una concurrencia de 50).

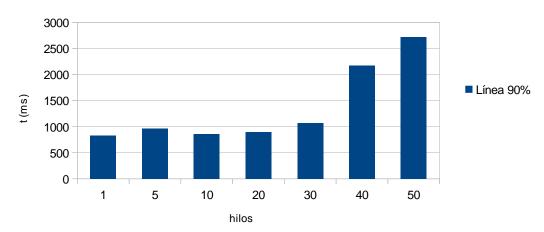
• 1, 5, 10, 20, 30, 40 y 50 hilos

- 15 minutos por grupo de hilo
- 500 ms de media entre peticiones
- 250 ms de desviación estándar entre peticiones

Los resultados fueron los siguientes:

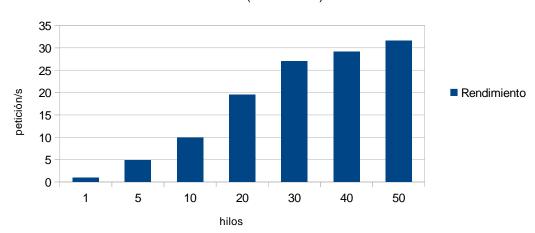


Global (tiempo de espera del 90%)



IDEAndalucía

Global (rendimiento)



En cuanto al rendimiento para una concurrencia de 50 hilos se atendían 30 peticiones por segundo. No se continuó con más hilos porque para 30 ya se había superado ampliamente el rendimiento proporcionado por la anterior arquitectura.

3.2. DESCRIPCION DE LA ARQUITECTURA

La arquitectura está compuesta por los siguientes grupos de servidores:

3.2.1. BALANCEADORES 2 SEVIDORES DE BALANCEO

Se utiliza un balanceo software proporcionado por HAproxy, tecnología que es utilizada en grandes sitios de internet en todo el mundo (Instagram, Twitter, GitHub, etc.) Con esta operación, se ha conseguido sustituir un balanceo hardware (F5), que es un fabricante con un licenciamiento de coste elevado, sin pérdida de funcionalidad.

3.2.2. SERVIDORES DE MAPAS: MAPSERVER

Se ha implantado una de las versiones más recientes de MapServer, correspondiente a la rama 6.x.

Se ha compilado el software para poder dar soporte a formatos propietarios como ECW, MrSID o JPEG2000; asimismo, se ha incluido soporte para OpenJPEG, lo que obliga a usar una versión específica de GDAL.

Junto con el servidor web se utiliza tecnología multihilo frente a multiproceso, lo que permite atender más peticiones simultáneas con un menor consumo de memoria.

Además, se hace uso de FastCGI frente a CGI. Este modo de funcionamiento hace que los procesos mapserver que sirven el contenido no tengan que iniciarse cada vez que llega una petición, si no que los procesos se encuentran ejecutándose a la espera de tramitar peticiones. Esto disminuye notablemente la latencia.

Al mismo tiempo se ha dado soporte a PostGIS 2.1.

Se ha optado por desplegar un mayor de nodos en lugar aumentar recursos indiscriminadamente (RAM, CPU....), y en caso de un incremento de la demanda permite aumentar el número de nodos

3.2.3. SERVIDORES DE MAPAS: GEOSERVER/GEOWEBCACHÉ.

Para algunos servicios como la descarga WFS se utiliza Geoserver en modo activo/activo. En el caso de los servicios cacheados se ha optado por una instalación standalone de Geowebcache, que es más flexible que la integrada con Geoserver. Estos servicios se prestan en modo activo/activo gracias al uso de de GFS2.

En el caso de Geoserver basta con compartir el directorio de datos entre los nodos y desactivar la funcionalidad de cache, de forma que la configuración se realiza en un solo nodo y se replica en el otro recargando el catálogo de objetos.

Para Geowebcache hay que compartir el directorio de almacenamiento de teselas entre los nodos y utilizar un backend compartido. En el caso que nos ocupa se ha optado por PostgreSQL como repositorio para la base de datos de cuota de disco, que al predeterminado Berkely DB, que no se puede usar en cluster activo/activo al no permitir accesos concurrentes.

3.2.4. BASES DE DATOS: SERVIDORES POSTGRESOL/POSTGIS.

La base de datos se ha desplegado en alta disponibilidad activo/activo.

Se ha realizado una migración de todos los datos, desde PostGIS 1.3 a PostGIS 2.1 Aunque lo más notable de la nueva arquitectura de base de datos es el uso de la función de replicación en línea (streaming replication) disponible a partir de PosgreSQL versión 9. Esto, combinado con una capa de pooling y balanceo, permite tener un cluster activo/activo en lectura y escritura para las bases de datos, aumentando considerablemente el rendimiento de los servicios vectoriales.

Cada una de las instancias del cluster de PostgreSQL es independiente, en el sentido de que no se comparte el almacenamiento que alberga las bases de datos. En la instancia que juega el papel de V Jornadas Ibéricas de Infra-estruturas de Dados Espaciais

maestro hay que activar los registros WAL (Write Ahead Log). De esta forma, se almacenan en un directorio configurable las transacciones realizadas sobre las bases de datos de la instancia.

El nodo esclavo se configura de forma que se conecta a la instancia maestra para la replicación en línea. También se le da acceso al directorio en el que se encuentran los registros WAL, de forma que se garantiza la integridad de los datos entre instancias aunque se pierda la conexión de replicación (por ejemplo, en un reinicio del servicio). Finalmente, para que el nodo esclavo acepte sentencias de tipo SELECT se configura en modo hot standby.

Para el balanceo de las peticiones se utiliza pgpool-II. Este software balanceo ofrece multitud de funcionalidades, pero en el caso que nos ocupa solo se utilizan las de balanceo y pooling. Con una configuración relativamente sencilla este software está repartiendo cargas de hasta 10000 sentencias SQL por segundo.

3.2.5. SERVIDORES PARA APLICACIONES, CLIENTES Y VISUALIZADORES VARIOS

En este grupo se encuentran las aplicaciones que consumen los distintos servicios proporcionados por los sistemas anteriormente mencionados: visualizadores, cliente de descarga, clientes del servicio de catálogo, etc.

3.2.6. SERVIDORES PARA EL NOMENCLÁTOR GEOGRÁFICO DE ANDALUCÍA.

En lo que respecta al Nomenclátor Geográfico no se ha hecho ninguna modificación respecto a la arquitectura anterior.

3.2.7. SERVIDORES DE DESCARGA.

Dos servidores (1+1) activo/pasivo, para el portal de descargas Line@ y los catálogos.

Debido a limitaciones técnicas de las versiones del software de catálogo Geonetwork empleadas no es posible configurar en modo activo/activo el servicio de catálogo, ya que se producen bloqueos en la escritura concurrente de los índices de Lucene utilizados para las búsquedas.

3.2.8 COMUNICACIONES.

Las comunicaciones se han mantenido tal como estaban, manteniendo un alto nivel de prestaciones, favorecida por la economía de escala de la red corporativa que permite absorber los picos demanda, permitiendo razones de transferencia al exterior de unos 80 Mbps (máxima registrada por ahora).

3.2.9. ALMACENAMIENTO

Se ha mejorado notablemente el sistema de almacenamiento, tanto en dimensión como en velocidad, al usar GFS en lugar de NFS para el acceso a los ficheros de datos compartidos entre servidores.

Frontend Frontend Service of Se

IDEAndalucía - Arquitectura septiembre 2014

Figura 4. Arquitectura de la IDE Andalucía en septiembre de 2014

4. CONCLUSIONES

La nueva arquitectura desplegada para soportar los servicios de la IDE Andalucía, va a garantizar la suficiente capacidad para los retos que el Plan Estadístico y Cartográfico de Andalucía ha señalado para los próximos años. En un horizonte de largo plazo, en el que manteniéndose la evolución de la demanda analizada unido a los nuevos retos del Plan, es más que previsible que se alcance el millón de vistas/año.

Está garantía está basada en las siguientes características:

Simplicidad: En la Nueva arquitectura es más fácil la gestión de la operación, mantenimiento y despliegue de nuevas versiones, siendo también más óptimo uso de los recursos (memoria, procesador).

Eficiencia: La migración de a una arquitectura completamente virtualizada ha permitido la optimización de recursos, desde el punto de vista económico se ahorra espacio en los CPD y se reducen los costes eléctricos y de refrigeración. Su flexibilidad permite redimensionar los recursos asignados según la demanda, de esta manera se elimina la infrautilización de recursos que en algunos casos se producía en las máquinas físicas

Escalabilidad: La arquitectura está preparad para hacerse más grande sin perder calidad en los servicios ofrecidos.

Disponibilidad: Se ha sustituido, con importante ahorro de costes, el balanceador hardware por balanceo software. Se ha completado la alta disponibilidad activo/activo en todos los niveles (la capa de presentación visores web, capa de servicios de mapas y capa de datos). Esto hace que ante fallos en alguna parte del sistema solo se pueda producir cierta degradación temporalmente, pero nunca una caída total.

REFERENCIAS

- [1] Sánchez F., Pardo, E., Redondo, M., Zabala, A.: Renovación de arquitectura, aplicaciones y datos en IDE Andalucía VI Jornadas Técnicas de la IDE de España JIDEE2008. Tenerife, 5 de noviembre de 2008.
- [2] Pardo, E., Redondo, M.: IDEAndalucía: más servicios y aplicaciones para nuevas políticas. VI Jornadas Técnicas de la IDE de España JIDEE2009. Murcia, 4 de junio de 2009.
- [3] Junta de Andalucía. Ley 3/2013, de 24 de julio, por la que se aprueba el Plan estadístico y Cartográfico de Andalucía 2013-2017. BOJA número 154, de 7 de agosto de 2013.
- [4] Geoportal Infraestructuras de Datos Espaciales de Andalucía. www.ideandalucia.es

AUTORES

GUTIERREZ José Angel

email: <u>joseangel.gutierrez@juntadeandalucia.es</u> Organización: Instituto de Estadística y

Cartografía de Andalucía

Departamento: Servicio de Informática

MOLINA, Antonio

email:antonio.molina@juntadeandalucia.es

Organización: Instituto de Estadística

Cartografía de Andalucía

Departamento: Servicio de Informática

VILLAR, Agustín

email:agustint.villar@juntadeandalucia.es

Organización: Instituto de Estadística

Cartografía de Andalucía

Departamento: Servicio de Infraestructuras

Geográficas.

MARTIN, Abel

email:<u>abel.martin.ruiz.ext@juntadeandalucia.es</u>

Organización: Instituto de Estadística

Cartografía de Andalucía

Departamento: Servicio de Informática

PARDO, Emilio

email: emilioj.pardo@juntadeandalucia.es

Organización: Instituto de Estadística y

Cartografía de Andalucía

Departamento: Servicio de Infraestructuras

Geográficas.

У