

Universidad de las Ciencias Informáticas
Facultad 2



**Título: Procedimiento para los procesos de
almacenamiento y salvados en Centros de
Procesamientos de Datos.**

Trabajo de Diploma para optar por el título de
Ingeniero Informático

Autor: Yoelvys Martínez Hidalgo

Tutor: Ing. Orlando Beneitez Ramírez

Ciudad de la Habana, Junio de 2012

“Año 54 de la Revolución”

PENSAMIENTO.



“Hombre recogerá quien siembre escuela.”

José Martí

DECLARACIÓN DE AUTORÍA.

Declaro ser autor de la presente tesis y reconozco a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Firma Autor

Yoelvys Martínez Hidalgo

Firma Tutor

Orlando Beneitez Ramírez

DATOS DE CONTACTO.

Nombre y Apellidos: Orlando Beneitez Ramírez

Ciudadanía: Cubano

Institución: Universidad de las Ciencias Informáticas.

Título: Ingeniero en Telecomunicaciones.

Categoría: Adjunto a la producción

E-mail: orlandobr@uci.cu

Cargo: Especialista General

AGRADECIMIENTOS.

A mis padres, mi hermana, mima, Vicente, Yasniel, Orelvys, tío nene, tía negra, Yami, Imilda, Zoila, Nico, Barbarita, Maribel, Guillermo, Maritza, Wilian, Maira, Maricel y familia.

A mis tíos y primos maternos y paternos y sus familias.

A toda las personas de mi pequeño barrio, a mis vecinos Olivia, Vidal, Juanito, Maiquel, Pepe, Carmita, Rosa.

A todos mis compañeros de la universidad, los del equipo de futbol 11 y futbol sala, los de mi grupo de primer año (2101) y los que han compartido apartamento conmigo estos 5 años

A todos mis amigos de Venezuela con los que compartí buenos y malos momentos.

A Betty por sus consejos y soportar mis malas acciones y siempre estar ahí.

A mis amigos de la universidad Yadier, Dasiel, Oscar Moncada, Oscar Calderín, Orelvys, Yoan, Laritza, Puchi, Huse, Danae, Yoanny, Manuel, Shampoo, dani, Ernesto, Wilber, Charchabal, Yei, Yanara, Francisco, Duany, Pipo, Juan A .

A los socios del doble, por las horas de angustia y cuero.

A las amigas de la universidad que conocí en mi viaje a venezuela, Melixa y Yuya,

A Yenia Garcia Vargas por su apoyo, cariño y amor incondicional

A todos los profesores que he conocido durante este tiempo en especial a las profes Aime y Daimara.

A los profes Orlando, Álvaro, Alexei, Adrián, Miguel, pues sin su ayuda nada hubiese sido posible.

DEDICATORIA.

A mis padres y mi hermana, por todo el amor que me han dado y su apoyo en los momentos difíciles.

A mi madre más anciana, que me demostró que madre no es una sola y me ayudó todo lo que pudo durante mis estudios.

A mi tercera madre “mima”, Vicente, Orelvys y Yasniel, por el cariño y apoyo que me han dado desde pequeño.

A mis abuelos y familia paterna, por su amor y cuidado cuando los visito.

A mi familia materna, por siempre estar ahí cuando lo he necesitado y su apoyo incondicional.

A toda las personas de mi barrio, por su ayuda y su apoyo durante mi carrera.

A Dani y su familia, por su comprensión durante el tiempo que fuimos novios y ayudarme dentro de sus posibilidades.

A todos mis amigos de Venezuela, por ser mi gran familia y tenerme como su hijo más pequeño, en especial a Betty, Vilma, Raiza, Arturo, Raúl, Ezequiel, Made, Eva, Carmen y Cari.

A mis amigos de la UCI, por los momentos felices y divertidos que pasamos juntos.

A una persona muy especial en mi vida, una persona que me ayudó en Venezuela y aun aquí lo hace, una persona cuyo cariño y comprensión fue y es insuperable, una persona que sin palabras rebuscadas te endulza el corazón y hace más agradable tu día , a ti Yenia, dedico esta tesis con todo mi corazón.

GLOSARIO DE TÉRMINOS

Bytes: término utilizado en la informática comúnmente como unidad básica de almacenamiento de datos.

FTP: del inglés *File Transfer Protocol*: protocolo utilizado para tener acceso a un anfitrión [host] de Internet, y posteriormente para transferir archivos entre ese anfitrión y la computadora que se está utilizando.

GNU/Linux: es uno de los términos empleados para referirse a la combinación del núcleo o *kernel* libre similar a Unix denominado Linux.

GPL: del inglés *General Public License*: es una licencia orientada principalmente a proteger la libre distribución, modificación y uso de software.

iSCSI: abreviatura de *Internet SCSI*: es un estándar que permite el uso del protocolo SCSI sobre redes TCP/IP.

ISO: del inglés *International Organization for Standardization*: es el organismo cuya función principal es la de buscar la estandarización de normas de productos y seguridad para las empresas u organizaciones a nivel internacional.

LDAP: del inglés *Lightweight Directory Access Protocol*: es un protocolo a nivel de aplicación que permite el acceso a un servicio de directorio ordenado y distribuido para buscar diversa información en un entorno de red. Habitualmente, almacena la información de autenticación (usuario y contraseña) y es utilizado para autenticarse aunque es posible almacenar otra información (datos de contacto del usuario, ubicación de diversos recursos de la red, permisos, certificados, entre otros datos).

Licencia BSD: es la licencia de software otorgada principalmente para los sistemas BSD (*Berkeley Software Distribution*). La licencia BSD al contrario que la GPL permite el uso del código fuente en software no libre.

NAS: del inglés *Network Attached Storage*: es una tecnología de almacenamiento dedicada a compartir la capacidad de almacenamiento de un computador (Servidor) con ordenadores personales o servidores clientes a través de una red (normalmente TCP/IP).

NFS: del inglés *Network File System*: es un protocolo de nivel de aplicación, según el Modelo OSI. Es utilizado para sistemas de archivos distribuido en un entorno de red de computadoras de área local.

Posibilita que distintos sistemas conectados a una misma red accedan a ficheros remotos como si se tratara de locales.

RSYNC: es una aplicación de software para sistemas de tipo Unix que ofrece transmisión eficiente de datos incrementales, que opera también con datos comprimidos y cifrados.

SAMBA: es una implementación libre del protocolo de archivos compartidos de Microsoft Windows (antiguamente llamado SMB, renombrado recientemente a CIFS) para sistemas de tipo UNIX y también permite validar usuarios haciendo de Controlador Principal de Dominio (PDC).

SAN: del inglés Storage Area Network: es una red dedicada al almacenamiento que está conectada a las redes de comunicación de una compañía.

SAS: del inglés Serial Attached SCSI: es un sistema de interconexión serie entre dispositivos SCSI.

SCSI: del inglés Small Computers System Interface: es una interfaz estándar para la transferencia de datos entre distintos dispositivos del bus de la computadora.

SMB/CIFS: Protocolo de red (que pertenece a la capa de aplicación en el modelo OSI) que permite compartir recursos entre computadoras

SNMP: del inglés Simple Network Management Protocol: es un protocolo de la capa de aplicación que facilita el intercambio de información de administración entre dispositivos de red. Permite a los administradores supervisar el funcionamiento de la red, buscar y resolver sus problemas, y planear su crecimiento.

SSH: del inglés Secure Shell: es el nombre de un protocolo y del programa que lo implementa, y sirve para acceder a máquinas remotas a través de una red.

Windows: el término Windows describe colectivamente todas o parte de varias generaciones de productos de sistema operativo de Microsoft

VSS: del inglés Volume Shadow Copy Service: permite a los administradores realizar copias instantáneas de volúmenes de datos críticos, sin producirse interrupciones en el servicio ni incoherencias en los archivos respaldados.

Hardware: conjunto de los componentes que integran la parte material de una computadora.

Software: *equipamiento lógico o soporte lógico* de un sistema informático.

CD: del inglés Compact Disc: *es un soporte digital óptico utilizado para almacenar cualquier tipo de información (audio, imágenes, vídeo, documentos y otros datos).*

Kernel: es un software que constituye la parte más importante del sistema operativo. Es el principal responsable de facilitar a los distintos programas acceso seguro al hardware de la computadora o en forma básica, es el encargado de gestionar recursos, a través de servicios de llamada al sistema.

RAM: del inglés Random Access Memory: es allí donde se cargan todas las instrucciones que ejecutan el procesador y otras unidades de cómputo. Se denominan "de acceso aleatorio" porque se puede leer o escribir en una posición de memoria con un tiempo de espera igual para cualquier posición, no siendo necesario seguir un orden para acceder a la información de la manera más rápida posible.

VMware: empresa que proporciona la mayor parte del software de virtualización disponible para ordenadores compatibles X86.

Snapshot: es una copia instantánea del estado de un sistema en un momento determinado.

Deduplicación: se refiere a la deduplicación de datos. Esta técnica de respaldo elimina los datos redundantes almacenados, guardando una única copia idéntica de los datos, y reemplazando las copias redundantes por indicadores que apuntan a esa única copia.

Demonio: un demonio, *daemon* o *dæmon* (de sus siglas en inglés *Disk And Execution MONitor*), es un tipo especial de proceso informático no interactivo, es decir, que se ejecuta en segundo plano en vez de ser controlado directamente por el usuario.

Fibre Channel: canal de fibra. Es una tecnología de red utilizada principalmente para redes de almacenamiento, disponible primero a la velocidad de 1 Gbps y posteriormente a 2, 4 y 8 Gbps.

PKI: del inglés Public Key Infrastructure: es una combinación de hardware y software, políticas y procedimientos de seguridad que permiten la ejecución con garantías de operaciones criptográficas como el cifrado, la firma digital o el no repudio de transacciones electrónicas.

TLS: del inglés Transport Layer Security: es un protocolo criptográfico que proporcionan comunicaciones seguras por una red, comúnmente Internet.

RESUMEN.

En la actualidad se hace necesario contar con un plan de almacenamiento y salva que se ajuste a las necesidades de la empresa, y que agilicen el proceso de recuperación ante una catástrofe. *Softwares* para la realización de tareas como las de salvos y replicación hacen posible que la recuperación parezca una tarea cotidiana y el tiempo en ejecutarla es ínfimo comparado el que se tendría que emplear para restablecer el sistema desde cero, sin tener en cuenta los costos que representaría para la empresa la pérdida de dicha información. No se puede dejar pasar por alto el importante papel que juegan los *softwares* que brindan las facilidades para el almacenamiento de la información, aunque estos no cuentan con una funcionalidad de recuperación ante un desastre proporcionan un entorno de administración que hace posible la asignación de los permisos de acceso a los recursos de almacenamiento creados. Muchas veces contando con las herramientas para llevar a cabo un proceso de almacenamiento y salva se pueden cometer errores producto a una mala planificación o mala distribución de los recursos. Para eliminar estos inconvenientes el presente trabajo tiene como objetivo crear y documentar un procedimiento que brinde una guía de las etapas, roles, actividades y herramientas que se deben tener en cuenta para realizar un adecuado almacenamiento y salva de la información.

Palabras Clave: procedimiento, Centro de Procesamiento de Datos, almacenamiento, salva, Bacula, Openfiler.

ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA.....	5
1.1 Procedimiento	5
1.2 Arreglo redundante de discos independientes	5
1.2.1 Ventajas del RAID	5
1.2.2 Niveles de RAID	6
1.3 Centro de Procesamiento de Datos.....	7
1.4 Salva o respaldo de datos en un CPD.....	9
1.4.1 Aspectos a tener en cuenta para realizar un respaldo	9
1.4.2 Métodos de almacenamiento	10
1.4.3 Norma ISO 17799.....	11
1.4.4 Clasificación de respaldos	12
1.5 Herramientas para gestión de almacenamiento compartido	14
1.5.1 Openfiler.....	14
1.5.2 FreeNAS.....	15
1.6 Análisis de herramientas de salva en CPD.....	16
1.6.1 Bacula	19
1.6.2 Areca Backup	21
1.6.3 Backup PC.....	22
1.7 Herramientas de salvadas en infraestructura de virtualización.	24
1.7.1 Herramientas libres.....	25
1.7.1.1 VZDump	25
1.7.2 Herramientas privativas.	25
1.7.2.1 VMware Data Recovery	25
1.7.2.2 Veeam Backup & Replication	26
1.8 Conclusiones.....	27
CAPÍTULO 2: DESCRIPCIÓN DEL PROCEDIMIENTO.....	28
2.1 Alcance	28
2.2 Roles	28

2.3	Descripción general del procedimiento	28
2.4	Posibles distribuciones de los recursos	37
2.5	Conclusiones.....	40
CAPÍTULO 3: EVALUACIÓN DEL PROCEDIMIENTO		41
3.1	Actividades por etapas	41
3.1.1	Etapa Planificación	41
3.1.2	Etapa Despliegue	42
3.1.3	Etapa Gestión.....	47
3.1.3.1	Ejecución del trabajo SALVA_MV_LUN.....	47
3.1.3.2	Restauración de archivos a través del comando “restore”	48
3.1.3.3	Ejecución de un trabajo de salva a un servidor de Base de Datos en caliente.....	51
3.2	Conclusiones.....	53
CONCLUSIONES GENERALES.....		54
RECOMENDACIONES.....		55
REFERENCIAS BIBLIOGRAFICA		56
BIBLIOGRAFIA.....		58

Índice de Tablas

Tabla 1.1 Tipos de respaldo. Características.	13
Tabla 1.2 Ejemplo de secuencia de respaldo GFS.....	13
Tabla 1.3 Herramientas de salva.....	19
Tabla 2.1 Requisitos de la herramienta Openfiler.....	30
Tabla 3.1 Resultados del análisis técnico.....	41
Tabla 3.2 Característica del demonio Bacula Client en la MV1.	43
Tabla 3.3 Característica del demonio Bacula Storage en la MV1.	44
Tabla 3.4 Característica del demonio Bacula Client en la MV2.	44
Tabla 3.5 Característica del demonio Bacula Director en la MV1.	46

Índice de Figuras

Fig. 2.1 Etapas que componen el procedimiento.....	29
Fig. 2.2 Despliegue con almacenamiento externo.....	37
Fig. 2.3 Despliegue con almacenamiento interno, portable.....	39
Fig. 2.4 Despliegue con almacenamiento interno, estático.....	39
Fig. 3.1 Despliegue de los recursos.....	42
Fig. 3.2 Ejecución manual trabajo SALVA_MV_LUN.....	47
Fig. 3.3 Resultado de la ejecución del trabajo SALVA_MV_LUN.....	48
Fig. 3.5 Ejecución del trabajo PRUEBA_MV2.....	49
Fig. 3.6 Empaquetamiento de los archivos de restauración.....	50
Fig. 3.7 Archivos recuperados.....	50
Fig. 3.8 Resultado de la ejecución del trabajo SALVA_BD.....	52
Fig. 3.9 Bases de datos salvadas mediante el script.....	53

INTRODUCCIÓN

En la actualidad es primordial para las empresas asegurar la continuidad de sus negocios, almacenando toda la información clave en un ambiente que proporcione las condiciones tecnológicas para su seguridad, recuperabilidad y para operar sin interrupciones, ni pérdidas de tiempo.

Por tal motivo, las empresas están cada día más interesadas en los Centros de Procesamiento de Datos (CPD). El tener la información y las aplicaciones más esenciales siempre disponibles se convierte entonces en la principal función de los Centros de Datos, por lo tanto, ya no son sólo un centro de acopio donde se acumulan cientos de megas de datos, sino que su misión es tener dicha información en constante funcionamiento.

Es preciso señalar, que la información es el activo más importante con que cuentan las compañías, ya que sin ella, éstas no pueden funcionar. Asegurar la información que permanece en los centros de datos va más allá de protegerla ante posibles amenazas externas. Se trata más bien, de garantizar niveles de servicio que permitan obtener información cuando se necesite, es decir, que esté disponible las 24 horas del día durante los 7 días de la semana.

Hoy, un aspecto relevante para certificar la disponibilidad de los datos, es contar con planes de continuidad correctos, como por ejemplo, los de recuperación de desastres, ya que pueden facilitar resolver dificultades para que las organizaciones puedan seguir funcionando ante cualquier amenaza, manteniendo así su continuidad y evitando cuantiosas pérdidas.

A esto se suma, el factor humano que es esencial para el resguardo de la información, es decir, contar con personal capacitado y con procedimientos establecidos basados en las normas internacionales, como por ejemplo, la ISO 17799 y la ISO 27001, que son una guía empresarial para mantener de forma segura la información de los Centros de Datos.

Por lo tanto, el reto está en adoptar medidas que abarquen y proporcionen ventajas a toda la empresa, es decir, aspectos organizacionales, legales, funcionales y comerciales. Sumándosele la utilización de las herramientas adecuadas que otorgan alta disponibilidad de las aplicaciones en los Centros de Datos y que garantizan la continuidad con soluciones de almacenamiento, respaldo y recuperación de sistemas.

En la Universidad de las Ciencias Informáticas (UCI) teniendo en cuenta el auge que ha tenido los centros de datos y la importancia que representa para una empresa tener su información almacenada en un lugar

seguro, surge la necesidad de contar un grupo de especialista en el tema de los centros de datos. Por tal motivo es creado el Grupo de Centro de Datos actualmente perteneciente al Centro de Telemática de la Facultad 2 cuyo objetivo es brindar servicios de consultoría, diseño de Centros de Datos, gestión de suministros, montaje de equipamiento, instalación de sistemas, configuración de servicios y soporte técnico.

Actualmente el grupo de trabajo a la hora del despliegue de un CPD no cuenta con un procedimiento que les brinde una idea de las etapas por las cuales debe transitar para garantizar de forma adecuada un almacenamiento y salva de la información.

Teniendo en cuenta lo anteriormente planteado surge como **problema científico** la siguiente interrogante: ¿Cómo garantizar un adecuado proceso de almacenamiento y salvas en los Centros de Procesamiento de Datos desplegados por el Grupo de Centro de Datos?

El **Objeto de Estudio** de esta investigación está dirigido a los procesos de almacenamiento y salvas de la información.

El **Campo de Acción** queda enmarcado en los procedimientos y técnicas utilizadas para el almacenamiento y salvas en los centros de procesamiento de datos.

Se persigue como **Objetivo General** elaborar un procedimiento que le brinde al Grupo Centro de Datos una guía para llevar a cabo los procesos de almacenamiento y salva de la información en centros de procesamiento de datos.

Para cumplimentar el objetivo general se definen las siguientes **Tareas de Investigación**:

- ✓ Descripción de los tipos de arreglos de discos.
- ✓ Descripción de los principales aspectos de un CPD.
- ✓ Descripción de los principales dispositivos de almacenamiento.
- ✓ Descripción de los tipos de respaldos que existen.
- ✓ Descripción de las herramientas utilizadas a nivel mundial para el almacenamiento y salvas en un CPD.
- ✓ Descripción de los roles y actividades propuestas para lograr el almacenamiento y salva de la información.
- ✓ Definición de los componentes propuestos para el almacenamiento y salvas de la información.

- ✓ Demostración de la validez del procedimiento propuesto.

Para apoyar el desarrollo de la investigación se emplean los siguientes métodos científicos:

➤ **Métodos Teóricos:**

Análítico sintético: Este método permitirá analizar las teorías y los documentos referentes al objetivo de la investigación, facilitando de esta forma la extracción de los elementos más importantes relacionados con el objeto de estudio. También posibilitará construir el camino a seguir, a partir del análisis detallado de cada uno de los documentos previamente mencionados.

➤ **Métodos Empíricos:**

Entrevista: Se realizarán varias consultas a los desarrolladores del proyecto Centro de Datos de la UCI, para obtener el conocimiento necesario sobre el estado actual del proceso llevado a cabo para el almacenamiento y salva de la información.

El presente trabajo de diploma consta de 3 capítulos donde se describe todo el proceso de investigación.

Capítulo 1 FUNDAMENTACIÓN TEÓRICA: Aborda los conceptos que son necesario para la comprensión de este trabajo, principales métodos, técnicas y herramientas utilizadas a nivel mundial para el almacenamiento y salva en CPD.

Capítulo 2 DESCRIPCIÓN DEL PROCEDIMIENTO: Describe cada uno de los roles, actividades, artefactos y herramientas que forman parte del procedimiento propuesto para lograr una contribuir auna adecuada distribución de los recursos.

Capítulo 3 EVALUACIÓN DEL PROCEDIMIENTO: Se enfoca en los aspectos relacionados con la demostración de la validez del procedimiento presentado en el presente documento.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA

Se profundizan los conceptos fundamentales para el lograr dar una propuesta acertada a las necesidades de un CPD a la hora de realizar tareas como almacenamiento y salva. Se realiza un estudio sobre los métodos, normas o estándares, técnicas y herramientas utilizadas para lograr un adecuado almacenamiento y salva de la información contenida en los CPD.

1.1 Procedimiento

Sucesión cronológica de operaciones concatenadas entre sí, que se constituyen en una unidad de función para la realización de una actividad o tarea específica dentro de un ámbito predeterminado de aplicación. Todo procedimiento involucra actividades y tareas del personal, determinación de tiempos de métodos de trabajo y de control para lograr el cabal, oportuno y eficiente desarrollo de las operaciones. (1)

1.2 Arreglo redundante de discos independientes

La tecnología RAID -*Arreglo Redundante de Discos Independientes / Redundant Array of Independent Disks*- asegura la integridad de los datos ante una eventual avería de uno de los discos, asegurando un funcionamiento continuo y permitiendo incluso la sustitución de la unidad defectuosa sin necesidad de detener los procesos que se estén ejecutando. (2)

1.2.1 Ventajas del RAID

La implementación de un RAID en el entorno empresarial proporciona a la empresa ventajas notables que hacen valer el costo de la inversión, estas son (2):

Tolerancia a fallos: Un RAID protege contra la pérdida de datos y proporciona recuperación de datos en tiempo real con acceso no interrumpido en caso de que falle un disco.

Mejora del Rendimiento / Velocidad: Un RAID consta de dos o más discos duros que ante el sistema principal funcionan como un único dispositivo. Los datos se desglosan en fragmentos que se escriben en varias unidades de forma simultánea. Este proceso, denominado fraccionamiento de datos, incrementa notablemente la capacidad de almacenamiento y ofrece mejoras significativas de rendimiento. RAID permite a varias unidades trabajar en paralelo, lo que aumenta el rendimiento del sistema.

Mayor Fiabilidad: Las soluciones RAID emplean dos técnicas para aumentar la fiabilidad: la redundancia de datos y la información de paridad. La redundancia implica el almacenamiento de los mismos datos en

más de una unidad. Aunque este planteamiento es muy eficaz, también es muy costoso, ya que exige el uso de conjuntos de unidades duplicados. El segundo planteamiento para la protección de los datos consiste en el uso de la paridad de datos. EL RAID utiliza un algoritmo matemático para generar información de paridad.

Alta Disponibilidad: El RAID aumenta el tiempo de funcionamiento y la disponibilidad del sistema. Para evitar los tiempos de inactividad, debe ser posible acceder a los datos en cualquier momento. La disponibilidad de los datos depende de dos aspectos: la integridad de los datos y tolerancia a fallos.

1.2.2 Niveles de RAID

La elección de los diferentes niveles de RAID va a depender de las necesidades del usuario en lo que respecta a factores como seguridad, velocidad, capacidad, coste, etc. Cada nivel de RAID ofrece una combinación específica de tolerancia a fallos (redundancia), rendimiento y coste, diseñadas para satisfacer las diferentes necesidades de almacenamiento. No hay un nivel de RAID mejor que otro; cada uno es apropiado para determinadas aplicaciones y entornos informáticos. (2)

RAID 0: “La más alta transferencia, pero sin tolerancia a fallos”.

También conocido como "fraccionamiento". Reparte los datos en pequeños segmentos que se distribuyen entre varias unidades, como no existe redundancia no ofrece tolerancia a fallo, por tanto un fallo de cualquier disco del arreglo tendría como resultado la pérdida de los datos y sería necesario restaurarlos desde una copia de seguridad. Permiten una transferencia simultánea de datos a todos los discos, con lo que se mejora la velocidad en las operaciones de lectura y escritura la cual aumenta en relación con el número de discos que forman el conjunto. Representa una gran ventaja en operaciones secuenciales con ficheros de gran tamaño.

RAID 1: “Redundancia. Más rápido que un disco y más seguro”.

Se basa en la utilización de discos adicionales sobre los que se realiza una copia síncrona de los datos que se están modificando. El RAID 1 ofrece una excelente disponibilidad de los datos con redundancia total de los mismos. Para ello, se duplican todos los datos de una unidad en otra. De esta manera se asegura la integridad de los datos y la tolerancia al fallo, pues en caso de avería, la controladora sigue trabajando con los discos no dañados sin detener el sistema. El RAID 1 es una alternativa costosa para los grandes sistemas, ya que las unidades se deben añadir en pares para aumentar la capacidad de almacenamiento. Sin embargo, el RAID 1 es una buena solución para las aplicaciones que requieren

redundancia cuando hay sólo dos unidades disponibles. Se necesita un mínimo de dos unidades para implementar una solución RAID 1. (2)

RAID 0+1, RAID 0/1 ó RAID 10. "Ambos mundos".

El nivel de RAID 0+1 fracciona los datos para mejorar el rendimiento, pero también utiliza un conjunto de discos duplicados para conseguir redundancia de datos. Las unidades se deben añadir en pares cuando se aumenta la capacidad, lo que multiplica por dos los costes de almacenamiento. Una configuración RAID 0+1 utiliza un número par de discos (4, 6, 8) creando dos RAID 0. Cada RAID 0 es una copia exacta del otro, de ahí la estructura en RAID 1. El RAID 0+1 es una excelente solución para cualquier uso que requiera gran rendimiento y tolerancia a fallos, pero no una gran capacidad, por un tema de costes. (2)

RAID 3: "Acceso síncrono con un disco dedicado a paridad".

Dedica un único disco al almacenamiento de información de paridad. Las operaciones de lectura y escritura acceden a todos los discos al mismo tiempo, por lo cual el RAID 3 es mejor para sistemas de un sólo usuario con aplicaciones que utilicen ficheros de gran tamaño. El RAID 3 ofrece altas tasas de transferencia, alta fiabilidad y alta disponibilidad, a un coste intrínsecamente inferior que un RAID 1. Sin embargo, su rendimiento en escritura es menor o igual al de un disco debido a que la información de paridad ha de escribirse siempre en el mismo disco. Se necesita un mínimo de tres unidades para implementar una solución RAID 3. (2)

RAID 5: "Acceso independiente con paridad distribuida".

Ofrece tolerancia al fallo, pero además, optimiza la capacidad del sistema mediante el uso de paridad distribuida. Esto lo consigue mediante el cálculo de información de paridad y su almacenamiento alternativo por bloques en todos los discos del conjunto. La información del usuario se graba por bloques y de forma alternativa en todos ellos. Se necesita un mínimo de tres unidades para implementar una solución RAID 5. (2)

El RAID 5 es la solución más económica por megabyte, que ofrece la mejor relación de precio, rendimiento y disponibilidad para la mayoría de los servidores. (2)

1.3 Centro de Procesamiento de Datos.

Se denomina CPD a aquella ubicación donde se concentran todos los recursos necesarios para el procesamiento de la información de una organización. También se conoce como centro de

cómputo (Iberoamérica) o centro de cálculo (España) o centro de datos por su equivalente en inglés data center. Dichos recursos consisten esencialmente en unas dependencias debidamente acondicionadas, computadoras y redes de comunicaciones. Suelen ser creados y mantenidos por grandes organizaciones con objeto de tener acceso a la información necesaria para sus operaciones. Prácticamente todas las compañías que son medianas o grandes tienen algún tipo de CPD, mientras que las más grandes llegan a tener varios. (3)

Entre los factores más importantes que motivan la creación de un CPD se puede destacar el garantizar la continuidad del servicio a clientes, empleados, ciudadanos, proveedores y empresas colaboradoras, pues en estos ámbitos es muy importante la protección física de los equipos informáticos o de comunicaciones implicadas, así como servidores de bases de datos que puedan contener información crítica.

Si no hay CPD, no hay información. Si no hay información, no hay conocimiento. Sin conocimiento, no se existe. Por ello, las organizaciones son cada vez más conscientes de la importancia de tener un CPD que garantice un confort y una seguridad a sus activos más valiosos: la información. (4)

La evolución actual de los CPD se basa en tres pilares fundamentales (5):

➤ Consolidación.

La consolidación de las infraestructuras de la red de datos de cara al usuario y de la red de almacenamiento de respaldo logra mayor eficacia administrativa y aumenta el uso, lo que incrementa la amortización de los activos y reduce el coste total de propiedad.

➤ Virtualización

Aumenta la productividad y agilidad de la empresa, al desligar el entorno de aplicaciones de las limitaciones de un hardware concreto. De esta forma, los recursos informáticos, de red y de almacenamiento, se pueden asignar a las aplicaciones de la forma que más se ajuste a las necesidades de la organización.

➤ Automatización

Permite gestionar el CPD como un sistema adaptable ya que facilita el aprovisionamiento de recursos, al mismo tiempo que agiliza la solución de los problemas y facilita la recuperación posterior a las amenazas de la seguridad.

1.4 Salva o respaldo de datos en un CPD.

Las interrupciones se presentan de formas muy variadas: virus informáticos, fallos de electricidad, errores de hardware y software, caídas de red, hackers, errores humanos, incendios, inundaciones, etc. Aunque no se pueda prevenir cada una de estas interrupciones, la empresa sí puede prepararse para evitar las consecuencias que éstas puedan tener sobre su negocio. (6)

Salvar la información significa copiar el contenido lógico de nuestro sistema informático a un medio que cumpla con una serie de exigencias (6):

Ser confiable: Minimizar las probabilidades de error. Muchos medios magnéticos como las cintas de respaldo, los disquetes, o discos duros tienen probabilidades de error o son particularmente sensibles a campos magnéticos, elementos todos que atentan contra la información que se ha respaldado allí.

Estar fuera de línea, en un lugar seguro: Tan pronto se realiza el respaldo de información, el soporte que almacena este respaldo debe ser desconectado de la computadora y almacenado en un lugar seguro tanto desde el punto de vista de sus requerimientos técnicos como humedad, temperatura, campos magnéticos, como de su seguridad física y lógica. No es de gran utilidad respaldar la información y dejar el respaldo conectado a la computadora donde potencialmente puede haber un ataque de cualquier índole que lo afecte.

La forma de recuperación sea rápida y eficiente: Es necesario probar la confiabilidad del sistema de respaldo no sólo para respaldar sino que también para recuperar. Hay sistemas de respaldo que aparentemente no tienen ninguna falla al generar el respaldo de la información pero que fallan completamente al recuperar estos datos al sistema informático. Esto depende de la efectividad y calidad del sistema que realiza el respaldo y la recuperación.

1.4.1 Aspectos a tener en cuenta para realizar un respaldo

Para realizar un proceso de respaldo se hace necesario llevar a cabo un análisis de la información para determinar (7):

- El tipo de resguardo, dependiendo del objetivo
- Su ubicación y distribución, es decir en qué equipos y rutas de disco se encuentra
- Su masa total, que es la suma de los tamaños de cada conjunto de datos

- La frecuencia de actualización de la información, o el lapso de tiempo en el que ésta es alterada notablemente.

1.4.2 Métodos de almacenamiento

Los métodos varían para cada caso, siendo los más comúnmente utilizados (8):

Backup en cinta

Con la instalación de un dispositivo de cinta se puede implementar una solución de resguardo periódico muy funcional, con escasos requerimientos de administración. Es el método más utilizado para pequeños servidores con juegos de datos centralizados.

Discos externos

De la misma forma que con las cintas, el proceso de resguardo se programa para ser realizado en forma desatendida en unidades de disco extraíbles.

Este método es muy útil en pequeños grupos de trabajo con datos descentralizados, ya que el disco puede trasladarse de una computadora a otra para recolectar la información durante el proceso de resguardo.

Instantánea en CD - Disco Compacto / Compact Disc -

Un conjunto de datos predeterminado se graba en CD en el lugar, para ser archivado. Se utiliza en redes pequeñas para conjuntos de datos descentralizados.

Resguardo rotativo

Método económico pero poco fiable utilizado en pequeños grupos de trabajo, donde se resguarda periódicamente la información del disco de un puesto en un directorio del disco de otro.

Extracción

Un conjunto de datos se extrae a una unidad portable y se traslada a un equipo seguro, fuera del lugar. Sucesivos conjuntos se van almacenando y, cada determinado lapso de tiempo se optimiza y organizan, se graban a CD o DVD *-Disco Versátil Digital / Digital Versatile Disc-* y se entregan para su almacenamiento. Este método se utiliza para generar históricos y otros tipos de resguardo.

1.4.3 Norma ISO 17799

Estándar internacional de alto nivel para la administración de la seguridad de la información, fue publicado por la ISO -*Organización Internacional de Normalización / International Organization for Standardization*- en diciembre de 2000 con el objeto de desarrollar un marco de seguridad sobre el cual trabajen las organizaciones.

El ISO 17799, al definirse como una guía en la implementación del sistema de administración de la seguridad de la información, se orienta a preservar los siguientes principios de la seguridad informática:

Confidencialidad: Asegurar que únicamente personal autorizado tenga acceso a la información.

Integridad: Garantizar que la información no será alterada, eliminada o destruida por entidades no autorizadas.

Disponibilidad: Asegurar que los usuarios autorizados tendrán acceso a la información cuando la requieran.

Este estándar dedica un epígrafe a lo referente con los respaldos, persiguiendo como objetivo mantener la integridad y disponibilidad de la información y los medios de procesamiento de información. Establece que se deben tener en cuenta los siguientes elementos a la lograr realizar un respaldo eficiente (9):

- a) Se debiera definir el nivel necesario de respaldo de la información.
- b) Se debieran producir registros exactos y completos de las copias de respaldo y procedimientos documentados de la restauración.
- c) La extensión (por ejemplo, respaldo completo o diferencial) y la frecuencia de los respaldos debiera reflejar los requerimientos comerciales de la organización, los requerimientos de seguridad de la información involucrada, y el grado crítico de la información para la operación continua de la organización.
- d) Las copias de respaldo se debieran almacenar en un lugar apartado, a la distancia suficiente como para escapar de cualquier daño por un desastre en el local principal.
- e) A la información de respaldo se le debiera dar el nivel de protección física y ambiental apropiado consistente con los estándares aplicados en el local principal; los controles aplicados a los medios en el local principal se debiera extender para cubrir la ubicación de la copia de respaldo.

- f) Los medios de respaldo se debieran probar regularmente para asegurar que se puedan confiar en ellos para usarlos cuando sea necesaria en caso de emergencia.
- g) Los procedimientos de restauración se debieran chequear y probar regularmente para asegurar que sean efectivos y que pueden ser completados dentro del tiempo asignado en los procedimientos operacionales para la recuperación.
- h) En situaciones cuando la confidencialidad es de importancia, las copias de respaldo debieran ser protegidas por medios de una codificación.

1.4.4 Clasificación de respaldos

Copias de Información

Los respaldos son sólo duplicados de archivos que se guardan en dispositivos de almacenamiento de alta capacidad. Los archivos que son respaldados pueden variar desde archivos del sistema operativo, bases de datos, hasta archivos de un usuario común. Existen varios tipos de softwares que automatizan la ejecución de estos respaldos, pero el funcionamiento básico de estos paquetes depende del denominado bit de archivo, el cual indica si el archivo ha sido modificado. El bit de archivo es activado en los archivos (o bloques) cada vez que sean modificados y es mediante este bit que se llevan a cabo los tres tipos de respaldos comúnmente utilizados (10):

- *Respaldo Completo ("Full")*: Guarda todos los archivos que sean especificados al tiempo de ejecutarse el respaldo. El bit de archivo es eliminado de todos los archivos (o bloques), indicando que todos los archivos ya han sido respaldados.
- *Respaldo de Incremento ("Incremental")*: Cuando se lleva a cabo un Respaldo de Incremento, sólo aquellos archivos que tengan el bit de archivo serán respaldados; estos archivos (o bloques) son los que han sido modificados después de un Respaldo Completo. Además cada Respaldo de Incremento que se lleve a cabo también eliminará el bit de archivo de estos archivos (o bloques) respaldados.
- *Respaldo Diferencial ("Differential")*: Este respaldo es muy similar al "Respaldo de Incremento", la diferencia estriba en que el archive bit permanece intacto.

<i>Respaldo</i>	<i>Archivos respaldo en</i>	<i>Bit de archivo</i>	<i>Ventajas</i>	<i>Desventajas</i>
<i>Completo ("Full")</i>	Todos	Eliminado en todos los archivos	Con este respaldo únicamente es	Tiempo de Ejecución

			posible recuperar toda la información	
<i>De Incremento ("Incremental")</i>	Archivos con el bit de archivo activo. (Aquellos que hayan cambiado desde el último Respaldo Completo)	Eliminado en los archivos que se respaldan	Velocidad	Requiere del último Respaldo Completo y de todos los Respaldos de Incremento que le siguieron para recuperar el Sistema
<i>Diferencial ("Differential")</i>	Archivos con el bit de archivo activo. (Aquellos que hayan cambiado desde el último Respaldo Completo)	Intacto	Sólo requiere del último Respaldo Completo y del último respaldo Diferencial	Ocupa mayor espacio en discos comparado con Respaldos de Incremento

Tabla 1.1 Tipos de respaldo. Características.

Secuencia de Respaldo GFS (Abuelo-Padre-Hijo / Grandfather-Father-Son)

Esta secuencia de respaldo es comúnmente utilizadas y consiste en respaldos completos cada semana y respaldos de incremento o diferenciales cada día de la semana. Suponiendo la siguiente semana:

<i>Domingo (1)</i>	<i>Lunes (2)</i>	<i>Martes (3)</i>	<i>Miércoles (4)</i>	<i>Jueves (5)</i>	<i>Viernes (6)</i>	<i>Sábado (7)</i>
Diferencial/De Incremento o NADA	Diferencial/De Incremento	Diferencial/De Incremento	Diferencial/De Incremento	Diferencial/De Incremento	Completo	Diferencial/De Incremento o NADA
<i>Domingo (8)</i>	<i>Lunes (9)</i>	<i>Martes (10)</i>	<i>Miércoles (11)</i>	<i>Jueves (12)</i>	<i>Viernes (13)</i>	<i>Sábado (14)</i>
Diferencial/De Incremento o NADA	Diferencial/De Incremento	Diferencial/De Incremento	Diferencial/De Incremento	Diferencial/De Incremento	Completo	Diferencial/De Incremento o NADA

Tabla 1.2 Ejemplo de secuencia de respaldo GFS

En caso de fallar el sistema el jueves 12 será necesario el respaldo completo del Viernes 6 y si se utilizaron respaldos diferenciales, sólo el respaldo diferencial del miércoles 11. Si se utilizaron respaldos de incremento: se necesitarán todos los respaldos de incremento desde el sábado 7 hasta el miércoles 11.

1.5 Herramientas para gestión de almacenamiento compartido

En los CPD se hace necesario contar con servidores dedicados al almacenamiento cuya función principal es brindar un espacio físico a los servidores de virtualización para que sea utilizado como contenedor de las máquinas virtuales. Para una mejor administración en estos servidores de almacenamiento se utilizan herramientas que permiten la gestión de permisos y asignación de tamaño físico de almacenamiento a usuarios específicos. A continuación se describe las principales características de dos herramientas libres de gran uso a nivel mundial que sirven como Sistema Operativo y como herramienta de gestión de almacenamiento compartido en servidores con este propósito.

1.5.1 Openfiler.

Es una distribución Linux orientada a servidores de almacenamiento creada por Xinit Systems, distribuida bajo licencia GPL. Su principal ventaja es la de dotar a los administradores de sistemas de una administración vía web que simplifica y facilita la creación de recursos, el manejo de usuarios y la asignación de permisos, lo que se traduce en una baja curva de aprendizaje. (11)

Openfiler convierte tu red en un centro de almacenamiento, organiza los datos de tu empresa, añade más espacio a cualquier máquina en la red o simplemente se utiliza para cualquier función que requiera más espacio. Ofrece una interfaz de administración bien amigable que puede gestionar toda la configuración, incluyendo servicios como LDAP *-Protocolo Ligero de Acceso a Directorios / Lightweight Directory Access Protocol-*, SAMBA, NFS *-Sistema de Archivos de Red / Network File System-*, RSYNC, iSCSI, FTP *-Protocolo de Transferencia de Archivos / File Transfer Protocol-*. Usa el kernel de Linux 2.6 el cual le permite soportar los últimos microprocesadores, tarjetas de red y hardware de almacenamiento. Sistema de archivos con capacidad de hasta 60TB, sistemas de archivos en línea e incremento de volumen. (11)

Costo Efectivo

A través de su interfaz web, el administrador puede realizar diversas tareas administrativas tales como la creación de volúmenes, las cuotas de la red, la asignación de cuota de disco para usuarios y grupos y la gestión de RAID (11).

Protocolos soportados

Openfiler proporciona capacidades de almacenamiento tanto a nivel de bloque y a nivel de archivo. En el nivel de bloque, el almacenamiento puede ser exportado a través de (11):

- iSCSI Target
- Fibre Channel

En el nivel de archivo, Openfiler soporta:

- NFS
- SMB/CIFS (Common Internet File System)
- FTP
- rsync

El almacenamiento puede ser asignado a través de cualquiera de los protocolos soportados y al mismo tiempo permite que los datos se accedan desde una base heterogénea de clientes en la red, ya sean Linux, Windows o Mac. También cuenta con el servicio iSCSI initiator el cual le permite conectar a un servicio iSCSI Target de otro servidor de almacenamiento. (11)

1.5.2 FreeNAS

FreeNAS básicamente es un sistema operativo ligero, gratuito y de código abierto basado en la estructura del sistema operativo FreeBSD, admite el uso compartido a través de Windows, Apple, y los sistemas de tipo UNIX. FreeNAS 8 incluye ZFS - *Sistema de Archivo Zettabyte / Zettabyte File System*-, que soporta altas capacidades de almacenamiento e integra los sistemas de archivos y la gestión de volumen en una sola pieza de *software*. (12)

El sistema de archivos zettabyte es un sistema de archivos avanzada de código abierto desarrollada originalmente para Solaris. FreeNAS utiliza actualmente la versión de ZFS 15 como su sistema de archivos recomendado para dispositivos con 6 GB de RAM - *Memoria de Acceso Aleatorio / Random Access Memory* - o más. (12)

ZFS admite la adición de las cuotas de los conjuntos de datos, con el fin de evitar que un usuario pueda usar una capacidad mayor que la asignada para la tarea que desempeña y también puede asignar más

capacidad en caso de ser necesario, esto se puede hacer para utilizar plenamente la capacidad disponible cuando muchos usuarios no utilizan la totalidad de su cuota de la capacidad del sistema. (12)

En cualquier momento, una copia instantánea (snapshot) del sistema de archivos ZFS puede ser tomada, preservando todos los archivos del sistema en ese momento contra la eliminación hasta que la instantánea expira o se elimina. Esto puede hacerse manualmente o programada regularmente. Estas instantáneas actúan como puntos de roll-back en caso del borrado accidental de datos críticos. (12)

Estas copias instantáneas también ofrecen un modo de copia de seguridad de un sistema de ficheros ZFS que es posible utilizarlos para realizar copias de seguridad incremental de un sistema de archivos a otro. El uso compartido de archivos es lo que mejor sabe hacer FreeNAS. Es compatible con SMB / CIFS (intercambio de archivos de Windows), NFS (intercambio de archivos en Unix) y AFP (intercambio de archivos de Apple), así como FTP, iSCSI, y otros métodos de compartir datos a través de la red. FreeNAS 8 está administrado totalmente a través de una interfaz de navegador web después de la instalación inicial. Todos los servicios que FreeNAS ofrece y todas sus opciones se pueden configurar mediante esta interfaz gráfica. (12)

Servicios soportados por FreeNAS (12):

- LDAP
- SMB/CIFS (Common Internet File System)
- SSH
- NFS
- RSYNC
- iSCSI Target

1.6 Análisis de herramientas de salva en CPD

Actualmente existen numerosas herramientas para la salva automática de la información y cada una cuenta con variadas características que las hacen diferenciarse entre ellas pero también tienen puntos en común. A continuación se llevará a cabo el análisis de tres herramientas utilizadas para la realización de salvadas en CPD, teniendo en cuenta los siguientes aspectos:

Software Libre

El *software* libre es una cuestión de la libertad de los usuarios de ejecutar, copiar, distribuir, estudiar, cambiar y mejorar el *software*. Más precisamente, significa que los usuarios de programas tienen las cuatro libertades esenciales. (13)

- La libertad de ejecutar el programa, para cualquier propósito (libertad 0).
- La libertad de estudiar cómo trabaja el programa, y cambiarlo para que haga lo que usted quiera (libertad 1). El acceso al código fuente es una condición necesaria para ello.
- La libertad de redistribuir copias para que pueda ayudar al prójimo (libertad 2).
- La libertad de distribuir copias de sus versiones modificadas a terceros (libertad 3). Si lo hace, puede dar a toda la comunidad una oportunidad de beneficiarse de sus cambios. El acceso al código fuente es una condición necesaria para ello.

Un programa es *software* libre si los usuarios tienen todas esas libertades. Entonces, debería ser libre de redistribuir copias, tanto con o sin modificaciones, ya sea gratis o cobrando una tarifa por distribución, a cualquiera en cualquier parte. (13)

También debería tener la libertad de hacer modificaciones y usarlas en privado, en su propio trabajo u obra, sin siquiera mencionar que existen. Si publica sus cambios, no debería estar obligado a notificarlo a alguien en particular, o de alguna forma en particular. (13)

Encriptación de Datos

En un Sistema de Comunicación de Datos a través de una red no segura como lo es internet, es de vital importancia asegurar que la información no viaje en texto claro, manteniendo su autenticidad, integridad, confidencialidad y el no repudio de la misma entre otros aspectos.

Compresión de Datos

Este aspecto cobra gran importancia en casos donde el volumen de información a salvar es grande y mantener copias periódicas trae consigo una gran inversión en el almacenamiento, con la compresión de datos se aprovecha al máximo el espacio con que cuenta la infraestructura de almacenamiento de nuestro CPD. Como desventaja esto hace un poco más lento el proceso de salva aunque es un tiempo que será económicamente bien recompensado.

Tipos de Respaldo

Como ya se describió existen tres tipos de respaldo: completo, incremental y diferencial. Todos tienen sus ventajas y desventajas pero combinados entre ellos se puede elaborar un plan de salva eficiente que responda a los intereses del CPD, por eso es necesario que esta característica esté presente en la herramienta seleccionada.

Planificador de Tareas

Para la automatización del sistema de respaldo esta característica ocupa un notable lugar, realizando una planificación de los momentos en que se realizarán la salva no será necesario la ejecución de cada tarea de forma manual, acotando el trabajo simplemente la configuración de las tareas, dejándole el resto a la herramienta de salva, resaltando que el personal tendrá como prioridad garantizar que el respaldo se haya realizado correctamente.

Respaldo de permisos de archivos

Para mantener la protección sobre los archivos del sistema, luego de una restauración, es necesario respaldar el archivo junto con su información de control de acceso. Respaldo permisos en ambientes GNU/Linux, y listas de control de acceso en ambientes Windows. Si no se dispone de este mecanismo el administrador del sistema deberá otorgar los permisos nuevamente, luego de la restauración del archivo.

Respaldo de archivos abiertos

En sistemas donde un servicio se brinda las 24 horas al día, los 7 días a la semana, y existe una gran concurrencia de usuario es posible que a la hora de realizar el respaldo algunos archivos permanezcan abiertos, muchas herramientas cuentan con la potencial funcionalidad de respaldar los archivos y mantener su coherencia.

Facilidad y eficiencia en la restauración de archivos

La facilidad para restauración de archivos viene en gran medida con la posibilidad de contar con una interfaz que provea facilidades para buscar y restaurar los archivos deseados.

Las herramientas seleccionadas son:

Herramienta	Licencia
Bacula	GPL

Areca Backup	BSD
BackupPC	GPL

Tabla 1.3 Herramientas de salva.

1.6.1 Bacula

Análisis de la Herramienta

Bacula es una colección de 5 demonios que cooperan entre sí para llevar a cabo un trabajo de salva, aunque pueden estar desplegados de forma independiente se necesita llevar a cabo una correcta y detallada configuración para lograr su comunicación. Para agrupar los volúmenes donde se almacenará la información salvada y poder aplicar configuraciones al grupo, Bacula utiliza un recurso llamado pool (piscina).

A continuación se describen los demonios que conforman la colección de Bacula:

Demonio Director: Es el encargado de llevar a cabo los trabajos de salva de forma automática, siguiendo la planificación que le fue programada, indicándole al demonio cliente qué archivos se van salvar y dónde se almacenaran los mismos, limitándose solamente a dar las órdenes, pues el tráfico de los archivos se realiza desde el demonio cliente hacia el demonio de almacenamiento. Los trabajos pueden ser ejecutados de forma opcional, manualmente a través de la consola, no así para el proceso de restauración el cual no acepta de una automatización, siendo algo lógico analizando que no se puede predecir el momento en que hará falta la utilización de este tipo de trabajo.

Demonio de Almacenamiento (Storage Daemon): Es el encargado del manejo de los dispositivos físicos donde se almacenan los datos recibidos de los demonios clientes, junto con los datos, son recibido también los atributos de los archivos salvados. Es el encargado de hacer llegar al director los atributos para que sean enviados al demonio catálogo.

Demonio Cliente (Client Daemon): Se encuentra instalado en el cliente que contiene los archivos a salvar, es el encargado del empaquetamiento de los archivos indicados por el director y enviarlos una vez salvados, al almacenamiento establecido por el propio director.

Catálogo (Catalog): Es el encargado de manejar la base de datos que contiene toda la información referente a los trabajos programados, todos los volúmenes y las pool creadas. Quedan registrados todos

los procesos de salva y la información de todos los archivos que lo componen, con su fecha, tamaño, lugar donde se restauran, lugar donde se almacenan físicamente; en la base datos no quedan guardados los archivos en cuestión, solo sus atributos, por eso en caso de ser borrada, los respaldos permanecerían intactos y se podría realizar la restauración usando otra herramienta que provee Bacula, bextract.

Consola (Console): Es el medio que permite la interacción entre el administrador del sistema y el módulo del director, de esta manera puede acceder a todas las funcionalidades de Bacula.

Resumen de las características principales de Bacula (14)

- Permite respaldar sobre varios volúmenes.
- Arquitectura Cliente-Servidor.
- Compuesto por 5 componentes: Director (Servidor), Consola, Cliente (File Daemon), Almacenamiento (Storage Daemon) y Catálogo.
- Compresión de datos: GZIP.
- Encriptación: MD5, SHA1.
- Soporte para MySQL, postgresQL y SQLite.
- Incluye servicio de catálogo que le permite al administrador localizar y restaurar cualquier archivo así como ver los respaldos realizados hasta el momento.
- Requiere realizar configuraciones en sus componentes.
- Soporte limitado para Windows (solo cliente).
- Tipos de respaldo: Completo, Incremental y Diferencial.
- Restauración de respaldos mediante el propio Bacula. La ventaja es que incluye características como la de restaurar lista de control de acceso ACL (siglas en inglés) para cada archivo.
- Respalda sobre cinta, disco, USB, DVD.
- Interfaz gráfica Gnome, y varias interfaces web disponibles.
- Programador de tareas con ejecución simultánea.
- Muy rápida restauración de archivos individuales.

- Encriptación PKI -*Infraestructura de Clave Pública / Public Key Infrastructure*- para datos y comunicación sobre TLS -*Seguridad de la Capa de Transporte / Transport Layer Security*.
- Soporte para archivos mayores a 2 GB y arquitecturas de 64 bits.
- Testeo de unidades de cinta.
- Soporte para VSS, que permite realizar respaldos de archivos abiertos en Windows.
- Soporte para respaldo de bases de datos en línea (En Windows utiliza VSS y en Linux usa scripts previstos por la comunidad Bacula).

1.6.2 Areca Backup

Análisis de la Herramienta

Areca Backup es un *software* de copia de seguridad de archivos desarrollado en Java. Se incluye un mecanismo de operación que garantiza la integridad de una copia de seguridad.

El *software* permite a los usuarios seleccionar los archivos o directorios de copia de seguridad, filtrar, encriptar y comprimir su contenido, y almacenarlos en una ubicación de copia de seguridad.

Para llevar a cabo la creación y ejecución de respaldos en Areca, se definen grupos de respaldo configurables. Los grupos de respaldo consisten de un conjunto de destinos, en los cuales se configurarán los siguientes puntos:

- Origen de los datos (archivos y directorios a respaldar).
- Destino de los datos (carpeta en disco, unidad de red, FTP, etc...).
- Filtros
- Modo de Almacenamiento
- Compresión
- Encriptación

Pueden haber varios destinos definidos para un mismo grupo lo cual permite categorizar y visualizar de manera esquemática todos los respaldos creados.

Resumen de las características principales de Areca Backup (15)

- Desarrollado en Java.
- Licencia pública GNU 2.
- Posee interfaz gráfica e interfaz en línea de comandos.
- Compresión de datos: ZIP/ZIP64.
- Encriptación de datos: AES 128 y de AES 256.
- Filtros de Búsqueda (por extensión, directorio, expresión regular, tamaño, fecha, etc.).
- Tipo de respaldo: incremental, completo y diferencial.
- Modo de almacenamiento estándar, imagen y delta (solamente las partes modificadas de los archivos serán almacenadas).
- Ejecución simultánea de respaldos.
- Respaldos sobre discos, unidades de red, USB, FTP (con SSL/TLS).
- Unión de archivos contiguos para disminuir el espacio de almacenamiento.
- Mecanismo de transacción, el cual genera procesos de respaldo transaccionales, asegurando la integridad de los datos.
- Respaldo y recuperación de permisos, links simbólicos y ACL de archivos (solo Linux).
- Versionado de archivos y explorador de historial: permite buscar y restaurar más de una versión existente del archivo.
- No dispone de planificador de tareas pero permite usar tanto el de Windows como el de Linux.
- Envío de informes a cuentas de e-mail.
- Generación de reportes de respaldo.
- Registro de actividades de usuarios (Log).
- No respalda archivos abiertos.

1.6.3 Backup PC

Análisis de la Herramienta

Backup PC es una herramienta para respaldo de información disco a disco, con una arquitectura que solo requiere la presencia de un servidor con el *software* instalado. El protocolo SMB estándar se utiliza para extraer los datos a respaldar desde los clientes de Windows mientras que en los clientes Linux y MacOSX, se utiliza RSYNC o FTP. (16)

Resumen de las características principales de BackupPC (16)

- Desarrollado en Perl.
- Código Abierto.
- Inteligente estrategia de manejo del agrupamiento de archivos (*pooling*) minimiza el acceso y espacio de almacenamiento en disco. Idénticos archivos de varios respaldos de la misma o diferente PC sólo se almacenan una vez.
- Deduplicación de archivos permite almacenar solamente una copia de archivos duplicados eliminando toda posible redundancia.
- Soporte de compresión opcional para reducir aún más el almacenamiento en disco. Dado que sólo los archivos nuevos necesitan ser comprimidos.
- No se necesita instalar el *software* en el cliente.
- Extracción de datos a respaldar mediante tar, rsync.
- Una potente interfaz de usuario web que permite a los administradores ver los archivos de registro, configuración, estado actual y permite a los usuarios iniciar y cancelar las copias de seguridad y navegar y recuperar archivos de respaldos.
- Soporta entornos móviles, como computadoras portátiles conectados a la red.
- Soporte para redes DHCP.
- Parámetros de configuración flexibles permiten varias copias de seguridad a realizarse en paralelo, directorios donde hacer el respaldo, programación de respaldos completos o incrementales y envío de notificaciones *vía e-mail*.
- Parámetros de configuración se pueden establecer en todo el sistema, o también en función de cada PC.
- Debe ser instalado en servidores Linux, Solaris o Unix.

- Probado en clientes Linux, Freenix, Solaris, Win95, Win98, Win2000 y WinXP.
- No respalda ACL de archivos Windows.
- No respalda archivos abiertos.
- Máximo en el tamaño del respaldo.
- Compresión de datos: Permite ahorrar entre el 30 y 40 % de espacio pero requiere instalar zlib y zip.
- Encriptación RSA.
- Tipos de respaldo: Completo, Incremental y Diferencial.

1.7 Herramientas de salvallas en infraestructura de virtualización.

Como se mencionaba anteriormente uno de los pilares de los CPD es la virtualización, esto es producto a las múltiples ventajas que provee tener una infraestructura de virtualización dentro de las que se encuentran: reduce el costo total en materia de hardware y espacio físico, provee una mayor facilidad para la creación de entornos de pruebas que permiten poner en marcha nuevas aplicaciones sin impactar a la producción, un fallo general en una máquina virtualizada no afecta al resto de las máquinas virtuales. Como estas, existen muchas más ventajas que han hecho que las grandes empresas del mundo inviertan en su utilización en sus CPD.

Actualmente existen muchas soluciones para la creación de un entorno de virtualización y una de las empresas que ha escalado hasta la cima es VMware, tanto con sus herramientas para la virtualización como las que complementan la misma, realizando tareas de salva y replicación. A pesar de sus múltiples ventajas estas presentan una limitante para aquellas empresas que no cuentan con el capital necesario para pagar las licencias y poder explotar al máximo sus funcionalidades, pues aunque en muchas ocasiones existen versiones gratis, sus mayores fortalezas se encuentran en las de pago. Empresas como Symantec han desarrollado soluciones de virtualización que proporcionan capacidades de infraestructura líderes del sector, que son compatibles con VMware. Es válido resaltar que el campo de la virtualización no ha sido ocupado solamente por empresas propietarias, pues soluciones como Xen de la empresa Citrix, VirtualBox de Oracle, openVZ de Parallels y Proxmox de Proxmox Server Solutions, todas bajo el principio de *software* libre, han sabido ganarse la confianza en el mundo de la virtualización de grandes empresas, ejemplo de ellas Google y Amazon utilizando Xen como herramienta de virtualización.

A pesar que el objetivo que persigue este trabajo es la elaboración de un procedimiento para el almacenamiento y salva bajo el principio del *software* libre se tendrán en cuenta también herramientas privativas utilizadas en el mundo para la realización de salvas en CPD en infraestructura virtualizada.

1.7.1 Herramientas libres.

1.7.1.1 VZDump

Proxmox Virtual Environment es una plataforma de virtualización código abierto fácil de usar para el funcionamiento de dispositivos y máquinas virtuales, desarrollado y mantenido por Proxmox Server Solutions GmbH (17).

La copia de seguridad o *backup* es uno de los rasgos más importante integrados a Proxmox VE a través de la utilidad VZDump la cual es capaz de realizar *snapshots* o copias instantáneas de máquinas virtuales en ejecución que incluyen los archivos de configuración. VZDump almacena la copia de seguridad en el disco en un único archivo, el cual puede ser copiado directamente a un almacenamiento remoto. (18)

Proxmox VE provee una interfaz web para definir las tareas de respaldo. El primer paso es definir una la ubicación donde se realizará el almacenamiento del respaldo, pudiendo utilizar un recurso compartido de NFS en una NAS/SAN, accediendo a través del menú 'VZDump Backups'. Después de configurado la ubicación del almacenamiento se pueden planificar los trabajos de respaldos a realizar a través de la misma interfaz web. (18)

1.7.2 Herramientas privativas.

A continuación se describen los principales rasgos de dos herramientas utilizadas para la realización de salva, en el caso de la segunda se complementa la solución con la funcionalidad de llevar a cabo la replicación de datos, estas herramientas forman parte de las soluciones dadas por VMware para satisfacer la necesidad de los clientes de realizar las tareas antes mencionadas.

1.7.2.1 VMware Data Recovery

Es una solución de respaldo y recuperación fiable y fácil de desplegar que las organizaciones pueden utilizar al virtualizar su infraestructura a fin de proporcionar una primera línea de protección de los datos en su entorno virtual.

VMware Data Recovery permite realizar lo siguiente: (19)

- Backup de la imagen completa de las máquinas virtuales.
- Recuperación completa e incremental de máquinas virtuales, así como de archivos y directorios individuales.

Proceso de respaldo:

Se realiza un copia instantánea de las maquinas que se quieran salvar garantizando la captura del estado completo en el momento en que se crea la instantánea, durante el proceso de almacenamiento, VMware Data Recovery deduplica el flujo de bloques de datos con el fin de eliminar datos redundantes antes de escribir los datos de respaldo en el disco de destino. (19)

Proceso de recuperación:

Durante el proceso de recuperación de una máquina virtual completa, VMware Data Recovery presenta los puntos de restauración de cada máquina virtual. Una vez seleccionado el punto de restauración deseado, VMware Data Recovery inicia el proceso de restauración de la máquina virtual en el *host* o en el *pool* de recursos de destino recuperando los bloques de datos específicos. Al sobrescribir una máquina o disco virtual existente, VMware Data Recovery sólo transfiere los bloques de datos que han cambiado. VMware Data Recovery también permite la restauración de archivos. Este proceso de recuperación es similar, con la diferencia de que se inicia en la consola de la máquina virtual. (19)

1.7.2.2 Veeam Backup & Replication

Veeam es una empresa proveedora de software de gestión, respaldo y replicación de máquinas virtuales para VMware. Veeam Backup & Replication ofrece respaldo y recuperación rápida, flexible y fiable de las aplicaciones virtualizadas y los datos. Unifica el respaldo y la replicación en una única solución. Con esta solución de respaldo VMware, no hay agentes que administrar, ni atender a los trabajos de respaldo, y puede verificar la recuperabilidad de cada respaldo, en todo momento. (20)

Administración centralizada

Desde una consola Web puede administrar las licencias de Veeam a través de múltiples servidores, recibir estadísticas sobre los tamaños de los respaldos y editar/clonar los trabajos desde una ubicación centralizada. (20)

Compresión y deduplicación integradas

Incluido sin cargo alguno, ahora puede administrar grandes volúmenes de datos generados por respaldos basados en imágenes con la posibilidad añadida de comprimir o deduplicar los datos consiguiendo un rendimiento mejorado al reducir la cantidad de datos que se envía por la red. (20)

Sin software agente

No hay que licenciar o desplegar, administrar, monitorizar o actualizar agentes en los hosts o en las máquinas virtuales. (20)

1.8 Conclusiones

- Existen empresas como VMware, IBM, Symantec, etc, que cuentan con su propio procedimiento para realizar el despliegue de centros de datos, con el inconveniente que no se encuentran publicados en el internet por motivos comerciales, pues son utilizados únicamente por su personal cuando se les contrata, por tal motivo no aparece reflejado en el estado del arte de la investigación realizada la descripción de otros procedimientos. Aunque se refleja un fragmento de la norma ISO 17799, encargada de establecer las normas o pasos a tener en cuenta para llevar a cabo el proceso de seguridad y continuidad de la información en una institución. .
- Como herramienta de almacenamiento se utilizará Openfiler teniendo en cuenta que la administración del servidor se realiza a través de una interfaz web usando SSL (https) para su comunicación y brinda la posibilidad, a través del servicio iSCSI initiator, de adicionar a un servidor principal el volumen de almacenamiento compartido por otros servidores a través del servicio iSCSI Target.
- De las herramientas libres de salvadas anteriormente descritas, Bacula permite el respaldo de bases de datos en uso, programación de tareas de respaldo y restauración, búsqueda de archivos, interfaz gráfica donde realizar configuraciones y visualizar el estado de las tareas de respaldo. Las dos herramientas restantes Areca y BackupPC tienen las siguientes desventajas: no posee planificador para ejecutar respaldos, no respalda los permisos de los archivos en sistemas operativos Windows, no puede respaldar archivos abiertos y no cuenta con soporte para base de datos en línea. Las ventajas que posee Bacula sobre sus dos oponentes hacen que sea la herramienta que cumpla con las exigencias que se tuvieron en cuenta durante el análisis realizado, siendo así la seleccionada para llevar a cabo el proceso de salva en los CPD.

CAPÍTULO 2: DESCRIPCIÓN DEL PROCEDIMIENTO

El presente capítulo presenta y describe un procedimiento para llevar a cabo los procesos de almacenamiento y salva en centros de datos teniendo en cuenta las posibles condiciones tecnológicas que se puedan presentar. Se llevan a cabo una serie de actividades que conforman un ciclo de vida para cada proceso y se complementan con la utilización de las herramientas anteriormente seleccionadas.

2.1 Alcance

Garantiza los procesos de almacenamiento y salva en los CPD con la utilización de las herramientas seleccionadas, en él se describe el personal encargado de la ejecución de las tareas que serán llevadas a cabo en las diferentes etapas por las cuales se encuentra compuesto. Es capaz de adaptarse a las condiciones tecnológicas con que cuente el centro de datos donde se decida utilizar, siempre y cuando se cumpla con los requerimientos mínimos que se establecen a lo largo de su descripción, brindándole las soluciones que se ajusten con su equipamiento.

2.2 Roles

Durante el ciclo de vida del procedimiento se necesita de un personal especializado cuya función sea llevar a cabo cada una de las actividades que tributan a la obtención de adecuados proceso de almacenamiento y salva. Teniendo en cuenta lo antes planteado se han definido como roles que intervienen el *especialista de almacenamiento* y el *especialista de copia de seguridad*, responsables del proceso de almacenamiento y salva respectivamente.

Especialista de almacenamiento: es la persona encargada de realizar el levantamiento de los recursos necesarios para llevar a cabo el proceso de almacenamiento, es su responsabilidad mantener actualizado el documento *asignación de almacenamiento* y junto al especialista de copia de seguridad debe elaborar el documento que recoge la asignación de los medios: *distribución de recursos*.

Especialista de copia de seguridad: es la persona encargada de la realización de las actividades que son llevadas a cabo en el proceso de salva y tiene la responsabilidad de mantener actualizada la *bitácora de configuración de bacula*.

2.3 Descripción general del procedimiento

Se encuentra dividido en tres etapas a través de las cuales transitan los procesos de almacenamiento y salva.



Fig. 2.1 Etapas que componen el procedimiento.

Aunque las etapas anteriormente definidas se tendrán en cuenta tanto en el proceso de almacenamiento como en el de salva, se hace necesario para una mejor organización, el análisis del ciclo de vida de cada proceso de forma independiente, excluyendo la etapa planificación, la cual será descrita de forma común para ambos procesos, pues se hace necesario la participación de los especialistas de almacenamiento y copia de seguridad de forma conjunta en las actividades que la componen.

➤ *Etapa Planificación*

La actual etapa es aplicable para un centro de datos que desee adquirir nuevos recursos tecnológicos para establecer una nueva infraestructura de almacenamiento y salva o para uno que desee utilizar los recursos con los que cuenta. Tiene como objetivo hacer un levantamiento del equipamiento necesario para llevar a cabo el proceso de almacenamiento y salva que requiere el centro de datos y conformar la propuesta que se ajuste a sus necesidades.

Roles que intervienen

Especialista de almacenamiento y especialista de copia de seguridad.

Artefactos que se generan

Levantamiento Técnico: documento donde queda reflejado un resumen del hardware disponible para el almacenamiento.

Distribución de recursos: documento donde queda descrito la distribución de los recursos tecnológicos que fueron asignados para el proceso de almacenamiento y salva.

Actividades

1. Realizar análisis técnico: se lleva a cabo un levantamiento de los recursos tecnológicos disponibles en el centro de datos para ser utilizado en el proceso de almacenamiento y salva, quedando reflejado en el documento *levantamiento técnico*. En caso que el centro de datos tenga como necesidad conocer los recursos mínimos de los cuales necesita para crear una nueva infraestructura, sería responsabilidad de los especialistas que ejecutan la actividad realizar una propuesta inicial teniendo en cuenta los requisitos que se describen a continuación.

Teniendo en cuenta las recomendaciones y requisitos que brinda el sitio oficial de la herramienta de almacenamiento y los conocimientos adquiridos a lo largo de la investigación realizada, es necesario tener en cuenta los siguientes aspectos:

- Según el sitio oficial de Openfiler en su sección requerimientos del sistema¹, para lograr un mejor rendimiento con el uso de su herramienta de almacenamiento Openfiler, es recomendado el uso de procesadores de arquitectura de 64 bit de la familia *Intel Xeon* o *AMD Opteron*. Aunque establecen una tabla donde se muestra los requisitos mínimos y los recomendados, los cuales son:

Requisitos mínimos	Requisitos recomendados
Procesador 32 bits a 1 GHz o superior	Procesador 64 bits a 1.6 GHz o superior
2 GB o más de RAM	2 GB o más de RAM
2 GB de espacio en disco para el área de memoria de intercambio	2 GB de espacio en disco para el área de memoria de intercambio
8 GB de espacio en disco para instalar el sistema operativo Openfiler	8 GB de espacio en disco para instalar el sistema operativo Openfiler
Interfaz de red Ethernet de 100 MB	Interfaz de red Ethernet de 1 GB
	Controlador de hardware RAID

Tabla 2.1 Requisitos de la herramienta Openfiler.

¹ <http://openfiler.com/products/system-requirements>

CAPÍTULO 2: DESCRIPCIÓN DEL PROCEDIMIENTO

- En caso de existir una infraestructura de virtualización se hace necesario la utilización de un servidor donde se encontrará instalada la herramienta Openfiler, cuya función principal será brindar a la infraestructura un lugar donde almacenar las máquinas virtuales.
- ✓ Haciendo referencia a lo planteado en la norma ISO 17799 en su epígrafe 10.5 citado textualmente *“las copias de respaldo se debieran almacenar en un lugar apartado, a la distancia suficiente como para escapar de cualquier daño por un desastre en el local principal”*, se hace necesario la inclusión de un servidor externo o un medio de almacenamiento portable que pueda ser transportado fuera del centro de dato principal después de realizado el almacenamiento de la salva.
- ✓ En caso de no existir la posibilidad de contar con un medio portable para el almacenamiento de las salvas, ni un servidor que se encuentre ubicado fuera del centro de datos principal, se propone incluir un servidor interno, el cual brindará a los servidores de salva un volumen de almacenamiento a través del protocolo NFS, su capacidad de almacenamiento mínimo viene dada por la suma total del tamaño de la información que vaya a ser salvada, la frecuencia con que cambia y el tiempo que va a permanecer en el servidor antes de ser remplazada por versiones más actualizadas.
- ✓ Un aspecto que no debe faltar en los servidores de almacenamiento es el arreglo de disco (RAID), como se describe en el capítulo anterior existen varios niveles RAID y cada cual es dedicado a un entorno con requerimientos específicos. Quedan como propuesta a utilizar los niveles RAID 0 o RAID 1, por ser soluciones económicas que brindan rapidez en la transferencia de datos y alta disponibilidad respectivamente.
- ✓ Para la selección de un dispositivo de almacenamiento portable, se debe tener en cuenta que el volumen total de la salva que se va a realizar no supere el máximo de almacenamiento del dispositivo.
- ✓ Para el proceso de salva se hace necesario como mínimo, la utilización de un servidor encargado del manejo de los dispositivos de almacenamiento interno, en caso de existir, y del acceso al volumen donde se encuentran almacenadas las máquinas virtuales de la infraestructura de virtualización para realizar la salva de cada una, y un servidor donde se encontrarán instalados el demonio director, catálogo y consola, de la herramienta bacula. Los servidores de salva no necesitan discos duros de grandes tamaños de almacenamiento pues solo será instalado el sistema operativo seleccionado (cualquier distribución de Linux para servidores) y los demonios de la herramienta de salva, Bacula.

2. Seleccionar servidores de almacenamiento y salva: teniendo en cuenta los resultados obtenidos en la actividad anterior, que se encuentran reflejados en el artefacto *levantamiento técnico*, se lleva a cabo una asignación de los recursos tecnológicos, para lograr determinar los servidores que serán utilizados para cada proceso y sus prestaciones finales, dejándolo plasmado en el documento *distribución de recursos*. En caso de que se utilice un medio de almacenamiento portable, se incluye en la descripción del servidor de salva que contendrá el demonio de almacenamiento y es responsabilidad del especialista de copia de seguridad su configuración y mantenimiento.
3. Seleccionar nivel RAID: después de asignados los servidores destinados al almacenamiento se hace necesario la elección del nivel RAID que se va a utilizar, siempre y cuando se cuente con al menos dos discos duro, si lo que se requiere es tener mayor volumen de almacenamiento y que las tareas de lecturas y escrituras se realicen con la mayor rapidez entonces el nivel a utilizar es RAID 0, pero si se requiere tener un respaldo de la información almacenada entonces se utilizará el RAID 1.
4. Definir local de almacenamiento remoto: si se cuenta con dispositivos de almacenamiento portable se hace necesario definir el local donde se mantendrán almacenado. Según lo planteado por la norma ISO 17799 citado textualmente “*a la información de respaldo se le debiera dar el nivel de protección física y ambiental apropiado (ver cláusula 9) consistente con los estándares aplicados en el local principal; los controles aplicados a los medios en el local principal se debiera extender para cubrir la ubicación de la copia de respaldo*”, el local debe cumplir con una serie de requisitos que contribuyen a la seguridad e integridad de la información que será guardada, la selección será responsabilidad del especialista de copia de seguridad basándose en los aspectos recogidos en el documento *nivel de protección física y ambiental* que no es más que una copia de la cláusula anteriormente mencionada, creada durante el tiempo de investigación.

Proceso de almacenamiento

➤ *Etape Despliegue*

Tiene como objetivo llevar a cabo el despliegue de la herramienta que hacen posible que los servidores de almacenamiento puedan brindar los servicios para los que fueron destinados, teniendo en cuenta la distribución que se encuentra descrita en el documento *distribución de recursos* de la etapa anterior.

Roles que intervienen

Especialista de almacenamiento.

Artefactos que se generan

Asignación de almacenamiento: documento donde se registran los principales aspectos de los volúmenes de almacenamiento que se comparten en los servidores, para así tener un mayor control de la asignación de los recursos y del acceso permitido.

Actividades

1. Instalar servidores: se lleva a cabo la instalación de los servidores que fueron definidos en el documento *distribución de recursos*, para la instalación del servidor de almacenamiento Openfiler, se encuentran definidos la secuencia de pasos a seguir en el *manual de instalación* en su epígrafe 1.3.1.
2. Configurar volumen de almacenamiento: se lleva a cabo la configuración de los servidores para compartir los volúmenes de almacenamiento, quedando reflejado en el artefacto *asignación de almacenamiento*. En el caso del servidor Openfiler, se tendrán en cuenta los pasos definidos en el *manual de instalación* a lo largo de su epígrafe 1.3.2. Si se seleccionó un servidor interno para almacenar las salvadas, entonces es necesario compartir el almacenamiento al servidor de salvadas siguiendo los pasos reflejados en el *manual de instalación* en el epígrafe 1.4.3 en la sección *servidor*.

➤ *Etapas Gestión*

Tiene como principal objetivo la adición de nuevos volúmenes de almacenamiento, actualización o eliminación de los ya existentes. Se lleva a cabo el soporte necesario para que no se vea afectada la disponibilidad de los servidores. La ejecución de las actividades que se proponen, por su naturaleza, vienen condicionadas por las diferentes situaciones que se puedan presentar.

Roles que intervienen

Especialista de almacenamiento.

Artefactos que se generan

Actividades

1. Gestionar volumen de almacenamiento: tiene como objetivo adicionar, actualizar o eliminar un volumen de almacenamiento, una vez terminada la actividad se tiene como responsabilidad la actualización del documento *asignación de almacenamiento*.

2. Realizar tarea de soporte: en caso de ocurrir una rotura en algunos de los servidores de almacenamiento es responsabilidad del especialista resolver el problema en el menor tiempo posible, e informar al especialista de copia de seguridad en caso de la pérdida de la información.

Proceso de salva

➤ *Eta*pa Despliegue

Se realiza la instalación y configuración de los servidores que fueron destinados para el proceso de salva, en caso de existir un medio de almacenamiento portable será configurado en la herramienta de salva para que lo utilice.

Roles que intervienen

Especialista de almacenamiento.

Artefactos que se generan

Bitácora de configuración de bacula: tiene como objetivo registrar los principales aspectos de configuración que se realizan sobre la herramienta de salva Bacula para lograr un el entendimiento de la distribución de cada uno de los demonios que la conforman. También quedan registrados los trabajos de salvas programadas teniendo en cuenta el cliente, los archivos a salvar, el horario, el tipo de salva, el dispositivo de almacenamiento y la pool donde será almacenado los datos salvados.

Actividades

1. Instalar herramienta de salva: primeramente se lleva a cabo la instalación de la distribución de Linux que se vaya a utilizar como sistema operativo en los servidores de salva descritos en el documento *distribución de recursos*, también se realiza la instalación de los demonios de bacula que estarán ubicados en cada servidor y los paquetes necesarios para poder realizar el almacenamiento de las salvas en un medio portable, en caso de existir y de ser necesaria dicha instalación. Los pasos para la instalación de cada demonio de bacula se encuentran definidos en el *manual de instalación* epígrafe 1.4.1.
2. Configurar herramienta de salva: se establecen los principales parámetros en cada uno de los demonios para poder lograr una comunicación entre ellos y se pueda llevar a cabo el proceso de salva y restauración a través de la creación de trabajos, los pasos a seguir para la ejecución de esta actividad

se encuentran en el *manual de instalación* a lo largo del epígrafe 1.4.2. Se registran los principales parámetros de configuración en el documento *bitácora de configuración de bacula*.

3. Configurar dispositivo de almacenamiento: si fue seleccionado un servidor interno para el almacenamiento de las salvas, entonces se hace necesario, una vez configurado el volumen en el servidor de almacenamiento, para poder acceder y comenzar a almacenar las salvas, montar en el servidor de salva que contiene el demonio de almacenamiento el volumen compartido, siguiendo los pasos que se encuentran definidos en el *manual de instalación* en el epígrafe 1.4.3 en la sección *pasos en el cliente*. En caso de que se utilice un dispositivo portable se debe realizar la configuración necesaria para que Bacula pueda almacenar los datos salvados.
4. Crear trabajos: tiene como objetivo realizar la programación de los trabajos encargados de realizar de forma automática la salva de la información quedando registrado en el documento *bitácora de configuración de Bacula*, también pueden ser ejecutados de forma manual en caso que se haga necesario. Antes de realizar la programación de los trabajos se debe realizar un estudio para poder seleccionar el horario en que será ejecutado, la frecuencia con que se debe realizar, los archivos que se van a salvar, el tipo de respaldo a utilizar, donde se va a almacenar, el período de tiempo por el cual se encontrará en el servidor y la utilización o no de compresión de los datos. A continuación se describe aspectos a tener en cuenta a la hora de realizar la programación de un trabajo.

Horario para realizar las salvas: el proceso de salva se realizará en el horario que menor tráfico de red exista en el centro de datos, el especialista de copia de seguridad será el encargado de solicitar a los especialistas del centro cuya función es la gestión de la red, un informe con los horarios picos del tráfico en la red.

Uso del factor compresión: trae consigo a la hora de ejecutar un trabajo de salva que se requiera de un menor espacio para almacenarlo, pero como factor negativo aumenta el tiempo que demora en culminar el trabajo, pues para comenzar la transferencia de los datos se necesita que el proceso de compresión termine, en centros de datos donde el espacio almacenamiento no es una limitante se propone obviar la compresión.

Tipo de respaldo y frecuencia: no existe un estándar que indique cuando se debe realizar una salva ni el tipo de respaldo que se debe utilizar, pues la selección viene dada por el dinamismo que tenga la información, es decir la frecuencia con la que cambia. Para contribuir al ahorro de almacenamiento se

propone realizar salvallas de tipo total y entre cada salva total realizar salvallas incrementales, la frecuencia de realización estará condicionada a la información que se vaya a salvar.

Periodo de tiempo por el cual se encontrará en el servidor: viene dado por el tiempo que necesite la empresa mantener el historial de los datos salvados, en caso de que no sea necesario, el periodo mínimo por el cual permanecerán en el servidor será el doble del intervalo de las salvallas de tipo completa del sistema con el objetivo que se garantice que exista una copia completa del sistema una vez borrado

➤ *Etapa Gestión*

Se lleva a cabo la adición de nuevos trabajos y actualización o eliminación de los ya existentes. Se llevan a cabo los trabajos de restauración en caso de ocurrir algún incidente que afecte el servicio que se encuentra respaldado. La ejecución de las actividades que se proponen, por su naturaleza, vienen condicionadas por las diferentes situaciones que se puedan presentar.

Roles que intervienen

Especialista de copia de seguridad.

Artefactos que se generan

Actividades

1. Gestionar trabajo: tiene como objetivo adicionar, actualizar o eliminar un trabajo. Para llevar a cabo las dos primeras acciones no se deben pasar por alto los aspectos que se deben tener en cuenta, anteriormente descritos, a la hora de realizar su programación. Es responsabilidad del especialista la actualización del documento *bitácora de configuración de Bacula*.
2. Ejecutar trabajo manualmente: ejecutar de forma manual un trabajo de salva es una opción recomendada en caso de ocurrir un error durante la ejecución programada en la herramienta. A diferencia de los trabajos de salva los trabajos de restauración no pueden ser automatizado, algo lógico teniendo en cuenta que no se puede predecir cuándo será necesario usarlos, por tanto, en caso de ocurrir una pérdida de la información en un sistema, los archivos serán recuperados ejecutando de forma manual un trabajo de restauración o seleccionando solo los archivos que se necesitan recuperar. El proceso de recuperación selectiva antes mencionado es posible gracias al comando “restore” que contiene la consola de Bacula, una detallada descripción de cómo utilizarlo se encuentra en el documento *manual de instalación* en su epígrafe 1.4.6.

3. Configurar nuevo dispositivo de almacenamiento: si decide incluir un nuevo dispositivo de almacenamiento o cambiar el existente, es responsabilidad del especialista su configuración y prueba antes de ser incluido en el proceso de salva existente.

2.4 Posibles distribuciones de los recursos

Durante la descripción del procedimiento se establecen los principales recursos con los que debe contar un centro de datos para poder crear una infraestructura de almacenamiento y salva. Para una mayor comprensión se muestran tres imágenes que representan la estructura de un centro de datos genérico, donde se encuentran desplegadas las posibles soluciones, las cuales serán explicadas a continuación.

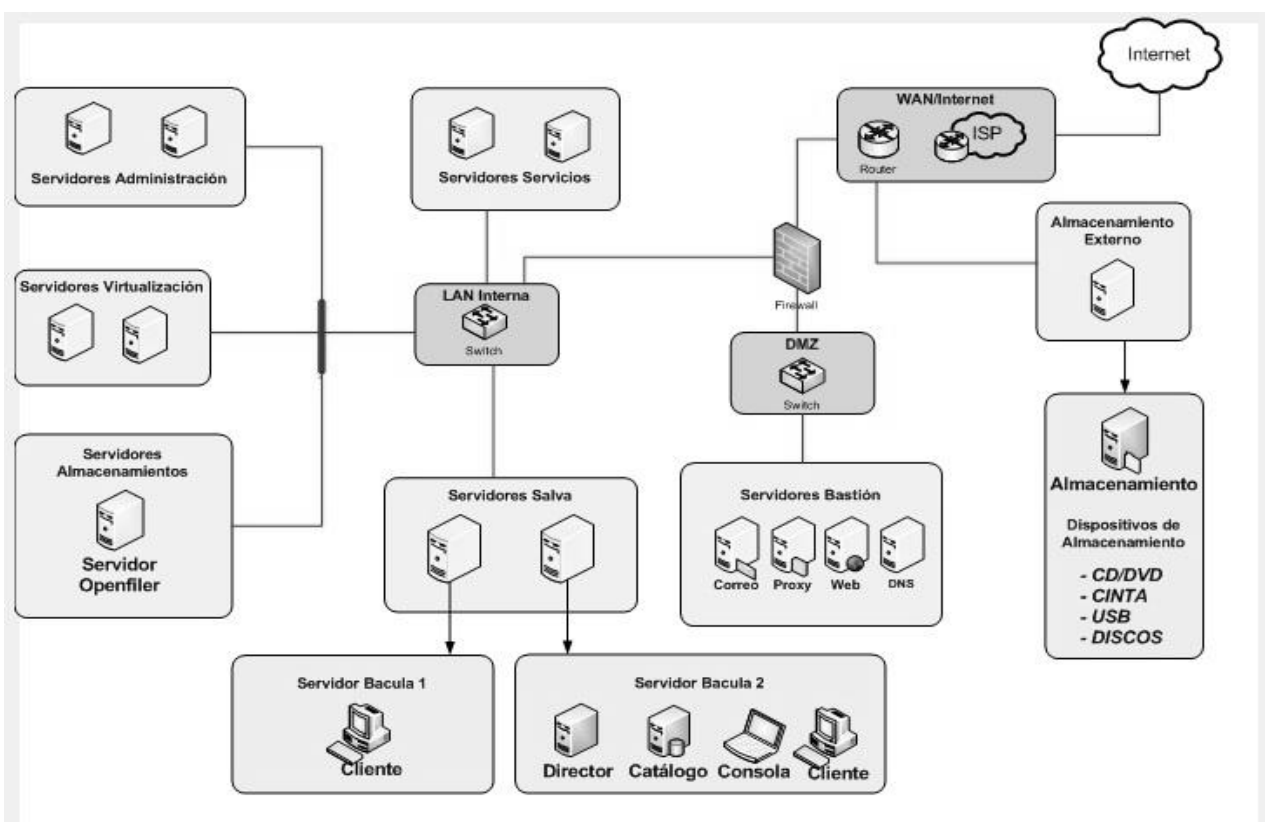


Fig. 2.2 Despliegue con almacenamiento externo.

En la figura 2.2 se muestra cómo quedarían distribuidos los recursos en caso de existir un servidor externo para el almacenamiento de las salvadas.

Servidor Openfiler: servidor cuya función principal es brindar un volumen donde almacenar las máquinas virtuales creadas por los servidores de virtualización.

Servidor Bacula 1: servidor donde quedará montado el volumen compartido por el servidor Openfiler a los servidores de virtualización, con el objetivo de poder acceder a las máquinas contenidas en él y realizar el proceso de salva, sin tener que utilizar una herramienta propia del programa de virtualización. En caso de ser una solución de VMware se necesitaría de la herramienta vmfs-tools para poder ver la información contenida en el volumen. Los pasos para llevar a cabo este proceso se encuentran definidos en el documento *manual de instalación* epígrafe 1.4.7. El demonio cliente será el encargado de empaquetar los archivos de las máquinas virtuales almacenadas en el volumen, que sean ordenados a salvar por el director.

Servidor Bacula 2: contiene el demonio director, catálogo, consola y cliente, su función principal es ser el centro de dirección del proceso de salva y se ubica un demonio cliente con el objetivo de poder respaldar la configuración de los demonios y la salva de la base de datos que contiene el catálogo.

Almacenamiento Externo: es una buena práctica que garantiza el almacenamiento de las salvas en un lugar apartado del centro de datos principal, el demonio almacenamiento es el encargado del manejo del medio físico que se vaya a utilizar para almacenamiento.

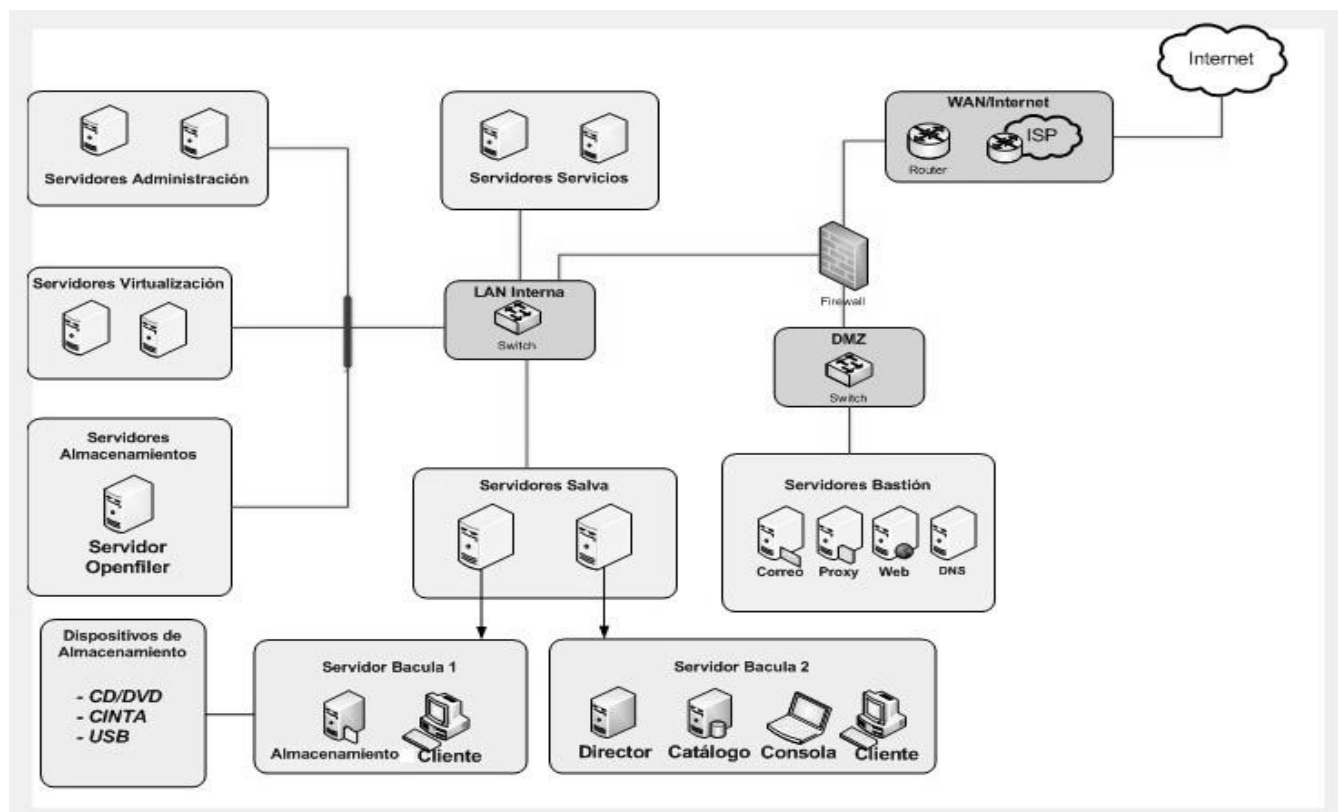


Fig. 2.3 Despliegue con almacenamiento interno, portable.

En la figura 2.3 se puede observar que se cuenta en el centro de datos con un almacenamiento que puede ser retirados una vez realizada la salva, aunque requiere de un mayor esfuerzo que el esquema con almacenamiento externo, debido a que los medios tienen que ser trasladados a un lugar seguro, es una solución que brinda la posibilidad de minimizar las probabilidades de perder la información respaldada en caso de una catástrofe en el local principal.

La descripción del despliegue es igual que el anterior solo que ahora el demonio de almacenamiento se encuentra ubicado en el *servidor de salva 1*.

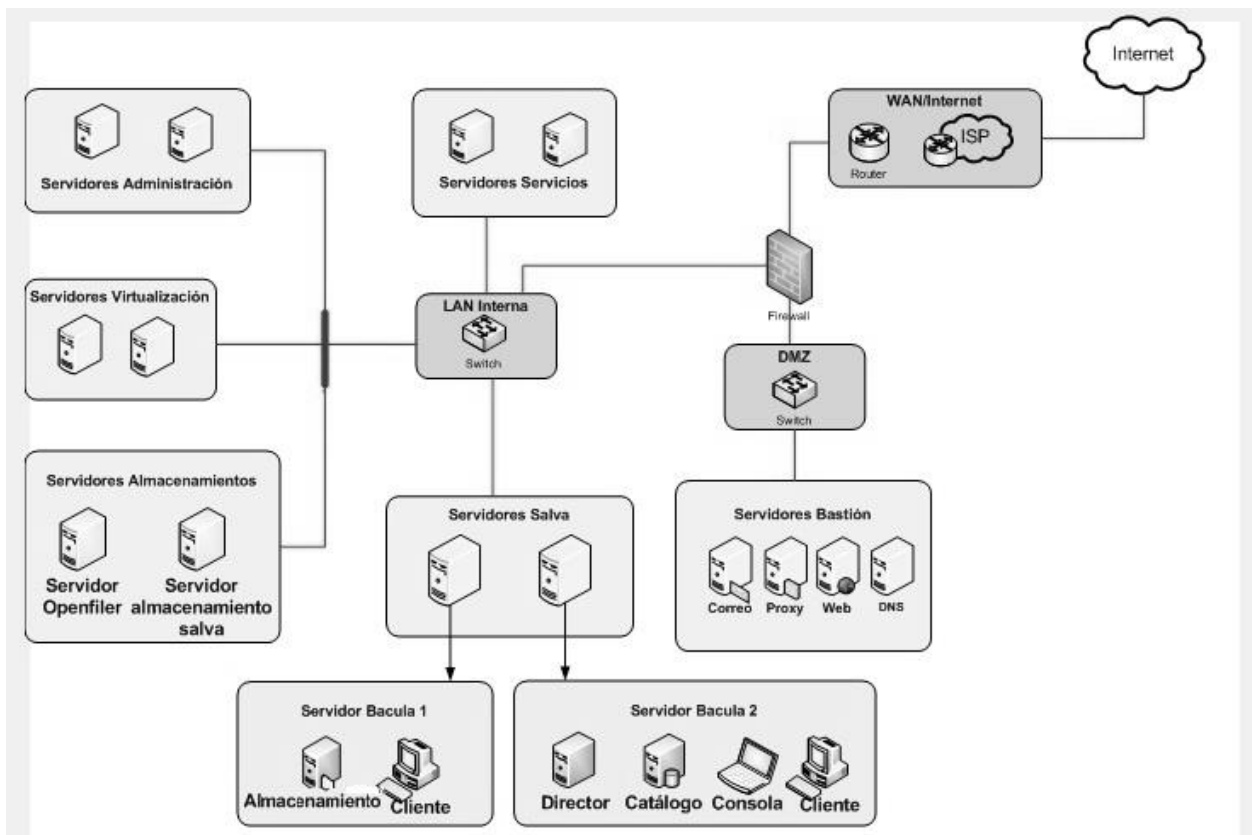


Fig. 2.4 Despliegue con almacenamiento interno, estático.

La figura 2.4 muestra la última posible distribución, el almacenamiento se realiza en un servidor interno, por lo tanto si ocurriese una catástrofe en el centro de datos pudiera perderse toda la información respaldada.

Servidor almacenamiento salva: es la única novedad, en él quedarán almacenadas las salvas realizadas, para que sea posible se comparte a través del protocolo NFS el volumen de almacenamiento que vaya a ser utilizado por el demonio de almacenamiento que se encuentra instalado en el *servidor bacula 1*.

2.5 Conclusiones

En el capítulo se describe las actividades que se deben llevar a cabo durante cada etapa para poder lograr desarrollar un proceso de almacenamiento y salva de acorde a las necesidades del centro de datos donde se vaya a utilizar. Fueron generados 4 artefactos que complementan la solución brindando un mayor control y organización de la asignación y configuración de los recursos que fueron utilizados.

CAPÍTULO 3: EVALUACIÓN DEL PROCEDIMIENTO

Los capítulos anteriores fundamentan desde el punto de vista teórico y describen a nivel técnico una propuesta de un procedimiento para el almacenamiento y salva en CPD. Sin embargo, la misma pudiera considerarse incompleta, en ausencia de los elementos que demuestren su validez. En el presente capítulo se llevarán a cabo una serie de prueba en un escenario virtual, las cuales a su vez sirven de base para demostrar el carácter válido de la misma, dando cumplimiento así a una de las tareas más importantes de esta investigación.

3.1 Actividades por etapas

A continuación se llevara a cabo el ciclo de vida del procedimiento en el centro de datos de la uci ubicado en la planta baja del docente 2, debido a que no existían el número los servidores reales se tomó la decisión por parte del grupo de trabajo de realizar las pruebas con máquinas virtuales (MV).

3.1.1 Etapa Planificación

Se llevó a cabo un estudio de los recursos disponibles en el centro de datos el cual arrojó que no se cuenta con dispositivos de almacenamiento portables ni un centro de datos externo, por lo tanto se decidió utilizar para el almacenamiento de las salvas un servidor interno.

Resultados del levantamiento:

	Máquina Virtual 1	Máquina Virtual 2	Máquina Virtual 3
Procesador	Intel® Xeon® CPU 3050 @ 2.13GHz	Intel® Xeon® CPU 3050 @ 2.13GHz	Intel® Xeon® CPU 3050 @ 2.13GHz
RAM	1 GB	512 MB	1 GB
Disco Duro	10 GB	10 GB	30

Tabla 3.1 Resultados del análisis técnico.

Teniendo las prestaciones identificadas en la actividad de análisis técnico se decidió realizar la siguiente distribución:

La máquina virtual 1, teniendo en cuenta que es el que posee mayor capacidad de disco duro fue seleccionada como servidor de almacenamiento de salva.

La máquina virtual 2, teniendo en cuenta que es el de menos prestaciones fue seleccionada como el servidor de salva que alberga el demonio director, catalogo, consola y cliente. Aunque sea el servidor donde se encuentre instalado el director no necesita de mayores prestaciones que el resto de los servidores que contienen otros demonios debido a que el tráfico mayor del proceso de salva ocurre entre los clientes y el demonio de almacenamiento. Aunque el servidor cuenta con un demonio cliente, los archivos que serán salvado por él no son de gran tamaño, no requiriendo de un gran procesamiento de información.

La máquina virtual 3, es usada como servidor de salva que alberga el demonio de almacenamiento y el cliente.

En el caso del servidor de almacenamiento donde estará instalada la herramienta Openfiler no queda reflejada su selección porque ya se encuentra instalada en un servidor real.

3.1.2 Etapa Despliegue

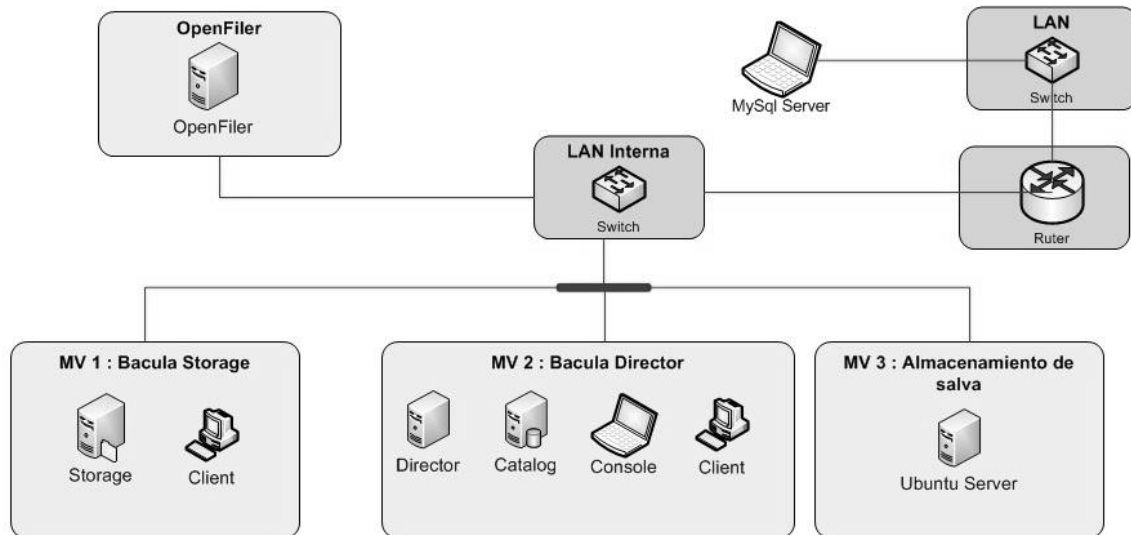


Fig. 3.1 Despliegue de los recursos.

Se instaló el sistema operativo Ubuntu Server 11.04 en cada una de las máquinas virtuales y después se realizó la instalación de cada demonio de Bacula que se encuentra desplegado en las MV que se pueden observar en la figura 3.1. También se compartió el volumen de almacenamiento de la MV 3 a través del protocolo NFS concediéndole acceso solo a la MV 1.

Como se puede apreciar en la figura 3.1 se encuentra fuera del centro de datos una laptop que tiene instalado un servidor MySQL para que se le realice una salva en ejecución. La dirección IP que se le asignó fue 10.51.18.132.

Configuración de los demonios

Una vez instalado el sistema operativo en los servidores de salvos se procedió a la configuración de los demonios. A continuación se muestra una serie de tablas con la información que contiene el documento “bitácora de configuración de Bacula”.

➤ *MV1 Bacula Storage*

Demonio Bacula Client (FileDaemon)

<i>Recurso Client</i>
Nombre: <i>mv1_storage</i>
Puerto: 9102
Dirección IP: 10.52.60.131
Contraseña: 0OZ0DMt1krC4QbEhgyuW-manHkpdGI5_v
Descripción: Encargado de empaquetar los archivos de la MV que se encuentran almacenadas en el volumen que fue montado en /media/san

Tabla 3.2 Característica del demonio Bacula Client en la MV1.

Demonio Bacula Storage

<i>Recurso Storage</i>
<p>Nombre: File</p> <p>Puerto: 9103</p> <p>Dirección IP: 10.52.60.131</p> <p>Contraseña: E__ll-9Z8jYXXl7d__KEd2feiykrl7H3j</p>
<i>Recurso Device</i>
<p>Nombre: FileStorage</p> <p>Tipo: File</p> <p>Dispositivo de almacenamiento: /mnt/salva</p>

Tabla 3.3 Característica del demonio Bacula Storage en la MV1.

➤ MV2 Bacula Director

Demonio Bacula Client (FileDaemon)

<i>Recurso Client</i>
<p>Nombre: mv2_director</p> <p>Puerto: 9102</p> <p>Dirección IP: 10.52.60.132</p> <p>Contraseña: /1M2JeVTulm/vlG7mdnuy/ai3lz92Euivuw4OaWmldNI</p> <p>Descripción: Encargado de empaquetar los archivos de configuración del servidor MV2 Bacula Director.</p>

Tabla 3.4 Característica del demonio Bacula Client en la MV2.

Demonio Bacula Director

<p><i>Recurso Director</i></p>
<p>Nombre: director</p> <p>Puerto: 9101</p> <p>Dirección IP: 10.52.60.132</p> <p>Contraseña de la consola: /1M2JeVTulm/vlG7mdnuy/ai3lz92Euivuw4OaWmldNI</p>
<p><i>Recurso FileSet</i></p>
<p>Nombre: PRUEBA</p> <p>Archivo a salvar: /media/prueba</p> <p>Archivo a excluir:</p> <p>Descripción: Contiene los archivos que serán salvados para mostrar posteriormente el funcionamiento de la restauración.</p> <p>Nombre: MV_LUN_1</p> <p>Archivo a salvar: /media/san/UbuntuServerApache-Yannier/</p> <p>Archivo a excluir: *.vmdk</p> <p>Descripción: Contiene los archivos a salvar de la MV UbuntuServerApache-Yannier que se encuentra almacenada en el servidor Openfiler, se acceden a ellos a través de la LUN montada en el directorio /media/san/ en la MV1; excluyendo los archivos con extensión vmdk que es la que le corresponde a los discos virtuales.</p> <p>Nombre: BD</p> <p>Archivo a salvar: /home/salva</p> <p>Archivo a excluir:</p> <p>Descripción: Contiene los archivos de las salvadas realizadas al servidor de base de datos.</p>

<i>Recurso Schedule</i>							
<p>Nombre: MV1</p> <p>Frecuencia: Lunes – Sábado 00:05 Tipo: Incremental Domingo 00.05 Tipo: Completo</p> <p>Nota: el horario establecido se debe a que en el horario de la noche la concurrencia de usuarios es menor teniendo en cuenta que los archivos a salvar son accedido en el horario de 8:00 am – 5:00 pm, por los estudiantes y ocasionalmente en las noches desde los laboratorio de producción. Debido a que los archivos a copiar cambian diariamente se hizo necesario realizar una salva de tipo completo los domingos y el resto de la semana de tipo incremental.</p>							
<i>Recurso: Pool</i>							
<p>Nombre: File</p> <p>Tiempo de Retención: 15 días</p> <p>Nota: como no es interés del centro de datos mantener un historial de los datos salvados, se establece 15 días como el tiempo que se almacenará en el servidor.</p>							
<i>Recurso: Job</i>							
<i>Nombre</i>	<i>Tipo</i>	<i>Cliente</i>	<i>FileSet</i>	<i>Schedule</i>	<i>Storage</i>	<i>Device</i>	<i>Pool</i>
<i>SALVA_MV_LUN</i>	<i>Backup</i>	<i>mv1_storage</i>	<i>MV_LUN_1</i>	<i>MV1</i>	<i>File</i>	<i>FileStorage</i>	<i>File</i>
<i>PRUEBA_MV2</i>	<i>Backup</i>	<i>mv2_director</i>	<i>PRUEBA</i>	<i>MV1</i>	<i>File</i>	<i>FileStorage</i>	<i>File</i>
<i>REST_PRUEBA</i>	<i>Restore</i>	<i>mv2_director</i>	<i>PRUEBA</i>		<i>File</i>	<i>FileStorage</i>	<i>File</i>
<i>SALVA_BD</i>	<i>Backup</i>	<i>mv2_director</i>	<i>BD</i>	<i>MV1</i>	<i>File</i>	<i>FileStorage</i>	<i>File</i>

Tabla 3.5 Característica del demonio Bacula Director en la MV1.

3.1.3 Etapa Gestión

Teniendo en cuenta que en la etapa de gestión las actividades definidas son ejecutadas en circunstancias dadas, se describen una serie de problemáticas y la forma en que fueron solucionadas.

3.1.3.1 Ejecución del trabajo SALVA_MV_LUN

Tiene como objetivo realizar una breve salva de una MV que se encuentra almacenada en el servidor Openfiler, excluyendo los discos virtuales que contiene dicho directorio pues el objetivo que se persigue es solo demostrar como la unión de bacula, open-iscsi y vmfs-tools hacen posible la realización de salvas al sistema de archivos definidos por VMware.

Durante la ejecución programada para el trabajo que se describe anteriormente, por fallo en el sistema se reinició el servidor donde se encuentra instalado el demonio de almacenamiento y el demonio cliente, ambos involucrados en el proceso de salva, trayendo consigo que ocurriese un error en el trabajo que se estaba ejecutando. Para poder solucionar el problema que se presentó existen dos posibles soluciones, esperar a su próxima ejecución programada o ejecutarlo de forma manual inmediatamente, teniendo en cuenta las consecuencias que puede traer la espera, se decidió optar por la segunda opción, la cual se describe a continuación:

Desde la consola de Bacula se ejecuta el comando "run" y se selecciona el trabajo SALVA_MV_LUN y se muestra a continuación un resumen del trabajo previo a su ejecución (ver figura 3.2).

```
Run Backup job
JobName: SALVA_MV_LUN
Level: Full
Client: mvl_storage
FileSet: MV_LUN_1
Pool: File (From Job resource)
Storage: File (From Job resource)
When: 2012-06-09 23:38:01
Priority: 10
OK to run? (yes/mod/no):
```

Fig. 3.2 Ejecución manual trabajo SALVA_MV_LUN.

A continuación, como todos los valores de los parámetros mostrados en el resumen son correctos se procede a la ejecución del trabajo.

```

Build OS:          i686-pc-linux-gnu ubuntu 11.04
JobId:            41
Job:              SALVA_MV_LUN.2012-06-09_23.41.04_03
Backup Level:     Full
Client:           mvl_storage 5.0.3 (04Aug10) i686-pc-linux-gnu,ubuntu,11.04
FileSet:          "MV_LUN_1" 2012-06-08 04:17:37
Pool:             "File" (From Job resource)
Catalog:          "MyCatalog" (From Client resource)
Storage:          "File" (From Job resource)
Scheduled time:   09-Jun-2012 23:38:01
Start time:       09-Jun-2012 23:41:06
End time:         09-Jun-2012 23:47:34
Elapsed time:     6 mins 28 secs
Priority:         10
FD Files Written: 14
SD Files Written: 14
FD Bytes Written: 14,412,601 (14.41 MB)
SD Bytes Written: 14,414,529 (14.41 MB)
Rate:            37.1 KB/s
Software Compression: 97.3 %
VSS:             no
Encryption:      no
Accurate:        no
Volume name(s):  vol102
Volume Session Id: 20
Volume Session Time: 1339178958
Last Volume Bytes: 461,934,056 (461.9 MB)
Non-fatal FD errors: 0
SD Errors:       0
FD termination status: OK
SD termination status: OK
Termination:     Backup OK

```

Fig. 3.3 Resultado de la ejecución del trabajo SALVA_MV_LUN.

Como resumen se puede observar que se salvaron 14.41 MB en un tiempo de 6 min 28 segundos con una compresión de 97.3 %.

3.1.3.2 Restauración de archivos a través del comando “restore”

A pesar que un trabajo de restauración ha sido creado, en ocasiones es necesario restaurar un sistema a un punto específico, puede darse el caso que la última ejecución del trabajo de salva se hiciera sobre el sistema ya defectuoso. Gracias al comando “restore” de Bacula, esta problemática puede ser resuelta en un pequeño lapso de tiempo. Para poder observar con mejor comprensión, se ejecutará tres veces el trabajo de salva *PRUEBA_MV2* pero antes de cada ejecución se creará una nueva carpeta en el directorio de salva “/media/salva” con los nombres: C1, C2 y C3 para cuando se realice la restauración poder observar que carpetas fueron recuperadas.

JobId	Name	StartTime	Type	Level	JobFiles	JobBytes	JobStatus
28	PRUEBA_MV2	2012-06-09 16:47:52	B	F	2	0	T
29	PRUEBA_MV2	2012-06-09 16:49:23	B	I	2	0	T
30	PRUEBA_MV2	2012-06-09 16:52:45	B	I	2	0	T

Fig. 3.5 Ejecución del trabajo PRUEBA_MV2

Como se puede observar en la figura 3.5 se ejecutaron 3 veces el trabajo PRUEBA_MV2, en la primera ejecución se realizó una salva de tipo completo, donde se encuentra en el directorio “/media/salva” la primera carpeta creada “c1”, en la segunda y tercera ejecución debido a que la salva se realizó de tipo incremental se le adicionó a la salva la segunda carpeta “c2” y la tercera carpeta “c3” respectivamente.

Asumiendo que el sistema presentó problemas minutos antes de que se ejecutara por tercera vez el trabajo, los archivos salvados en dicha ejecución contienen la información con errores, la solución a este problema es volver al punto anterior. En el caso de prueba la solución fue dada haciendo una restauración completa desde la salva de la ejecución anterior (JobId = 29) ejecutando en la consola de Bacula el comando “restore” y seleccionado la opción 12: “Select full restore to a specified Job date”, cuya función es la construcción del paquete de restauración con los archivos salvados desde el trabajo de salva completo que le antecede al trabajo cuyo JobId fue insertado y los trabajos incrementales que se realizaron entre el trabajo completo y el insertado. En este caso se seleccionó el trabajo con JobId = 29 y el paquete fue construido con los archivos salvados en ese trabajo y los archivos del trabajo se salva completa con JobId = 28 (ver figura 3.6).

```

First you select one or more JobIds that contain files
to be restored. You will be presented several methods
of specifying the JobIds. Then you will be allowed to
select which files from those JobIds are to be restored.

To select the JobIds, you have the following choices:
  1: List last 20 Jobs run
  2: List Jobs where a given File is saved
  3: Enter list of comma separated JobIds to select
  4: Enter SQL list command
  5: Select the most recent backup for a client
  6: Select backup for a client before a specified time
  7: Enter a list of files to restore
  8: Enter a list of files to restore before a specified time
  9: Find the JobIds of the most recent backup for a client
 10: Find the JobIds for a backup for a client before a specified time
 11: Enter a list of directories to restore for found JobIds
 12: Select full restore to a specified Job date
 13: Cancel
Select item: (1-13): 12
Enter JobId to get the state to restore: 29
Selecting jobs to build the Full state at 2012-06-09 16:49:23
You have selected the following JobIds: 28,29

Building directory tree for JobId(s) 28,29 ...
2 files inserted into the tree.

You are now entering file selection mode where you add (mark) and
remove (unmark) files to be restored. No files are initially added, unless
you used the "all" keyword on the command line.
Enter "done" to leave this mode.

cwd is: /
$

```

Fig. 3.6 Empaquetamiento de los archivos de restauración.

Con la opción “*mark*” se marcan todos los archivos del árbol creado y con “*done*” se especifica que ya culminó la selección de los archivos que forman parte del paquete de restauración, una vez terminado el proceso de restauración se muestra en la carpeta donde se almacenó la restauración los archivos que componen la misma. Como se muestra en la figura 3.7 solo fueron recuperadas las carpetas c1 y c2, dejando fuera la tercera carpeta que fue creada y que se encontraba respaldada en el trabajo que se ejecutó después de ocurrir el incidente.

```

root@bacula-director-yoelvis:/home/restaurar/media/prueba# ls
c1  c2
root@bacula-director-yoelvis:/home/restaurar/media/prueba#

```

Fig. 3.7 Archivos recuperados.

Es válido resaltar que si el proceso de restauración se realiza a través de la ejecución de un trabajo de este tipo la herramienta devolverá una restauración completa del cliente implicado en dicho trabajo, no dando la opción de elegir un punto de restauración como el comando *restore*.

3.1.3.3 Ejecución de un trabajo de salva a un servidor de Base de Datos en caliente

El escenario de prueba contaba con un servidor Mysql donde se encuentra la base de datos “mia”, la cual no podía ser detenida para realizar su respaldo, es por ello que se necesitaba de un proceso que realizara la salva de sus datos en caliente, es decir en funcionamiento. A continuación se describe la solución dada:

Para la ejecución de este trabajo se hizo necesario la creación de un *script* cuyo objetivo fue realizar la salva a todas las base de datos contenidas en el servidor Mysql, sin que este fuera detenido, este *script* pudo ser ejecutado desde el propio trabajo de Bacula gracias al parámetro con que cuenta el mismo “RunBeforeJob” al cual se le asigna el directorio donde se encuentra el script. El script fue ubicado en la MV2 Bacula Director en el directorio /usr/local/sbin/salva_bd y el contenido del mismo es:

```
#!/bin/bash

BACKUPLOCATION=/home/salva/

mkdir -p $BACKUPLOCATION

# back up all the mysql databases, into individual files so we can later restore
# them separately if needed.
mysql -h 10.51.18.132 -u root -pmysql -B -N -e "show databases" | while read db
do
    BACKUPFILE=$BACKUPLOCATION/$db.mysql
    /usr/bin/mysqldump --single-transaction -h 10.51.18.132 -u root -pmysql $db > $BACKUPFILE
done
```

Para garantizar que en el momento de ejecución del script se está accediendo a la base de datos se creó una mini aplicación web, encargada de hacer inserciones en la base de datos “mia” durante un tiempo aproximado de 15 segundos. En él se insertan datos a tres tablas distintas y para lograr establecer un tiempo entre cada inserción se insertó un ciclo con 100000 iteraciones. El código utilizado es el siguiente:

```
<?php
$link = mysql_connect ("10.51.18.132","root","mysql");
mysql_select_db ("mia",$link);
$val = 0;
```

```

for($i = 0;$i < 100;$i++)
{
for ($j = 0;$j < 100000;$j++) { }
    $val = $val +1;
mysql_query ("INSERT INTO uno (uno, dos)
VALUES ('$val', '$val')", $link);
for ($j = 0;$j < 100000;$j++){ }
mysql_query ("INSERT INTO dos (uno, dos)
VALUES ('$val', '$val')", $link);
for ($j = 0;$j < 100000;$j++) { }
mysql_query ("INSERT INTO tres (uno, dos)
VALUES ('$val', '$val')", $link);
}
?>

```

Una vez ejecutado la aplicación se realizó la ejecución del trabajo SALVA_BD, el cual antes de ejecutarse invoca automáticamente el script descrito anteriormente y como se puede observar en la figura 3.8 el trabajo es realizado con éxito.

```

Build OS:          1686-pc-linux-gnu ubuntu 11.04
JobId:             63
Job:               SALVA_BD.2012-06-11_02.24.39_10
Backup Level:     Full
Client:           "mv2_director" 5.0.3 (04Aug10) 1686-pc-linux-gnu, ubuntu, 11.04
FileSet:          "BD" 2012-06-11 02:09:19
Pool:             "File" (From Job resource)
Catalog:         "MyCatalog" (From Client resource)
Storage:         "File" (From Job resource)
Scheduled time:  11-Jun-2012 02:24:22
Start time:      11-Jun-2012 02:24:50
End time:        11-Jun-2012 02:24:51
Elapsed time:    1 sec
Priority:         10
FD Files Written: 6
SD Files Written: 6
FD Bytes Written: 177,895 (177.8 KB)
SD Bytes Written: 178,530 (178.5 KB)
Rate:            177.9 KB/s
Software Compression: 77.0 %
VSS:             no
Encryption:      no
Accurate:        no
Volume name(s):  vol102
Volume Session Id: 13
Volume Session Time: 1339317074
Last Volume Bytes: 2,174,610,084 (2.174 GB)
Non-fatal FD errors: 0
SD Errors:       0
FD termination status: OK
SD termination status: OK
Termination:     Backup OK

```

Fig. 3.8 Resultado de la ejecución del trabajo SALVA_BD.

Después de ejecutado el mismo se puede observar en la figura 3.9 los archivos creados por el script que fue ejecutado por el propio trabajo de salva, el cual se encargó de su almacenamiento.

```
root@bacula-director-yoelvis:/home# ls /home/salva/  
information_schema.mysql  mia.mysql  mysql.mysql  performance_schema.mysql  test.mysql  
root@bacula-director-yoelvis:/home#
```

Fig. 3.9 Bases de datos salvadas mediante el script.

3.2 Conclusiones

- Con las pruebas realizadas a lo largo de este capítulo se logró demostrar de una forma práctica la validez del procedimiento propuesto, llevando a cabo las actividades que forman parte de cada etapa.
- Se llevaron a cabo la ejecución de diferentes trabajos de forma manual para dar solución a las diferentes problemáticas planteadas.

CONCLUSIONES GENERALES.

- Se describieron los principales aspectos relacionados con el proceso de almacenamiento y salva en un CPD, los cuales sirvieron de base a la propuesta presentada para llevar a cabo el mismo.
- Se demostró que aunque el desarrollo de las herramientas que llevan a cabo estas tareas está abarcado por soluciones privativas, existen soluciones libres que brindan una solución factible para cumplir con el principal objetivo por el cual fue creado el Centro de Datos, la continuidad de los servicios.
- Se realizó una descripción de las actividades que serán llevadas a cabo a la hora de establecer las políticas de salva en un Centro de Datos y el personal encargado de su ejecución.
- Se confeccionó un Manual de Operaciones que cuenta con la descripción precisa de cómo llevar a cabo cada una de las tareas propuestas, quedando conformado un procedimiento que puede ser utilizado por el personal del proyecto Centro de Datos para futuros despliegues.
- Por todo lo antes expuesto se concluye que el objetivo principal propuesto para el presente trabajo ha sido cumplido satisfactoriamente, poniendo en práctica todas y cada una de las tareas propuestas para el desarrollo del procedimiento.

RECOMENDACIONES.

A partir de las conclusiones abordadas se listan las recomendaciones en vistas de posibles mejoras:

- Ampliar el alcance de la presente propuesta con la inclusión del proceso de replicación de la información en Centros de Procesamiento de Datos.
- Puesta en práctica de la presente propuesta en toda la infraestructura de un centro de datos.
- Realización del proceso de salva sobre otros dispositivos soportados por la herramienta propuesta.

REFERENCIAS BIBLIOGRAFICA

1. [En línea] [Citado el: 29 de Marzo de 2012.] <http://www.definicion.org/procedimiento>.
2. SMDATA. [En línea] [Citado el: 16 de Mayo de 2012.] <http://www.smdata.com/formacion.php>.
3. [En línea] [Citado el: 14 de Diciembre de 2011.] <http://itnews.ec/marco/000177.aspx>.
4. [En línea] [Citado el: 9 de Diciembre de 2011.]
<http://www.academiadelanzarote.es/Discursos/Discurso%2034.pdf>.
5. telindus. [En línea] [Citado el: 12 de Febrero de 2012.]
<http://www.telindus.es/propuestatecnologica/solucionesavanzadas/sistemasyalmacenamiento.asp>.
6. [En línea] 11 de Noviembre de 2011.
http://www.fing.edu.uy/tecnoinf/cursos/adminf/material/ADI_apuntes_2009.pdf.
7. [En línea] [Citado el: 11 de Noviembre de 2011.]
http://www.moraldonet.com.ar/services/backup_procedures.htm.
8. [En línea] [Citado el: 11 de Noviembre de 2011.]
http://www.moraldonet.com.ar/services/backup_methods.htm.
9. *ISO 17799. Código para la práctica de la gestión de la seguridad de la información*. 2005. pág. 74, ISO/IEC.
10. [En línea] [Citado el: 12 de Noviembre de 2011.] <http://www2.udec.cl/~dvilche/tarea4/cap2.htm> .
11. Openfiler. [En línea] [Citado el: 21 de Marzo de 2012.] <http://www.openfiler.com/products>.
12. FreeNAS. [En línea] [Citado el: 24 de Marzo de 2012.] <http://www.freenas.org>.
13. [En línea] [Citado el: 30 de Marzo de 2012.] <http://www.gnu.org/philosophy/free-sw.es.html>.
14. Bacula. [En línea] [Citado el: 21 de Diciembre de 2011.]
http://www.bacula.org/presentations/Presentacion_Bacula.pdf.
15. Areca Backup. [En línea] [Citado el: 21 de Diciembre de 2011.] <http://www.areca-backup.org/documentation.php>.
16. BackupPC. [En línea] [Citado el: 21 de Diciembre de 2011.]
http://backupper.sourceforge.net/faq/BackupPC.html#step_1__getting_backuppc.
17. [En línea] [Citado el: 15 de Mayo de 2012.] http://pve.proxmox.com/wiki/Main_Page.

18. [En línea] [Citado el: 15 de Mayo de 2012.] http://pve.proxmox.com/wiki/Backup_-_Restore_-_Live_Migration.

19. [En línea] [Citado el: 23 de Abril de 2012.]

http://www.neovalia.es/imagenes/ficheros/vmware/Datasheets/data_recovery_es.pdf.

20. Veeam. [En línea] [Citado el: 16 de Abril de 2012.] <http://www.veeam.com/es/vmware-esx-backup.html>.

BIBLIOGRAFIA.

1. Vadala, Derek. *Managing RAID on Linux*. s.l. : O'Reilly. 1565927303.
2. Tim Converse, Joyce Park, Clark Morgan. *PHP5 and MySQL Bible*. s.l. : Wiley. 0764557467.
3. *Advanced Encryption Standard - AES*. s.l. : Springer. 3540265570.
4. Alan Calder, Steve Watkins. *IT GOVERNANCE A Manager's Guide to Data Security and ISO 27001/ISO 27002*. s.l. : Kogan Page, June 2008. 0749452714.
5. IBM. [Online] [Cited: Junio 7, 2012.] <http://www.redbooks.ibm.com/redbooks/pdfs/sg245470.pdf>.
6. Preston, W. Curtis. *Backup & Recovery*. January 2007.
7. Corporation, EMC. *Information Storage and Management*. April 2009. 0470294213.
8. Hufferd, John L. *iSCSI The Universal Storage Connection*. November 2002. 020178419X.
9. Sims, Gary. *Learning FreeNAS*. September 2008. 1847194680.
10. Alger, Douglas. *Build the Best Data Center Facility for Your Business*. June 2005. 1587051826.
11. [Online] [Cited: Diciembre 21, 2011.] http://www.bacula.org/presentations/guia_practica_bacula-temporal.pdf.
12. [Online] [Cited: Diciembre 15, 2011.] <http://backuppc.sourceforge.net/>.
13. [Online] <http://backuppc.sourceforge.net/>.