



Sócrates: Servicios Web 2.0 y almacenamiento personales

Socrates: 2.0 web services and personal network disk

◆ Ática, Universidad de Murcia

Resumen

Sócrates es un conjunto de actuaciones de la Universidad de Murcia para fomentar la implantación y uso de software libre en la misma. Una de sus acciones es la de ofrecer una plataforma extensible para ir dotando de servicios **Web 2.0**, de forma cómoda y sencilla, a la comunidad universitaria y ofreciendo además un espacio de almacenamiento personal en red. Este artículo describe la solución desarrollada al respecto del almacenamiento y los servicios 2.0.

Palabras clave: servicios web 2.0, unidad de red, wiki, alta disponibilidad.

Summary

Sócrates is the name of a group of actions in order to promote and provide open-source services to our community. One of the major goals of the project is to offer a network unit and Web 2.0 services like wikis, to teachers and students.

Keywords: web services 2.0, network unit, wiki, high availability.

◆
El proyecto Sócrates se inició en el curso 2004/2005 en la Universidad de Murcia

1. Introducción

En el curso 2004/2005 se inició en la Universidad de Murcia (en adelante UM) el proyecto Sócrates, para ir creando aulas de acceso basadas totalmente en software libre. Este proyecto además dotaba a los profesores de una unidad remota en red. Desde entonces ha ido creciendo en varios aspectos, tanto en número de usuarios como en servicios. A continuación exponemos los principales retos y las soluciones que hemos ido aportando.

2. Objetivos

Entre los principales objetivos que teníamos cabría destacar por un lado los de **servicio** en sí:

Dotar a toda la comunidad universitaria (orden de decenas de miles de usuarios) de espacio de almacenamiento personal.

Flexibilidad para la autorización y personalización basada fundamentalmente en perfiles de usuario: alumnos, pdi, estudiantes de cierta carrera, centro, facultad, etc.

Inclusión de servicios avanzados: web, wikis, blogs...

Login único para todo el sistema.

Instalación y desinstalación de servicios por parte del usuario.

Del lado de **sistemas** resaltaríamos: sistema flexible para la gestión de espacio, alta disponibilidad de servicios (almacenamiento, samba, web,...) y un bajo impacto en tareas de administración como actualizaciones desatendidas.

3. Requisitos

Entre los requisitos con los que partíamos cabría destacar los siguientes:

De almacenamiento: Servidor de ficheros de red (**NAS**) de EMC tipo Celerra, conectado a una **SAN** también de EMC (Clarion). Estos sistemas son exportados bajo **NFS**.

El tamaño máximo por sistema de ficheros (en adelante FS) recomendado por EMC en el Celerra es de unos 500 GB, aunque puede llegar a 2 TB teóricos.

El proyecto inicial de unidad de red que ya llevaba tiempo en uso y estaba destinado sólo a profesores (PDI), ofrecía sólo **SAMBA/CIFS** como mecanismo de acceso, cosa que nos obligaba a mantener esta interfaz.

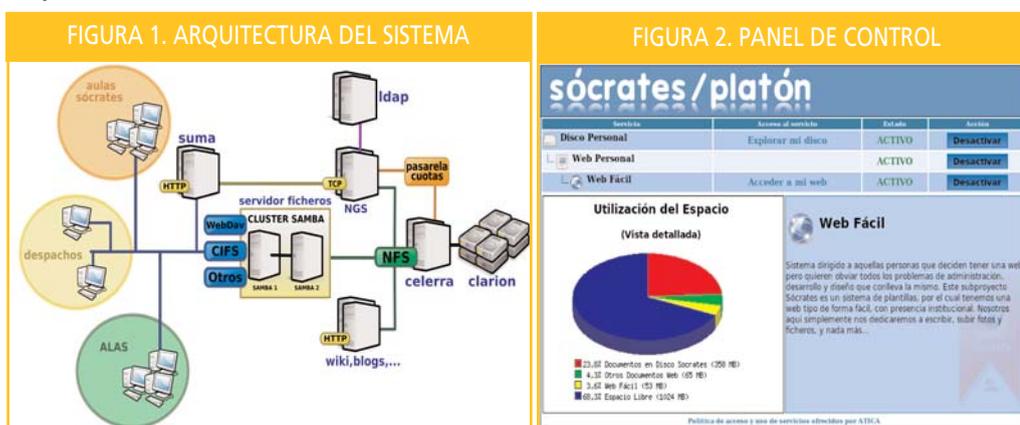
Los usuarios están en el **LDAP** (OpenLdap)[1] de la universidad, donde la información de cuenta fue heredada de la configuración del sistema de correo electrónico.

◆
Uno de los objetivos del proyecto es dotar a toda la comunidad universitaria de espacio de almacenamiento personal

4. Descripción de la solución

A continuación describimos los elementos más importantes de la solución.

Arquitectura del sistema



Espacio de almacenamiento

Uno de los primeros y mayores problemas fue la gran cantidad de espacio a direccionar, pues según la **cuota** inicial prevista, de 1GB para PDI y 512 MB para alumnos, había que poder almacenar más de 16 TB si toda la población lo usara al 100% y la población no creciera. EMC recomienda no utilizar FS de un tamaño superior a 500 GB y diseñamos todo sistema para poder trabajar con múltiples FS y no tener de entrada ninguna limitación teórica de espacio. Decidimos partir inicialmente con 26 FS, ampliables en caliente y un tamaño pequeño, que nos daban de entrada un direccionamiento máximo recomendable de hasta 12,7 TB, aunque basta con añadir más FS para seguir ampliando el espacio total.

Los estudios que hemos hecho posteriormente con datos de uso, muestran que la distribución de ocupación por usuario sigue una distribución que se asemeja a una de tipo Pareto (Power Law), y que disminuyen enormemente la cantidad práctica a direccionar.

Servicio de unidad de red mediante SAMBA/CIFS

SAMBA (CIFS) no es un protocolo que permita fácilmente el escalado ni el balanceo frente a otras alternativas como WebDav. Al tener que mantener el servicio, optamos por un clúster en modo activo/pasivo (A/P en adelante) y ofreciendo nombres de servicio distintos para los colectivos internos como las aulas de ordenadores. En caso de necesitar ampliar el número de servidores SAMBA, bastaría redirigir los nombres de servicio necesarios al nuevo clúster.

Compatibilidad con el Idap

La información de las cuentas de usuario están en el LDAP. Para no forzar a un mismo esquema de ficheros y no duplicar ramas de información en el árbol, decidimos crear un sistema índices, que tradujera el **homeDirectory** del usuario en el LDAP de las máquinas de correo, a su ubicación real en las máquinas SAMBA.

Gestión de usuarios y cuotas: nodo gestor Sócrates

El tratar con varios sistemas de ficheros trae consigo un problema de equilibrado de la ocupación y gestión interna de usuarios entre los FS. Los algoritmos de inserción y movimiento de los mismos que hemos utilizado, están basados en el algoritmo de Peterson sobre tablas dispersas abiertas, y han demostrado un comportamiento realmente bueno. Los sistemas de ficheros tiene una distribución de la ocupación según la normal estadística y con valores de ocupación cercanos. Además se han definido 2 niveles de uso por FS: uno que marca el porcentaje a partir del cual no pueden insertarse nuevos usuarios en él (sólo permite que crezca la ocupación de lo que ya están), y otro porcentaje superior al anterior que iniciaría los procedimientos bien de ampliación atendida o bien de reequilibrado de los FS. El **nodo Gestor Sócrates** (en adelante NGS), es el software encargado de

Uno de los principales problemas encontrados fue la gran cantidad de espacio a direccionar

La información de las cuentas de usuario están en el LDAP



El panel de control se integra como un módulo más de la plataforma de tele-enseñanza de la universidad (SUMA)

Con este servicio se consigue satisfacer la demanda por parte de nuestros usuarios

realizar en última instancia todas las tareas de alta, movimiento, activación y desactivación de servicios, así como enlazar con el LDAP y las cuotas del Celerra.

El estado de todos estos servicios está almacenado en el LDAP de la UM mediante los atributos definidos en el esquema de RedIris (**iris ldap schema**) [2] con lo que nos permite de forma sencilla la consulta, activación o desactivación con sólo modificar los valores allí almacenados.

Las cuotas las maneja el propio Celerra, y nos hemos limitado a crear un interfaz con el mismo para el establecimiento y gestión de las cuotas desde el NGS.

Alta disponibilidad de los servicios

Servidores de ficheros NFS: el NAS (Celerra) está constituido por dos nodos en modo A/P.

Servidores de ficheros SAMBA: hemos montado un clúster Debian en modo A/P basado en Heartbeat[3] conectado al LDAP.

Servidores web: de momento utilizamos el nodo que está en modo pasivo de los servidores de ficheros.

NGS: configuración A/P instalado en uno de los nodos del servidor de ficheros SAMBA.

Actualmente estamos en un proceso de migración a máquinas virtualizadas (VM), así que en futuras actualizaciones todas estas máquinas físicas se montarán sobre máquinas virtuales (VM) separando lo más posible las funciones.

5. Servicios ofrecidos

Además de la unidad de red, la arquitectura Sócrates permite la inclusión y parametrización de nuevos servicios que los usuarios que tengan permiso para ello pueden activar y desactivar de forma sencilla. Inicialmente se incluyen los siguientes:

- Web fácil: servicio de wiki basado en dokuwiki[4] con plantillas personalizadas.
- Web personal: servicio de hospedaje web tradicional con php.

6. Panel de control de usuario

Se integra como un módulo más de la plataforma de tele-enseñanza de la propia universidad (SUMA), y ofrece un panel de control sencillo a los usuarios donde activar o desactivar los servicios que tenga disponibles en función de su perfil con sólo pulsar un enlace, así como explorar los ficheros que tiene en su unidad remota. Ver figura 2.

7. Conclusión y trabajos futuros

Con este servicio se consigue satisfacer la demanda por parte de nuestros usuarios: unidad remota de almacenamiento, web tradicional o web fácil (wikis), permitiéndonos ampliar la oferta en un futuro (blogs, indexación...) de forma sencilla y cómoda para el usuario.

Referencias

[1] *OpenLdap*. Consultado en (<http://www.openldap.org>) 31-10-2007

[2] *Esquemas Ldap de Rediris*. Consultado en (<http://www.rediris.es/ldap/esquemas/>) 31-10-2007

[3] *HearBeat*. Consultado en (<http://www.linux-ha.org/>) 31-10-2007

[4] *Dokuwiki*. Consultado en (<http://www.splitbrain.org/projects/dokuwiki>) 31-10-2007

Miguel A. García Lax (glax-AT-um.es)
Fco. Javier García Ros (jgarcia-AT-um.es)
Jose Fco. Hidalgo Céspedes (jhidalgo-AT-um.es)
Alfonso Marín Marín (almarin-AT-um.es)
Antonio D. Merino Jiménez (amerino-AT-um.es)
Juan José Vidal Agustín (juanjova-AT-um.es)

Ática, Universidad de Murcia