Utilización de técnicas de replicación como estrategia de *backup* y *site* de contingencia

Por ***Lic. Julio César Panno***, Compañía General de Combustibles S.A.

La integración de herramientas de virtualización y de protección de la información debe utilizarse para resguardar datos y aplicaciones en un lugar alternativo o de contingencia, con el empleo de procedimientos desatendidos dentro de una ventana de tiempo compatible con las necesidades y los requerimientos del negocio. Además se debe asegurar que los datos estén disponibles con tiempo de recuperación mínimos.

Este trabajo ha sido galardonado con una Mención Especial en el área de Tecnología Informática de las V Jornadas de Geotecnología *realizadas durante el* IX Congreso de Exploración y Desarrollo de Hidrocarburos *del* IAPG.

En la actualidad, se destaca la información como uno de los recursos valiosos del que disponen las compañías para lograr un óptimo funcionamiento. Las operaciones de procesamiento de datos y la incorporación de la tecnología informática facilitan el camino hacia la generación del conocimiento en la búsqueda de soluciones a los problemas que se enfrentan de manera cotidiana.

Dentro de las organizaciones de TI podemos destacar cuatro pilares en el gobierno de la información:

- La organización
- La administración
- La protección
- El acceso

Organización de la información

Este punto comprende cómo se estructura la información, cuáles son los atributos que la describen, así como también establece de qué manera se agrupa y cómo se interrelaciona y se integra con el entorno organizacional.

Administración de la información

La administración determina cuáles son las aplicaciones y los procesos que gestionan la información, quiénes son los usuarios que la colectan, la actualizan y la utilizan. Las aplicaciones que organizan y administran esta información, así como los procesos de negocios en los que se montan las aplicaciones, constituyen el motor de las organizaciones sin las cuales se volverían inoperantes.

Protección

Detectar riesgos, mitigar y minimizar las vulnerabilidades que representa para las compañías la utilización de los sistemas informáticos, conlleva a las organizaciones de TI a estudiar las medidas de adopción más convenientes para salvar una circunstancia de tipo accidental o intencional que impida la continuidad operativa como consecuencia de algún inconveniente de cualquier naturaleza.

Acceso

En el ámbito informático, la gestión del *delivery* se ocupa del diseño, la planificación, la implementación y el mejoramiento de los flujos asociados a la entrega y la disponibilidad de la información y de las aplicaciones, contemplando de manera integral la posibilidad de ocurrencia de alguna contingencia.

Los avances en esta materia han vuelto a las organizaciones más dependientes de la tecnología informática, por eso la continuidad operativa depende exclusivamente de la posibilidad de contar "siempre" con la información y con las aplicaciones a pesar de las contingencias que pudieran ocurrir.

La separación geográfica para la custodia y el resguardo de la información y de las aplicaciones es clave en un esquema de contingencia. Mantener alejado el lugar de resguardo amplía enormemente la protección en caso de contingencias, tales como incendios, inundaciones, catástrofes naturales, explosiones e incidentes de seguridad, entre otras.

Tipologías de redes según su alcance

Redes LAN (*Local Area Network*). El Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) define una red local como “un sistema de comunicaciones de datos que permite a un número de dispositivos independientes comunicarse directamente entre sí, dentro de un tamaño moderado de área geográfica sobre un canal de comunicaciones física de moderada velocidad de datos”.

Los canales de comunicación son privados y generalmente se utilizan para conectar computadoras y otros dispositivos de red dentro de un mismo edificio.

Redes WAN (*Wide Area Network*). Representa la comunicación entre redes de área local a través de distancias geográficas extendidas. Por lo general, los canales de comunicación utilizados para actividad WAN son de propiedad de un tercero; por ejemplo, una compañía telefónica.

LAN *vs* WAN. Las palabras área geográfica limitada se utilizan en la definición de una red LAN para resaltar el hecho de que la L en LAN significa Local. Cuando las computadoras están conectadas a través de la ciudad o ciudades, países o continentes, la L se convierte en una W, indica un Red de área amplia o WAN.

Los canales de comunicación de las redes WAN son de menor velocidad que los utilizados en las redes LAN (Figura 1).

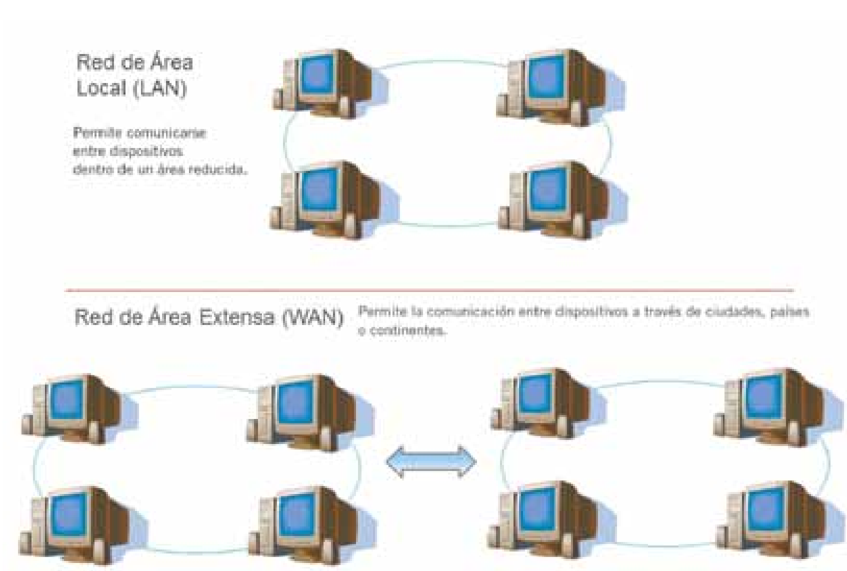

Figura 1. **Alcance de Redes LAN y WAN.**

Figura 2. **La virtualización permite ejecutar en un único servidor físico múltiples sistemas operativos.**

Ancho de banda

Capacidad de transportar datos a través de un enlace de comunicaciones. Esa capacidad se mide en cantidad de datos por unidad de tiempo, generalmente se utiliza bits por segundo. De esta manera podemos denominar a los enlaces de datos según su capacidad de transporte. Para los enlaces de datos tipo LAN podemos referirnos a velocidad, por ejemplo 100 Mbps (Megabits por segundo), 1 Gbps (Gigabit por segundo).

Para los enlaces tipo WAN, generalmente referimos a velocidades menores, como 10 Mbps, 20 Mbps, 30 Mbps.

Descripción de las tecnologías integradas

Para el desarrollo de este proyecto es esencial la adopción de tres tecnologías que, en forma conjunta, posibilitan el desarrollo de esta solución:

- Virtualización de servidores
- Almacenamiento centralizado de disco (*Data Storage*)
- Técnicas de Replicación

Virtualización de servidores. La arquitectura tradicional de servidores permite que se ejecute solo un sistema operativo a la vez. La virtualización de servidores desbloquea esta posibilidad mediante la abstracción del sistema operativo y las aplicaciones del *hardware* físico. Así, el uso de la virtualización de servidores permite que múltiples sistemas operativos puedan ejecutarse en un único servidor físico como “máquinas virtuales”, cada una con acceso a los recursos informáticos del servidor subyacente (Figura 2).

Dicho de otra manera, se refiere a la abstracción de los recursos de una computadora, llamada Hypervisor que crea una capa de abstracción entre el hardware de la máquina física (*Host*) y el sistema operativo de la Máquina Virtual (*Guest*), dividiéndose el recurso en uno o más entornos de ejecución.

Esta capa de *software* maneja, gestiona y arbitra los cuatro recursos principales de una computadora (CPU, Memoria, Dispositivos Periféricos y Conexiones de Red) y así podrá repartir dinámicamente dichos recursos entre todas las máquinas virtuales definidas en el computador central. De esta manera se pueden tener varios servidores virtuales ejecutándose en el mismo servidor físico.

Con la tecnología de virtualización de servidores podemos transformar *hardware* en *software*. De esta forma, podemos ejecutar diferentes sistemas operativos, como Windows, Linux, Solaris, incluso MS-DOS simultáneamente y en el mismo servidor físico.

Una máquina virtual está conformada por un conjunto de ficheros planos y binarios. Se puede definir específicamente como "un duplicado aislado de una máquina real". Por más semejanzas que tenga, las máquinas virtuales no son computadoras ordinarias. Por esto se descubre que el rendimiento de una máquina virtual no es el mismo que el de una computadora común. Este defecto es compensado por la capacidad de las máquinas virtuales de correr sobre cualquier equipo. Esto implica que una máquina virtual se puede apagar, transportar en un dispositivo de almacenamiento portable a otro equipo y la máquina virtual que se abre es exactamente la misma. La abstracción del *hardware* en el que procesa una máquina virtual es completa.

Un hipervisor o monitor de máquina virtual es una plataforma que permite aplicar diversas técnicas de control y de administración de los recursos para utilizar las diferentes máquinas virtuales al mismo tiempo sobre el mismo *hardware*.

Los hipervisores pueden clasificarse en dos tipos:

a) *Bare-Metal* (sobre metal desnudo): el hipervisor está instalado en un servidor físico sin la necesidad de que exista un sistema operativo (Windows o Linux) instalado previamente.
b) *Hosted*: este tipo de hipervisor necesita previamente de un sistema operativo instalado para poder ofrecer la funcionalidad descripta.

La mejora que ofrece esta característica es que se pueden correr procesos, que requieren de SO distintos al mismo tiempo. De esta forma, se pueden aprovechar las cualidades propias de cada sistema operativo sin tener que cambiar de *hardware*.

Utilizar máquinas virtuales ocasiona ahorros en el espacio de memoria física, mejora el aprovechamiento de los recursos disponibles y reduce costos de mantenimiento de equipo. Además, permite mejorar el aprovechamiento del equipo físico al utilizar los recursos que de otra forma estarían

ociosos porque, en general, nunca se llegan a utilizar todos los recursos de un servidor físico al mismo tiempo.

Racionaliza la proliferación de servidores, el espacio físico, el uso de energía y la refrigeración en las salas de cómputos.

Reduce drásticamente el tiempo de aprovisionamiento de nuevos servidores en máquinas virtuales, pasando a minutos en lugar de días o semanas necesarias para aprovisionar un servidor físico.

Asimismo, como la capacidad de procesamiento en los servidores ha aumentado de manera constante en los últimos años, y no cabe duda de que seguirá aumentando, la virtualización ha demostrado ser una tecnología muy potente para simplificar el despliegue de *software* y servicios, al dotar el Centro de Datos de más agilidad y flexibilidad.

Además, podemos aludir a la seguridad de la que nos provee las máquinas virtuales. Cada una es totalmente independiente de la computadora anfitrión, esto significa que varias personas pueden utilizar la misma máquina física, y no corren riesgo de perder o exhibir información confidencial, debido a que cada máquina virtual se encuentra aislada de los demás SO presentes en el equipo.

Almacenamiento centralizado de disco. El almacenamiento centralizado de discos también llamado *Data Storage* es un dispositivo dedicado para conservar los datos digitales y entregarlos cuando son solicitados por alguna aplicación y/o usuario.

Con la implementación de estos dispositivos centrales, todos los servidores se conectan al *Data Storage* para grabar y recuperar datos. Por ese motivo es esencial dotar a este dispositivo de componentes redundantes que garanticen su funcionamiento a pesar de fallas. También es fundamental la optimización del acceso a los datos para procesar múltiples operaciones de lectura y escritura en el dispositivo.

Esta tecnología cuenta con distintos conceptos tecnológicos, entre los principales están los arreglos de disco o RAID (del inglés *Redundant Array of Independent Disks*), traducido como "conjunto redundante de discos independientes", y redes de área de almacenamiento que permiten la adición en caliente de almacenamiento sin ninguna interrupción (tecnología "*Hot Swap*" descripta en la página siguiente).

Los datos son almacenados en unidades de almacenamiento de datos (discos duros o SSD) dispuestos en forma de arreglo configurando una unidad lógica de almacenamiento, así en lugar de ver varios discos duros diferentes, el sistema operativo ve uno solo. Si alguno de los discos del arreglo falla, este tipo de tecnología permite seguir operando sin pérdida de información, ya que la información se encuentra diseminada y/o replicada entre las distintas unidades de almacenamiento.

Esta implementación pueden soportar el uso de uno o más discos de reserva (*hot spare*), que son unidades preinstaladas que pueden usarse inmediatamente (y casi siempre automáticamente) tras el fallo de un disco del RAID. El *software* que administra el

RAID es el que se encarga de detectar el disco dañado, removerlo del arreglo e incorporar el disco de reserva en su lugar y reconstruir inmediatamente la información contenida en el disco dañado a partir de los datos contenidos en el resto del RAID.

A este concepto se suma la tecnología *Hot Swap* que permite instalar y desinstalar discos en caliente, es decir no se necesita apagar el dispositivo para realizar un cambio de discos, ya sea para reemplazar los fallados o bien para incorporar nuevas unidades de almacenamiento.

La unidad lógica configurada a partir del RAID, puede dividirse en volúmenes. Los volúmenes son como particiones de la unidad lógica que permiten ser presentados a diversos sistemas y/o servidores, la misma puede efectuarse en diferentes protocolos de acuerdo con las capacidades nativas de los sistemas operativos. Por ejemplo el protocolo NFS es el utilizado por sistemas Linux, mientras que el CIFS es el que utiliza Windows.

Sin embargo, esta eficiencia de almacenamiento puede venir asociado a una disminución del desempeño. Una gran cantidad de nodos haciendo demandas simultáneamente en el mismo equipo suelen enlentecer la experiencia de cada uno de ellos. Pero, los distintos fabricantes que ofrecen almacenamiento centralizado cuentan también con tecnología inteligente en controladoras, y distribuyen así la carga de trabajo física de los discos que permite el acceso simultáneo y garantizan una cantidad aceptable de entradas y salidas de datos necesaria para cumplir con las demandas de las aplicaciones (Figura 3).

El *Data Storage* se conecta a las redes de comunicación de una compañía y le permite a los servidores físicos o hipervisores acceder a los datos o los sistemas de archivo del *Data Storage* como si fueran locales o propios pero, a su vez, permite compartir estos datos entre varios equipos de la red. Esta característica facilita y posibilita la implementación de soluciones de virtualización de servidores, ya que las máquinas virtuales son almacenadas en el *Data Storage* y pueden ser utilizadas por uno o más hipervisores para su ejecución.

Al poder almacenar una máquina virtual en el *Data Storage* todas las facilidades y funciones del *Storage* pueden ser aplicadas no solo para los volúmenes de datos, sino también para los equipos virtuales.

Otra de las características asociadas a los dispositivos de almacenamiento centralizado es que pueden contar con la posibilidad de realizar *snapshots*. El *snapshot* es una copia de seguridad instantánea de un volumen de datos. Pero, ¿cómo resuelve esto la tecnología? Una de las maneras es cambiando la manera de almacenamiento de los datos en los volúmenes. La utilización de esta técnica incorpora la administración de los bloques de datos a través de tablas de punteros. Los punteros ubican los bloques de datos sobre el disco físico, pero cuando se graba nueva información, el bloque de datos no se actualiza, utiliza un nuevo bloque disponible en el volumen con la información incorporada y actualiza la tabla de punteros. Cuando se toma un *snapshot* no se realiza una copia de datos, sino de la tabla de punteros, esto demanda unos pocos segundos y no genera impacto significativo en el desempeño del *Data Storage*. Cada *snapshot* es una imagen congelada del volumen con acceso de solo lectura que refleja el estado de los bloques de datos al momento de la toma del Snapshot. Como los bloques no se alteran, las copias realizadas son

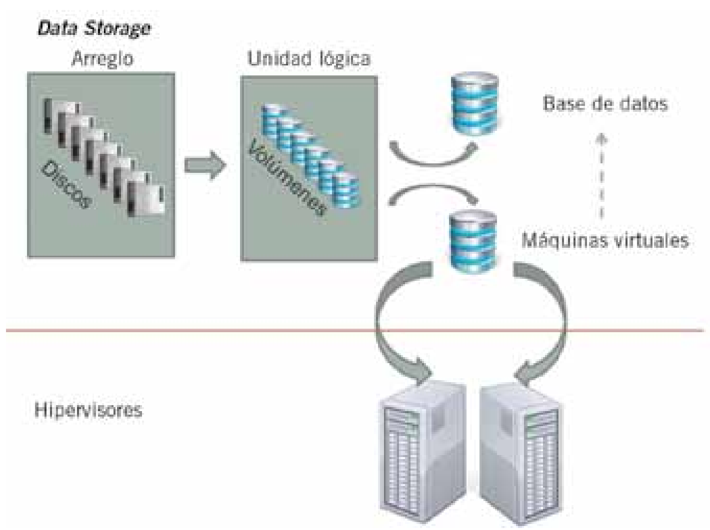

Figura 3. **El *Data Storage* permite el acceso de los servidores físicos o hipervisores de una compañía a los sistemas de archivo allí almacenados.**

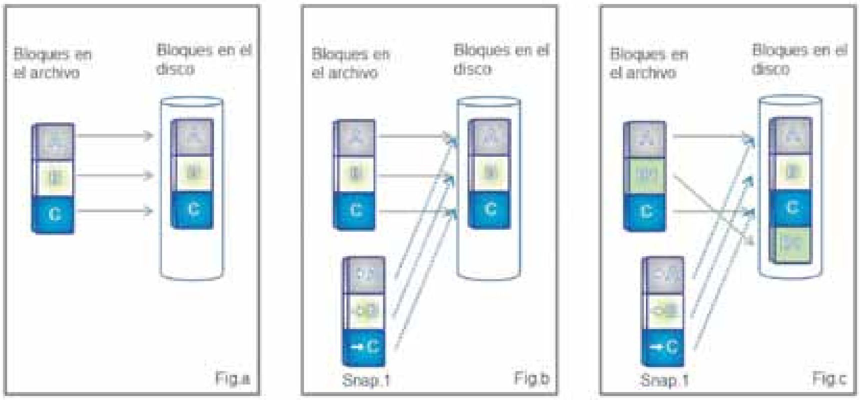

Figura 4. **a) se ejemplifica como se almacenan los bloques de un archivo y la tabla de los punteros. b) se muestra la creación del *Snapshot* a partir de la copia de la tabla de los punteros. c) se muestra como se actualiza la tabla de punteros cuando cambia un archivo sin alterar el *Snapshot*.**

perfectamente válidas. En la figura 4 se ejemplifica el funcionamiento.

Técnicas de replicación. Básicamente la replicación consiste en espejar un volumen de información origen en un volumen de información destino. El destino puede ser en el mismo *Data Storage* o en un *Data Storage* diferente. Si se trata de uno diferente pude estar ubicado en la red LAN o bien en una red WAN. En el volumen destino de la réplica es un volumen de lectura únicamente.

Para efectuar esta replicación se utilizan diferentes técnicas: la replicación sincrónica y la replicación asincrónica.

La replicación sincrónica es cuando los cambios en los bloques de información se realizan en el mismo momento tanto en el volumen origen, como en el destino; la transacción se completa cuando la actualización se concluye en ambos volúmenes. La replicación asincrónica es cuando se replica la información de los volúmenes en intervalos regulares de tiempo. La replicación sincrónica tiene aplicabilidad para entornos de red LAN, mientras que la asincrónica se utiliza para entornos WAN.

Para este caso, utilizamos la replicación de dos *Data Storage* en una ubicación geográfica distante en un entorno WAN con el fin de protección y *site* de contingencia. Por eso, se describirá con más detalle la replicación asincrónica.

Cuando espejamos asincrónicamente, se aplican las técnicas de *snapshot* dentro del proceso de replicación. Una copia se mantiene en el volumen origen como un punto indicador del estado en que se encuentran los volúmenes espejados. Cuando un nuevo proceso se inicia se realiza un nuevo *snapshot* y el mismo se compara con el *snapshot* anterior a fin de determinar los cambios ocurridos en el volumen, y esas actualizaciones son las que se realizan en el volumen destino. De esta manera, se logra una sincronización incremental entre cada sesión de replicación. Cuando el proceso de replicación asincrónico finaliza, el volumen destino se encuentra espejado al momento del inicio del proceso de replicación.

Como se mencionó, los volúmenes destinos son de lectura únicamente, pero si una situación de contingencia sucede, estos volúmenes pueden ser convertidos a lectura/escritura con el fin de poner en producción en *site* de contingencia (Figura 5).

Para reducir el tráfico de transferencia, la replicación asincrónica utiliza técnica de compresión y descompresión de datos optimizando de esta manera el ancho de banda requerido.

Descripción del escenario de aplicación

El escenario consta de dos sitios, denominados centro de datos A (CDA) y centro de datos B (CDB).

CDA es el centro de datos operativo, al que se conectan los usuarios o clientes dentro de un ambiente de red LAN tanto para consulta, como para actualización de la información. Dentro del mismo operan los servidores que contienen aplicaciones que se encuentran en producción.

CDB es el centro de datos de contingencia que contiene una réplica de los datos y de las aplicaciones de CDA, los usuarios solo se conectan en el caso de contingencia a través de un ambiente de red WAN. Dentro del mismo existen servidores en estado *stand by* que contienen las aplicaciones replicadas que solo se utilizarán en el caso de falla de CDA.

En ambos centros de datos los servidores se encuentran virtualizados y junto con los datos residen en *Data Storage* organizados en volúmenes de datos. Estos volúmenes pueden contener información legible para servidores Linux y para servidores Windows.

CDA está comunicado con CDB a través de un enlace de datos (Figura 6).

La información viaja normalmente en un solo sentido desde CDA hacia CDB con el fin de mantener actualizada la información.

En cada sesión de actualización, o también llamada de replicación, viaja la información adicionada en CDA desde la última sesión de replicación.

Al final de cada sesión de replicación, la información entre ambos centros de datos se encuentra espejada al momento del inicio de dicha sesión.

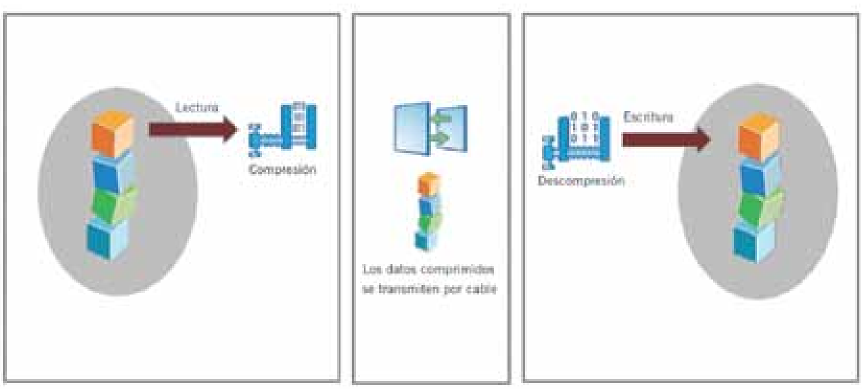

Figura 5. **Replicación. En el *Data Storage* Origen los datos son previamente comprimidos. En el proceso de Replicación transfiere los datos a través de la red. En *Data Storage* Destino los datos son descomprimidos y luego se graban en el dispositivo.**

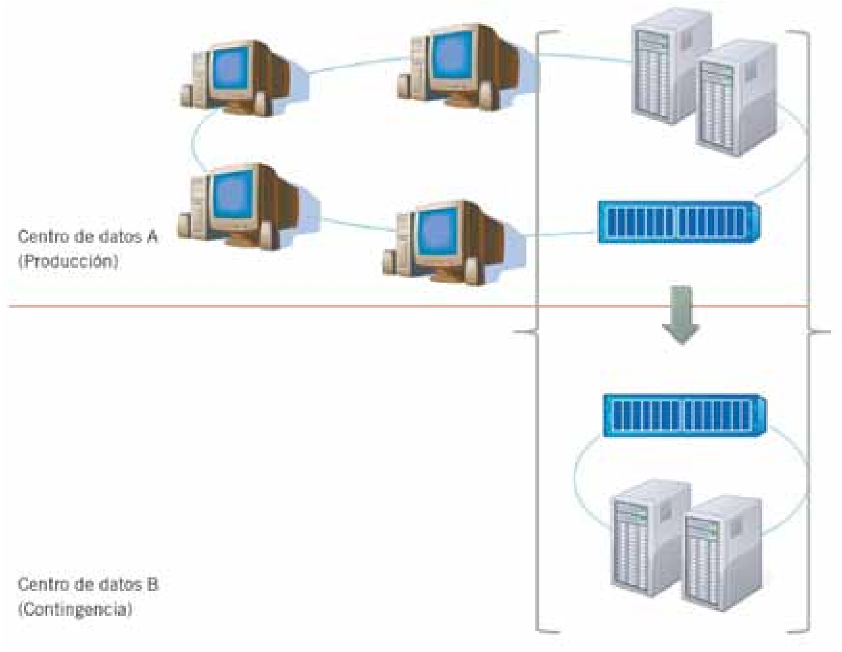

Figura 6. **Clientes en entorno LAN. Enlace punto a punto en una red local.**

El enlace de datos entre CDA y CDB puede ser establecido a través de un enlace punto a punto o a través de una VPN o red privada virtual.

Las redes VPN permiten implementar una red privada utilizando una red pública (Internet). Estas redes se utilizan para realizar conexiones punto a punto, para ello debemos tener acceso a la red pública desde los extremos que queremos conectar. La información que se trasfiere entre ambos extremos viaja encriptada.

Cuando CDA no se encuentre disponible, podemos ejecutar ciertos procedimientos que permitan poner operativo CDB y conectar a los clientes a este centro de datos (Figura 7).

Metodología

A continuación, se enumerarán las tareas que se deberán desarrollar para cada uno de los pilares en el gobierno de información descriptos. La idea de esta metodología es dar un panorama de cómo implementar esta técnica y los puntos que se deben considerar para lograr éxito en el desarrollo de un proyecto de estas características.

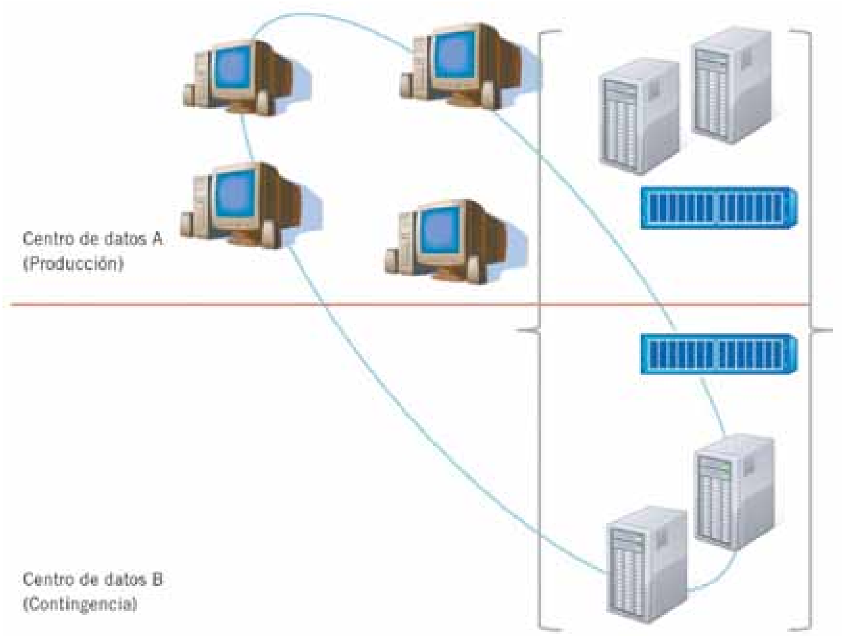

Figura 7. **Clientes en entorno WAN. Enlace entre áreas geográficas extendidas.**

Organización de la información

Catalogación de fuentes de datos

Para poder realizar tareas de cuidado y protección, primero debemos tener claro qué es lo debemos proteger. Por ese motivo el primer punto en la organización de la información es hacer un inventario de todas las fuentes de datos: bases de datos, documentos, imágenes, planillas, etcétera, que dispone la compañía y en la que se deben efectuar las acciones de protección. El producto final de esta tarea es un catálogo de todas las fuentes de datos.

Reorganización de las fuentes de datos

Una vez obtenido el catálogo, se debe analizar y, si es necesario, realizar una restructuración de las fuentes de datos con el objetivo de tener las fuentes de datos ordenadas y agrupadas en volúmenes; por ejemplo, podemos disponer de un volumen con los archivos privados de los usuarios, otros con los archivos públicos organizados por grupos de trabajo; podemos tener volúmenes por aplicaciones: Geología, Ingeniería, Producción, Contabilidad, etcétera. Como producto final de esta tarea se obtienen los volúmenes de datos y el tamaño que ocupa cada uno de ellos.

Administración de la información

Catalogación de aplicaciones

Al igual que en el caso de las fuentes de datos, también se debe catalogar todas las aplicaciones que utiliza la compañía e identificar a cada una de ellas en el servidor correspondiente. Es importante destacar que, en esta implementación, solo se podrán utilizar los servidores virtuales. Si los servidores fueran físicos, primero se debe planificar su virtualización. También debemos registrar qué fuente de datos es utilizada por cada aplicación. En el caso de que la fuente de datos se encuentre en el mismo servidor, se debe precisar esta situación en el catálogo, ya que en ese caso la pieza catalogada tendrá la doble función de aplicación y de fuente de datos. La finalidad de esta tarea es obtener el catálogo de aplicaciones ordenado por servidor e indicar si tienen incluida la fuente de datos.

Clasificación de las aplicaciones

Todos los servidores virtuales se almacenan en volúmenes del *Data Storage*. Con esta tarea debemos agrupar los servidores en volúmenes, un volumen puede contener uno o varios servidores. Las sincronizaciones en el *site* de contigencia se realiza por volúmenes de datos, con lo cual los servidores que se encuentren en el mismo volumen compartirán la misma sesión de sincronización. Como producto final de esta tarea se obtienen los volúmenes de aplicaciones y el tamaño que ocupa cada una de ellas.

una menor redundancia en el arreglo de discos. Además, se debe seleccionar la ubicación del *site* de contingencia, puede ser un lugar propio o de terceros, pero deberá estar ubicado en un lugar geográfico distinto. El producto final de esta tarea es el lugar de contingencia equipado.

Definir ventana de tiempo para la sincronización

De acuerdo con las necesidades del negocio, debemos definir nuestra ventana de tiempo de sincronización en función del tiempo de operaciones que podemos perder en el caso de ocurrencia de una contingencia. Es decir, al momento de ocurrencia de la contingencia y al activar el *site* alternativo, los volúmenes del *site* de contingencia contendrán la información actualizada en el momento de inicio de la última sesión de sincronización finalizada. Podemos definir una ventana de tiempo diferente para cada uno de los volúmenes. Otro

tre el inicio de una sesión de sincronización y la otra. Es decir, si tenemos una ventana de 6 horas, estimaremos el volumen de información entre esas 6 horas. Esta no es una tarea fácil de realizar, ya que a veces es muy difícil estimar volúmenes incrementales de información, porque dependemos del comportamiento de las aplicaciones y de cuanta información varía entre los volúmenes; no implica solo datos ingresados, sino también todos los bloques de datos que han variados en ese lapso. La mejor manera de abordar esta tarea es haciendo una simulación de cada volumen en el entorno de producción. Se debe crear un volumen espejo en el *Data Storage* de producción, realizar las sincronizaciones en la ventana de tiempo establecida y registrar las variaciones de volumen para cada uno de los ciclos. Es importante realizar varios ciclos de sincronización en diferentes días y horas, y calcular un promedio. Si no tenemos espacio físico suficiente en el entorno de producción, podemos realizar esta

Protección

Aprovisionamiento

Con todos los volúmenes ya catalogados ahora procedemos a calcular los recursos para el aprovisionamiento del *site* de contingencia. Debemos determinar, de acuerdo con las máquinas virtuales por espejar, la cantidad de los hipervisores necesarios y el espacio físico necesario en el *Data Storage* para contener a todos los volúmenes catalogados, ya sea los correspondientes a los volúmenes de las fuentes de datos o los correspondientes a las máquinas virtuales.

El *Data Storage* de contingencia podrá contener menores requisitos de seguridad que el de producción. Por ejemplo, si en producción tenemos doble controladora, en contingencia podemos tener solo una y también

factor que debemos considerar es la frecuencia de actualización de los volúmenes, podremos tener volúmenes que se actualizan permanentemente y otros que se actualizan con una menor frecuencia. Por ejemplo, podemos definir ventanas de tiempo de 24 horas, 12 horas 8 horas, 4 horas, 2 horas, etcétera. El producto final de esta tarea es la obtención de las ventanas de tiempo para cada uno de los volúmenes catalogados.

Estimar el delta de variación de información para la sincronización

Ya tenemos los volúmenes y los tiempos, ahora es el momento de estimar el tamaño de la información que se verá actualizada entre la ventana de tiempo. La ventana de tiempo indica además el tiempo que transcurre en-

tarea de un volumen a la vez o podemos montar primero el *Data Storage* de contingencia en nuestra red LAN antes de migrarla al *site* de contingencia, y de esta manera obtener todos los deltas de variación de los volúmenes.

Dimensionar el ancho de banda

Con las ventanas de tiempo más los deltas de variación entre las sesiones de sincronización podemos definir el ancho de banda necesario para tener sincronizados todos los volúmenes. Para optimizar el ancho de banda tenemos que tratar de ubicar las sesiones de sincronización espaciadas unas de otras de manera que la menor cantidad de volúmenes se estén sincronizando en el mismo momento. Podemos corregir alguna ventana de tiempo o correr el inicio de sincronización ajustando así lo definido en los puntos anteriores con el objetivo de optimizar las comunicaciones. Una vez definido el ancho de banda, podemos realizar simulaciones de sincronización con diferentes anchos de banda a fin de confirmar los valores estimados. Actualmente, existen *routers* con esta posibilidad y algunos son de uso libre.

Establecer un método de sincronización alternativo

Como todos los cálculos fueron efectuados con los valores promedios obtenidos, esto será útil para la mayor parte de las sesiones de sincronización, pero debemos tener un método de procesamiento alternativo para mantener la información espejada en el momento en que estos valores sean superados por un pico de volumen a transferir y que no pueda ser abastecido por el enlace principal. Si no contamos con una alternativa, se producirá un cuello de botella y la ventana de tiempo de sincronización quedará desbordada. Un método alternativo que podemos sugerir es la utilización de una sincronización de inicialización en cinta donde la totalidad de la información del volumen se copia a cinta (LTO 5), la misma se envía al *site* de contingencia y en ese lugar se vuelca el contenido de la cinta al *Data Storage* de contingencia, una vez finalizada se continúa con las sesiones de sincronización programadas. Existen herramientas de monitoreo que advierten cuando esta situación ocurre y ese es el momento para disparar el método de sincronización alternativo.

Establecer pruebas

Antes de poner en producción esta técnica se deben realizar pruebas de funcionamiento con el objetivo de corroborar en el terreno real la eficacia de todos los componentes involucrados en esta tarea. De las pruebas pueden surgir correcciones o mejoras que deben evaluarse previamente, ya que, una vez puesto en producción, estos ajustes suelen ser más complicados de realizar, además de poner en riesgo la disponibilidad del *site* de contingencia.

Acceso

Establece los procedimientos de operaciones normales y de monitoreo

Se deben establecer y documentar todas las actividades que se deban llevar a cabo para mantener operativo el *site* de contingencia. Se deben incluir todos los procesos de actualización de información involucrados, así como también se deben definir las herramientas de monitoreo y las alertas necesarias para controlar el funcionamiento adecuado de los procesos y de las comunicaciones. Completar esta documentación implica agregar los diagramas de plataforma y telecomunicaciones.

Establecer procedimientos de contingencia

Se deben incluir los procedimientos con el detalle paso a paso de las actividades que se realizarán al momento de ocurrir una contingencia en el Centro de Datos Principal, se debe incluir un plan de comunicaciones y definir el nivel gerencial que tomará la decisión de activar el Plan de Contingencia. Según el tipo de falla ocurrida, se podrá activar el plan de forma parcial o total. Para establecer qué elementos del plan se activarán, primero se deberá realizar una reunión técnica y determinar el alcance de la contingencia, el grado de afectación de los componentes del *site* principal y los componentes necesarios que se activarán para sobrellevar la contingencia. También es importante estimar el tiempo de restauración del *site* principal, es decir, de acuerdo con la gravedad del siniestro, cuánto tiempo pasará hasta que el *site* principal se encuentre plenamente operativo. También se deberá establecer cómo y desde que lugar se conectarán los usuarios en el entorno WAN al *site* de contingencia, si existen o no puestos de trabajos alternativos en otro edificio, en otra sucursal o bien si podrán tener actividad desde un sitio fuera de los dominios de la compañía.

Establecer procedimientos de restauración

Una vez finalizada la contingen-

cia, se deberán detallar los pasos necesario para restaurar el *site* principal. Por ejemplo, ¿cómo retornaran la información actualizada al *site* principal los volúmenes de información en el *site* de contingencia? Una de las posibilidades, depende del tiempo transcurrido, es realizar una sincronización inversa o bien se deberá optar por un resembrado total de los volúmenes. En ese caso, se deberá evaluar la utilización de la cinta o bien el traslado del *Data Storage* al ambiente LAN para actualizar los volúmenes del ambiente de producción.

Establecer pruebas de acceso desde cliente interno/oficinas externas

Realizar un plan de pruebas o testeo de los procedimientos de contingencia es fundamental para ajustar los detalles del plan al momento de su aplicación. De esta manera se asegura que en el momento de ocurrencia del siniestro, el plan se encuentre perfectamente probado y con garantía de aplicación exitosa. La simulaciones deben efectuarse con cierta periodicidad o cuando se efectúen cambios de importancia en el plan original.

Beneficios principales

Según lo descripto en este trabajo, los beneficios principales obtenidos con la adopción de esta técnica, en cuanto al resguardo de información y al mantenimiento de la continuidad operativa, son los siguientes:

1. Mejora de los estándares de seguridad.
2. Facilita la implementación de un *site* de contingencia.
3. Automatiza las operaciones de resguardo.
4. Minimiza el tiempo de indisponibilidad de los servicios de TI ante fallos graves.

Adicionales

A las ventajas obtenidas se pueden añadir algunos beneficios secundarios que están asociados y que agregan valor a la solución descripta, logrando una mejor relación en la ecuación de costos y beneficios.

- **Ambiente de pruebas**

Uno de los aspectos de utilización agregada del *site* de contingencia es emplear el mismo como ambiente de pruebas.

A toda la infraestructura que ha sido montada para utilización eventual podemos darle un uso alternativo y cotidiano.

La flexibilidad de utilización de tecnología de avanzada de los *Data Storage* permite crear *snapshot* o imágenes de *backups* de los volúmenes replicados en el *site* de contingencia, y a estos nuevos volúmenes los podemos configurar de lectura/escritura sin afectar el volumen original.

Esta tecnología permite montar rápidamente entornos de test *On the Flight*. El ciclo de vida de estos volúmenes puede ser de corto o de largo plazo y podemos crear varios volúmenes de prueba a partir de uno original creando de esta manera escenarios virtuales de prueba. A esta función se la denomina "Clonación de volúmenes" y no conlleva mayores costos y recursos adicionales.

Esta posibilidad está presente para ser aplicada tanto en los volúmenes de datos como en los volúmenes de aplicaciones.

- **Archivado a cinta remoto**

Más allá de todas estas técnicas de *back ups*, algunos entornos requieren el almacenamiento *Offline* de los datos. Cuando un dispositivo de cinta está unido a un *Data Storage*, los datos se pueden mover a la cinta periódicamente. Es decir, el *site* de contingencia se puede usar para realizar las operaciones de copia a cinta que antes se hacían en el *site* de producción, de esta manera se alivia el tráfico de datos y la carga de trabajo del *Data Storage* de producción.

Agradecimientos

A Darío Jalí de Acsys S.A. por la colaboración y el aporte técnico en este trabajo.

Bibliografía

Shawn Preissner (2014), *Technical Report Netapp*.

Virtualización (2014), Sitio web de vmware, *http://www.vmware.com*, acceso 14 de junio de 2014.

Srinath Alapati, Darrin Chapman (2009), *Technical Report Netapp*.

Jorge Nardelli (1984), Auditoría y Seguridad de los Sistemas de Computación. Gordon Smith (1998), La última Frontera.

Robert Murdick y John Munson (1988), Sistemas de Información Administrativa. Microsoft Press, 1996, Fundamentos de Redes.

BICSI LAN (1996), *Design Manual*.

Comité 802, 1980, *The Institute of Electrical and Electronics Engineers* (IEEE).