



SERIE: SISTEMAS DE ALMACENAMIENTO DISTRIBUIDO (V)

Propuesta de Sistemas de Almacenamiento Distribuido (SAD) en celdas de almacenamiento¹

Ricardo Marcelín Jiménez

Junio de 2012

Dado que las necesidades en el manejo de información de las organizaciones actuales son cada vez mayores, debemos estar siempre un paso adelante y pensar en soluciones creativas capaces de sostener este crecimiento. El almacenamiento es uno de los aspectos, tal vez el más importante, que debe ser atendidos con cuidado. El problema es que los sistemas de almacenamiento “en casa” resultan sumamente costosos, y no todas las organizaciones pueden pagarlos.

Sería un error pensar que únicamente las grandes empresas tienen altas necesidades de almacenamiento; imaginemos, por ejemplo, una pequeña clínica familiar que necesita almacenar los expedientes de sus pacientes de manera que puedan ser recuperados con eficiencia. En tal caso, se trataría de información sensible que debe ser tratada con cuidado, además de que el volumen de almacenamiento aumentará notablemente conforme pase el tiempo. ¿Se puede pensar en un sistema de almacenamiento de bajo costo que tenga altos

¹Este artículo fue redactado por Fernando Barajas con base en la investigación *Hacia los sistemas de almacenamiento distribuido de información, orientados por la naturaleza de los contenidos*, cuyo responsable es el Dr. Ricardo Marcelín Jiménez, quien colabora en proyectos de investigación aplicada del Fondo de Información y Documentación para la Industria INFOTEC.



estándares de calidad y sea lo suficientemente flexible como para considerar situaciones distintas? A lo largo de este artículo trataremos de explicar una propuesta que pueda solucionar esos cuestionamientos.

Gran parte de las organizaciones que necesitan almacenar información cuentan con una infraestructura moderna que se encuentra, en general, subutilizada. El apropiado manejo de estos equipos posibilitaría el hecho de disponer de capacidades de almacenamiento en el orden de los terabytes, e incluso de los petabytes. Sin embargo, para lograr esta tarea, se debe lidiar con una serie de dispositivos de distintos modelos, de diferente marca, con sus propios sistemas operativos y capacidades particulares. En otras palabras, para optimizar los sistemas que una organización ha adquirido, es necesario resolver su heterogeneidad sin pretender eliminarla, de tal forma que resulte un sistema distribuido, auto-organizado y carente de un dispositivo central. Esto debe ser así porque las capacidades de un dispositivo de esta naturaleza pronto serían insuficientes.

Toda propuesta de SAD debe tomar decisiones al menos en tres aspectos imprescindibles. El primero de éstos se refiere a las **necesidades de los usuarios** (requerimientos funcionales) del tipo: monitoreo de operaciones, gestión de metadatos, confidencialidad, etc.; y a los principios de diseño (requerimientos no funcionales), como la flexibilidad, la interoperabilidad y la confiabilidad. Por su parte, la **cantidad de información redundante**, el **balance de carga** en los dispositivos y la **disponibilidad** son dimensiones del segundo aspecto, se trata de parámetros de desempeños que garantizan la **calidad**. En el último aspecto, la



entidad básica de almacenamiento, es importante insistir en que no parece conveniente recargarse demasiado en dispositivos individuales, pues ellos se encuentran proclives a fallas; es mejor generar una memoria virtual que tenga un mayor tiempo de vida útil.

Bajo el entendido de que todos estos detalles deben ser tomados en cuenta, se propone el diseño y la construcción de "celdas de almacenamiento". La articulación de celdas de este tipo permitiría la implementación de un sistema de almacenamiento distribuido de bajo costo y alto desempeño.

Esta propuesta supone una organización por "celdas" o grupos de dispositivos de almacenamiento coordinados por un representante. Para entender esto, imaginemos el funcionamiento de un cuerpo humano, en donde cada dispositivo (como un servidor o una computadora) es una célula; un conjunto de estas células, será, bajo este concepto, un órgano. Dentro de cada órgano existirá una célula que coordine a las demás en un sistema P2P (peer to peer o entre pares). El conjunto de órganos será la estructura del sistema de almacenamiento. Se pueden agregar tantos órganos como se quiera, pues su relativa independencia los hace tolerantes a fallas aunque el sistema crezca; además de que la coordinación de sus tareas les permite afrontar grandes volúmenes de información. Las celdas de almacenamiento funcionarían como estos órganos.

En suma, se trata de un sistema cuyo centro está en todas partes. Cada celda es un SAD por sí misma, de ahí que pueda administrar con mayor facilidad la información y cumplir con los requerimientos sin aumentar demasiado los



costos. Los dispositivos de coordinación no pueden ser considerados como "centralizadores" del sistema, pues por sí mismos no cumplen con las tareas de administración, sino que simplemente tienen la información que indica cuál dispositivo cumple con determinada tarea.

En suma, la construcción de celdas de almacenamiento nos permite aprovechar los recursos con los que cada organización cuenta, sin la necesidad de descartar inversiones anteriores y sin las exigencias de equipos costosos. Esta organización nos da múltiples ventajas: en lugar de un solo dispositivo que administra una red enorme, tenemos muchas celdas auto-organizadas que, a su vez, se articulan en un gran sistema. Para decirlo en pocas palabras, ésta puede ser una solución interesante a los problemas de almacenamiento lo suficientemente flexible como para abarcar diversas organizaciones sin subestimar sus necesidades particulares.

Si te interesó el artículo, también puedes consultar:

- [Investigación "Hacia los sistemas de almacenamiento distribuido de información, orientados por la naturaleza de los contenidos"](#)
- [Artículos de Divulgación INFOTEC](#)
- [Proyectos de Investigación Aplicada en INFOTEC](#)



Esta obra está sujeta a la licencia **Atributo-No comercial-Sin obras derivadas 2.5 México** de Creative Commons. Puede copiarla, distribuirla y comunicarla públicamente siempre que cite a su redactor, autor y la institución que la publican (INFOTEC), no la utilice para fines comerciales ni haga con ella obras derivadas.

La licencia completa se puede consultar en:
<http://creativecommons.org/licenses/by-nc-nd/2.5/mx/>

**Ricardo Marcelín Jiménez**

r.marcelin.jimenez@gmail.com



Doctor en Ciencias Computacionales por la Universidad Nacional Autónoma de México. Profesor del Área de Redes y Telecomunicaciones, del Departamento de Ingeniería Eléctrica de la Universidad Autónoma Metropolitana, Unidad Iztapalapa. Como investigador, sus intereses son el almacenamiento distribuido, las redes inalámbricas de sensores y la simulación de eventos discretos. Actualmente, entre otras actividades, colabora con INFOTEC dirigiendo los proyectos “*Sistema de Imagenología Médica basada en Almacenamiento Distribuido*” y “*Sistema de Almacenamiento Distribuido basado en el Modelo de Nube*”.

INFOTEC es:

- Investigación - Educación - Soluciones integrales -