UTILIZACION DE TÉCNICAS DE REPLICACION COMO ESTRATEGIA DE BACKUP Y SITE DE CONTINGENCIA

Lic. Julio César Panno - julio_panno@cgc.com.ar
Compañía General de Combustibles S.A.

Keywords: Replicación, Virtualización, Contingencia, Backup, Data Storage

ABSTRACT

El objetivo buscado es el mantener reguardados datos y aplicaciones en un lugar alternativo o de contingencia utilizando procedimientos desatendidos dentro de una ventana de tiempo compatible con las necesidades y requerimientos del negocio.
La integración de herramientas de virtualización y de protección de la información permite contar con un escenario en donde los datos y las aplicaciones se encuentren protegidos y disponibles y con tiempo de recuperación mínimos independientemente del tipo de falla ocurrida.
Asimismo poder utilizar estas herramientas, en el caso de ocurrencia de una contingencia del sitio principal o de procesamiento, manteniendo de esta manera disponible los datos, aplicaciones y procesos vitales de la compañía.
El avance tecnológico y la evolución de las técnicas de replicación, permiten garantizar la fidelidad de los resguardos realizados entre los sitios y la infraestructura requerida en función de sus múltiples utilizaciones, permite lograr una mejor optimización de los costos en los presupuestos generales del área de TI.
Independientemente del objetivo principal, a partir de la incorporación de estas técnicas, redundan en beneficios adicionales siendo más amplio el ámbito de aplicación.
La experiencia lograda en este ámbito nos permite describir una metodología de implementación y los principales aspectos a tener en cuenta para lograr tener éxito con unas de los premisas básicas de la informática: “Resguardar el Valor de la Información” y unos de los fundamentos de las Organizaciones: “Mantener la continuidad Operativa”.

INTRODUCION

Hoy en día se destaca a “La información” como uno de los recursos valiosos con lo que disponen las compañías para lograr un su 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 se establece de qué manera se agrupa y como se interrelaciona y se integra con el entorno organizacional.

Administración de la información:
La administración determina cuales son las aplicaciones y los procesos que gestionan la información, quienes son los usuarios que la colectan, actualizan y utilizan. Las aplicaciones que organizan y administran esta información, así como también los procesos de negocios en los cuales 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, planificación, implementación y mejoramiento de los flujos asociados a la entrega y 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, motivo por el cual 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 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, 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 de la siguiente manera:
"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 que se utilizan son privados y se utilizan generalmente para conectar computadoras y otros dispositivos de red dentro de un mismo edificio.

Redes WAN

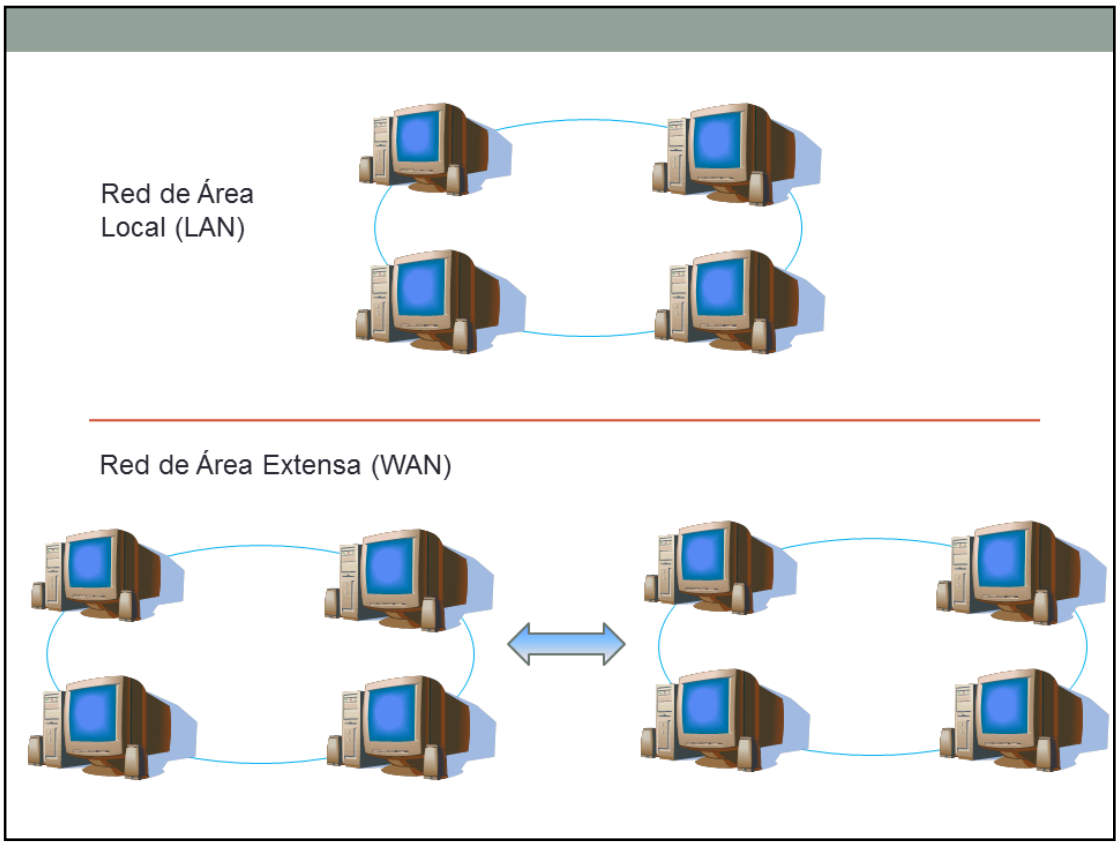
(Wide Area Network) que representa la comunicación entre redes de área local a través de distancias geográficas extendidas. A menudo, los canales de comunicación utilizados para actividad WAN son de propiedad de un tercero, como una compañía telefónica.

LAN vs. WAN

Las palabras área geográfica limitada se utilizan en la definición de una red LAN para resaltar el hecho de que el 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:

FIG.1



Ancho de Banda

Es la 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 de por ejemplo 100 Mbps (Megabits por segundo), 1 Gbps (Gigabit por segundo).
Para los enlaces tipo WAN generalmente referimos a velocidades menores como ser 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 utilizadas en forma conjunta posibilitan el desarrollo de esta solución:

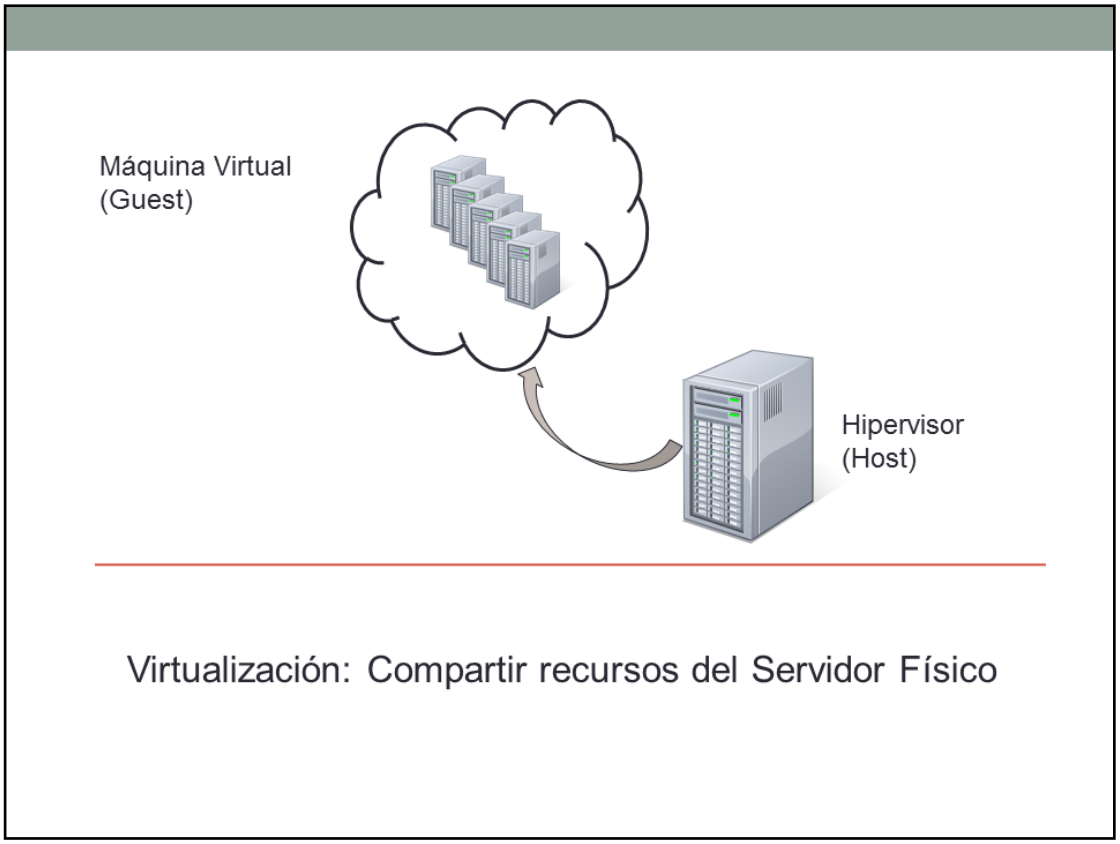
- ✓ Virtualización de Servidores
- ✓ Almacenamiento Centralizado de Disco (Data Storage)
- ✓ Técnicas de Replicación

Virtualización de Servidores

La arquitectura tradicional de servidores permite que se ejecuten sólo 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, 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.

FIG.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. Esto hace que se puedan 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 por ejemplo 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 la misma que la 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.

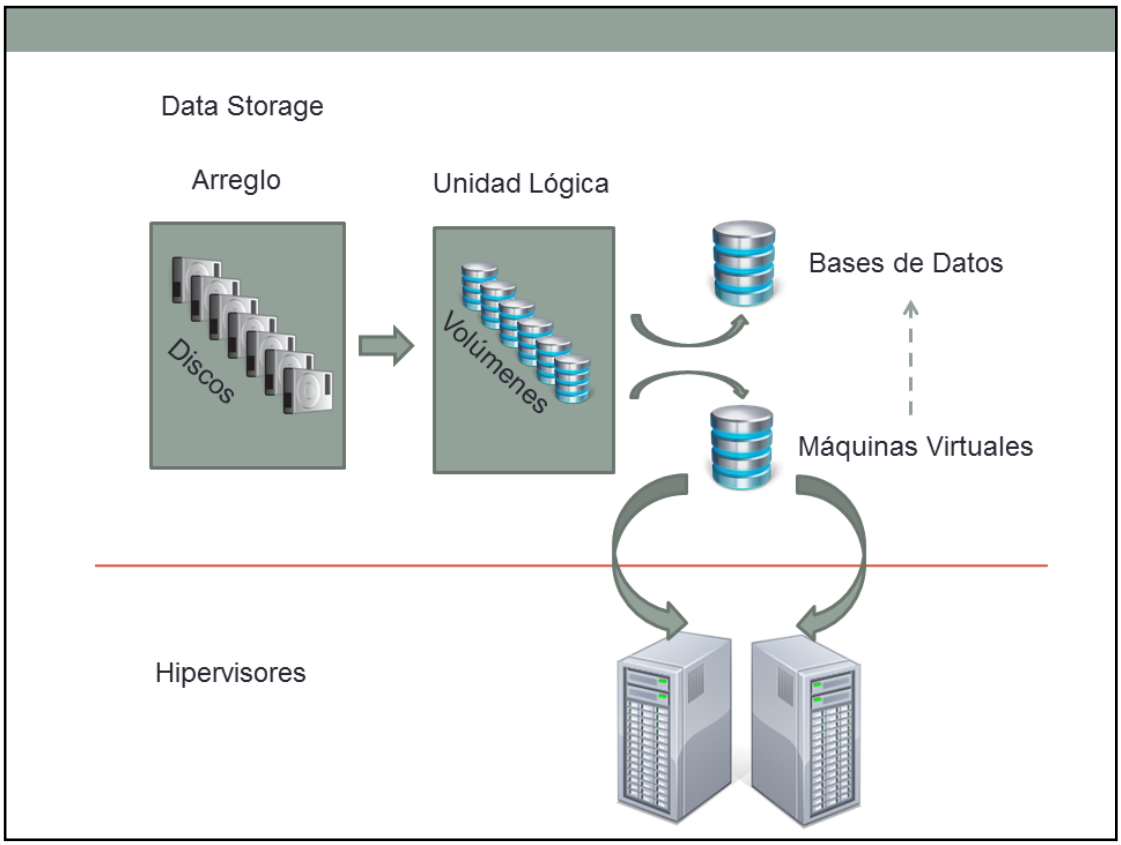
La mejoría que nos ofrece esta característica es que se pueden correr procesos, que requieren de SO distintos al mismo tiempo. De esta forma se pueden aprovechar las cualidades propias de cada sistema operativo, sin tener que cambiar de hardware.
Las razones del porque utilizar Máquinas Virtuales es porque ocasiona ahorros en espacio de memoria física, mejora el aprovechamiento de los recursos disponibles y reduce costos de mantenimiento de equipo. Nos permite mejorar el aprovechamiento del equipo físico porque en general, nunca se llega a utilizar todos los recursos de un servidor físico al mismo tiempo, utilizando los recursos que de otra forma estarían ociosos.
Racionaliza la proliferación de servidores, el espacio físico, el uso de energía y 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 que seguirá aumentando, la virtualización ha demostrado ser una tecnología muy potente para simplificar el despliegue de software y servicios, dotando al Centro de Datos de mucha más agilidad y flexibilidad.
Además, podemos aludir la seguridad que nos proveen 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 están en riesgo de perder o exhibir información confidencial. Lo anterior se debe 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 los mismos 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. Motivo por el cual es esencial dotar a este dispositivo de componentes redundantes que garantizan su funcionamiento a pesar de fallas. También es fundamental la optimización del acceso a los datos para poder procesar múltiples operaciones de lectura y escritura en el dispositivo.
Esta tecnología cuenta con distintos conceptos tecnológicos, siendo uno de los principales 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.
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. Este tipo de tecnología permite seguir operando sin pérdida

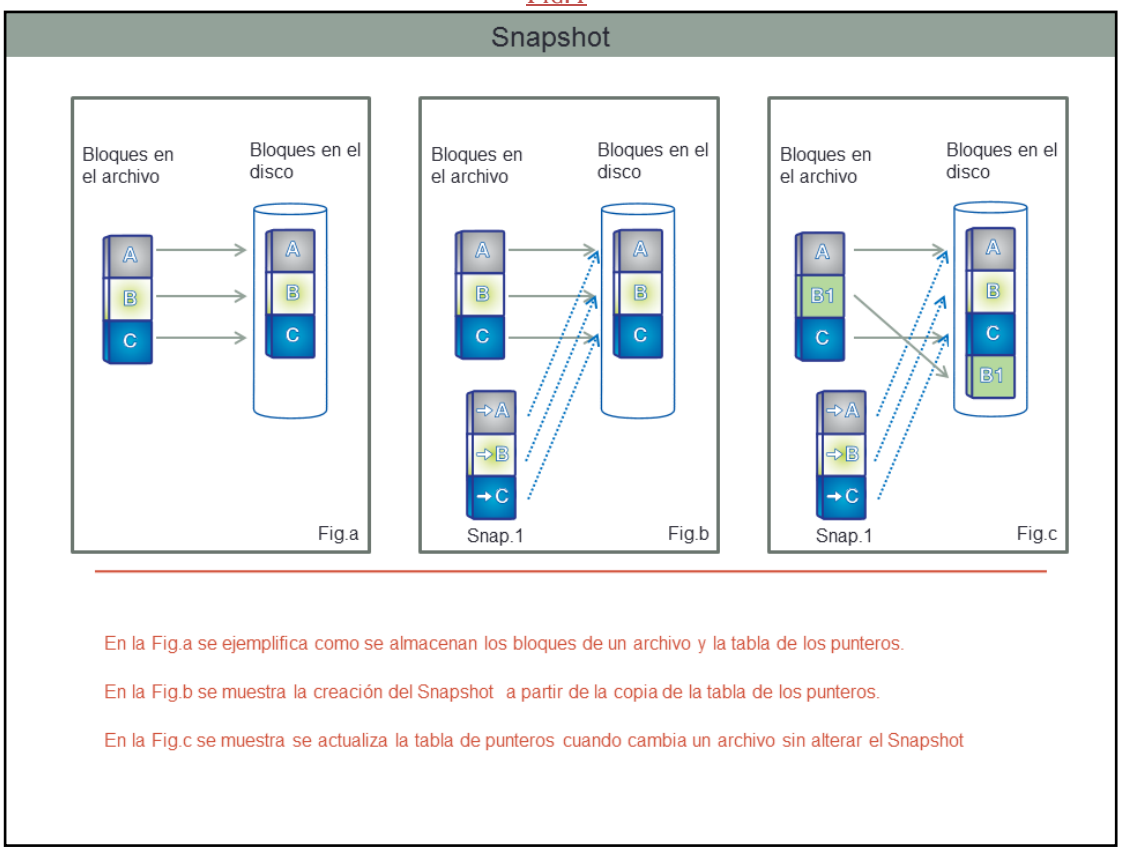
de información si alguno de los discos del arreglo falla, ya que la información se encuentra diseminada y/o replicada entre las distintas unidades de almacenamiento.
Esta implementación pueden soportar el uso de uno o más discos de reserva (hot spare), que son unidades preinstaladas que pueden usarse inmediatamente (y casi siempre automáticamente) tras el fallo de un disco del RAID. El software que administras 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 a 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. Sin embargo los distintos fabricantes que ofrecen almacenamiento centralizado cuentan también con tecnología inteligente en controladoras, pudiendo distribuir la carga de trabajo física de los discos que permite el acceso simultáneo y garantizar una cantidad aceptable de entradas y salidas de datos necesaria para cumplir con las demandas de las aplicaciones.

FIG.3



El Data Storage se conecta a las redes de comunicación de una compañía, permitiendo 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 permita la implementación de soluciones de virtualización de servidores, ya que las Máquinas Virtuales son almacenadas en el Data Storage pudiendo 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. ¿Cómo la tecnología resuelve esto? Una de las maneras para lograrlo 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 sólo 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 perfectamente válidas. En la figura de abajo se ejemplifica el funcionamiento.

FIG.4



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 puede tratarse de 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.

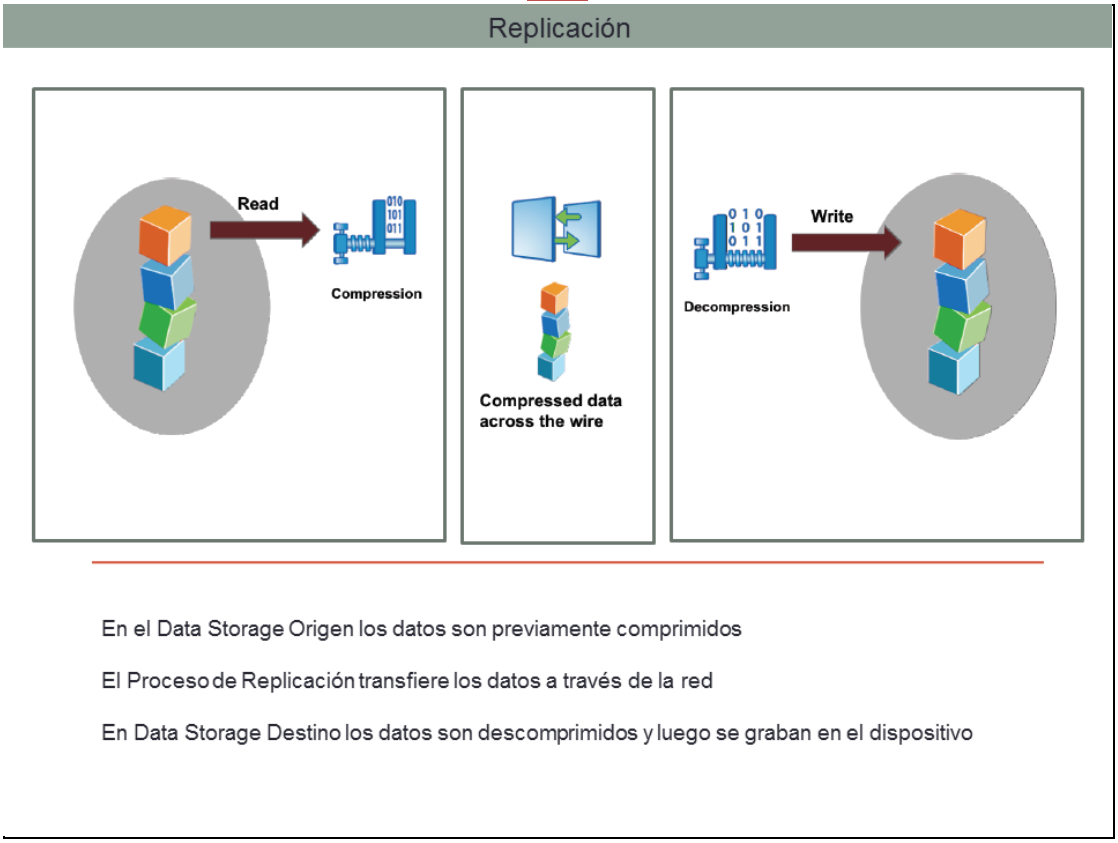
Para efectuar esta replicación existen diferentes técnicas utilizadas: La “Replicación Sincrónica” es cuando los cambios en los bloques de información se realizan en el mismo momento tanto en el volumen origen como en el destino y la transacción se completa cuando la actualización se concluye en ambos 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 rede 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. A tal fin se describirá con más detalle la Replicación Asincrónica.

Cuando espejamos asincrónicamente, se aplican las técnicas de Snapshop 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 con el fin de determinar los cambios sufridos 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 anteriormente 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.

FIG.5

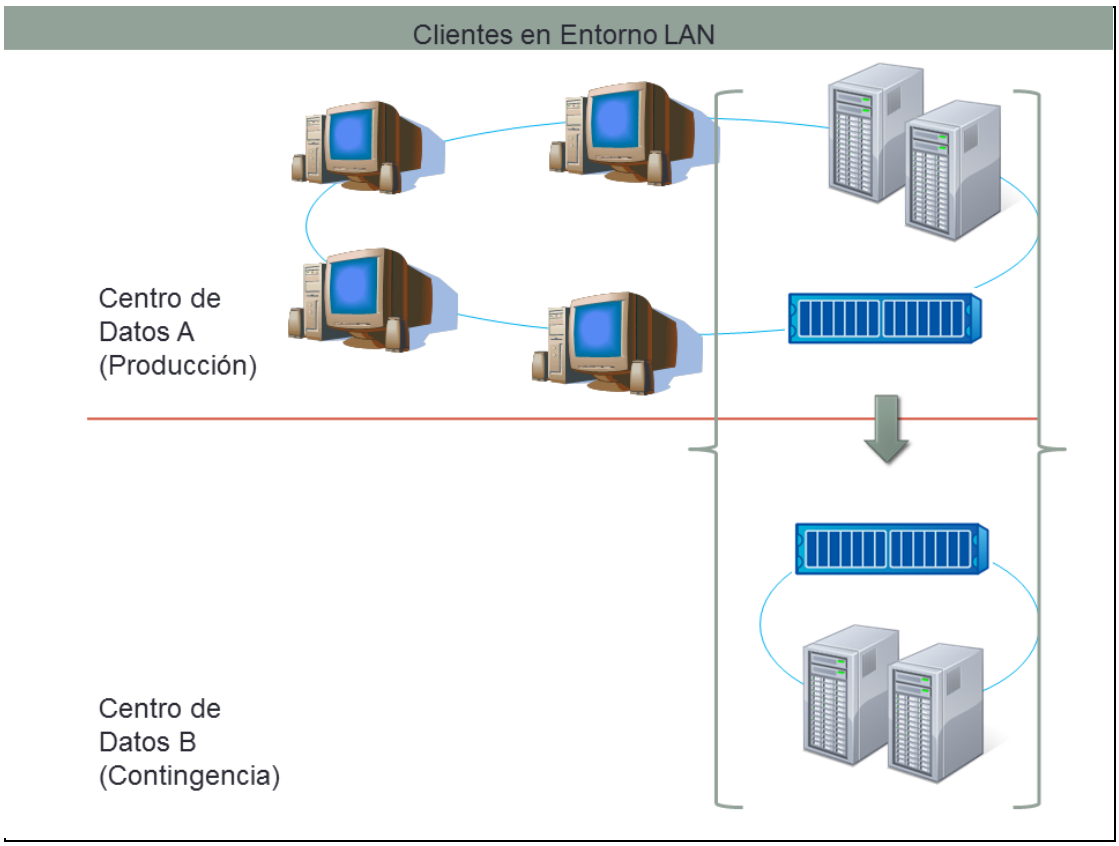


Para reducir el tráfico de transferencia, la Replicación Asincrónica utiliza técnica de compresión y descompresión de datos optimizando de esta manera el ancho de banda requerido.

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 cual 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 sólo 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 contiene las aplicaciones replicadas que sólo 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.

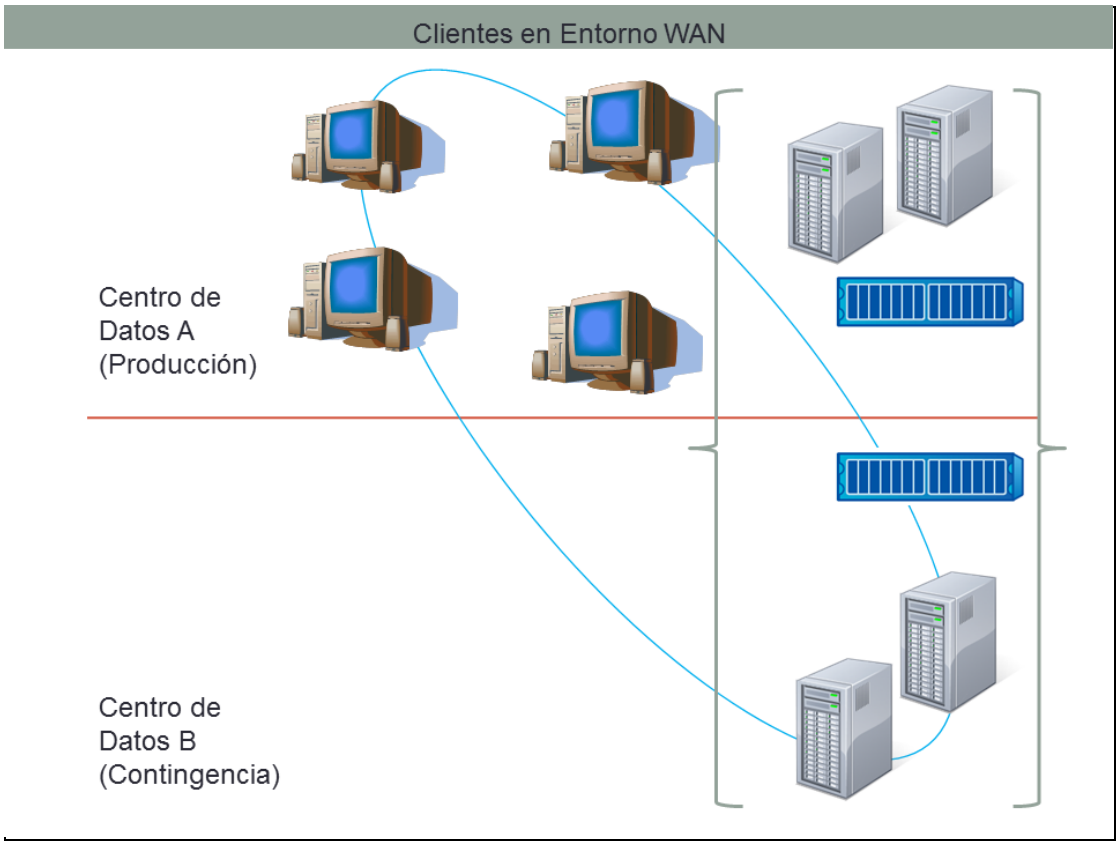
FIG. 6



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

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.
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 nos 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 nos permitan poner operativo CDB y los clientes conectarlos a este centro de datos.

FIG. 7



METODOLOGIA

Para cada uno de los pilares en el gobierno de información descriptos, se enumeran las tareas a desarrollar. La idea de esta metodología es dar un panorama de cómo implementar esta técnica y todos los puntos a tener en cuenta para tener éxito en el desarrollo de un proyecto de estas características.

ORGANIZACIÓN DE LA INFORMACIÓN

Catalogación de fuentes de datos

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

Reorganización de las fuentes de datos

Obtenido el catalogo, se debe analizar y de ser necesario realizar una restructuración de las fuentes de datos. El fin de esta tarea es tener las fuentes de datos ordenarlas 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, etc. El producto final de esta tarea es obtener los volúmenes de datos con los tamaños que ocupan cada uno de ellos.

ADMINISTRACIÓN DE LA INFORMACIÓN

Catalogación de aplicaciones

Del mismo modo que las fuentes de datos, se debe catalogar todas las aplicaciones que utiliza la compañía, identificando a cada una de ellas en el servidor correspondiente. Es importante destacar que sólo los servidores virtuales son los que se podrán utilizar en esta implementación. Si los servidores fueran físicos primero se debe planificar su virtualización. También debemos registrar cuál de las fuentes de datos es utilizada por cada aplicación. En el caso que la fuente de datos se encuentre en el mismo servidor se debe precisar esta situación en el catálogo ya que en este caso esta pieza catalogada tendrá la doble función de aplicación y fuente de datos. La salida de esta tarea es obtener el catálogo de aplicaciones ordenado por servidor e indicando si tienen incluida la fuente de datos.

Clasificación de las aplicaciones

Todos los servidores virtuales se almacenan en volúmenes del Data Storage, En esta tarea debemos agrupar los servidores en volúmenes, un volumen puede contener uno o varios servidores. La 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. El producto final de esta tarea es obtener los volúmenes de aplicaciones con los tamaños que ocupan cada uno de ellos.

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 a las máquinas virtuales a espejar, la cantidad de los hipervisores necesarios y el espacio físico necesario en el Data Storage para contener todos los volúmenes catalogados, ya sea los correspondientes a los volúmenes de las fuentes de datos y a 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 sólo una, y por ejemplo también una menor redundancia en el arreglo de discos. Se debe también seleccionar la ubicación del Site de Contingencia, podrá ser un lugar propio o de terceros, pero deberá estar ubicado en un lugar geográfico distinto. El producto final de esta tarea es obtener el lugar de contingencia ya equipado.

Definir ventana de tiempo para la sincronización

De acuerdo a 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 al momento de inicio de la última sesión de sincronización finalizada. Podemos definir una ventana de tiempo diferente para cada uno de los volúmenes. Otro factor a 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 hs, 12 hs. 8, hs. 4, hs. 2 hs. etc. El producto final de esta tarea es tener todas 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 nos indica además el tiempo que transcurre entre 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 es muy difícil a veces estimar volúmenes incrementales de información, ya que dependemos de cómo se comportan las aplicaciones y cuanta información varía entre los volúmenes, no implica datos ingresados sino también todos los bloques de datos que han variados en ese lapso de tiempo. La mejor manera de realizar esta tarea realizando 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 tarea de un volumen a la vez o si no podemos montar primeramente el Data Storage de contingencia e nuestra red LAN, antes de migrarla al Site de Contingencia, y de esta manera obtener todos los delta 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 poder 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 tal 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 de esta manera lo definido en los puntos anteriores con el fin de optimizar las comunicaciones. Definido el ancho de banda podemos realizar simulaciones de sincronización con diferentes anchos de banda con el fin de confirmar los valores estimados. Existen en la actualidad routers que tienen esta posibilidad y algunos son 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 poder 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 el de la utilización de una sincronización de inicialización en cinta en 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 nos advierten cuando esta situación ocurre y en ese momento debemos disparar el método de sincronización alternativo.

Establecer pruebas

Antes de poner en producción esta técnica se deben realizar pruebas de funcionamiento con el fin 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 riego 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 con el fin de controlar el funcionamiento adecuado de los procesos y de las comunicaciones. Para completar esta documentación, se deben 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 a realizar 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. De acuerdo al tipo de falla ocurrida se podrá activar el plan de forma parcial o total. Para establecer cuáles de los elementos del plan se activará, se deberá primeramente realizar una reunión técnica para determinar el alcance de la contingencia, el grado de afectación de los componentes del Site Principal y los componentes necesarios a activar para sobrellevar la contingencia. Es importante también poder estimar el tiempo de restauración del Site Principal, es decir, de acuerdo a la gravedad del siniestro cuando tiempo pasará hasta que el Site Principal se encuentre plenamente operativo. También se deberá establecer cómo los usuarios y desde que lugar se conectarán 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 contingencia, se deberán detallar los pasos necesario para restaurar el Site Principal. Por ejemplo ¿cómo los volúmenes de información en el Site de Contingencia retornaran la información actualizada al Site Principal? Una de las posibilidades, dependiendo del tiempo transcurrido es realizar una sincronización inversa o bien se deberá optar por un resembrado total de los volúmenes. En este 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 importante para poder ajustar los detalles del plan al momento de su aplicación. De esta manera nos aseguramos que al 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 bien cuando se efectúen cambios de importancia en el plan original.

BENEFICIOS

Beneficios Principales

Los beneficios principales obtenidos, con la adopción de esta técnica, en cuanto a resguardo de información y al mantenimiento de la continuidad operativa, según todo lo descripto anteriormente en el desarrollo de este trabajo son los siguientes:

1. Mejora de los estándares de seguridad.
2. Facilita la implementación de un Site de Contingencia
3. Automatización de las operaciones de resguardo.
4. Minimiza el tiempo de indisponibilidad de los servicios de TI ante fallos graves.

Beneficios Adicionales

Adicionalmente, a las ventajas obtenidas se pueden añadir algunos beneficios secundarios que vienen 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 poder utilizar el mismo como ambiente de pruebas.
  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 nos 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 al volumen original.
  Esta tecnología nos 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 n 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 descriptas, 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 utilizar para realizar las operaciones de copia a cinta que antes se realizaban en el Site de Producción, aliviando de esta manera el tráfico de datos y la carga de trabajo del Data Storage de Producción.

AGRADECIMIENTOS

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 Wmware, <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, Tthe Institute of Electrical and Electronics Engineers (IEEE)