

UNIVERSIDAD NACIONAL DE INGENIERÍA

FACULTAD DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA



**BACKUP DE SERVIDORES USANDO REDES IP, REDES SAN Y LOS
BENEFICIOS DE LA DEDUPLICACIÓN DE DATOS PARA ENTORNOS WAN**

INFORME DE SUFICIENCIA

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO DE TELECOMUNICACIONES

PRESENTADO POR:

MARIANO TORRES RIOS

**PROMOCIÓN
2008 - II**

**LIMA – PERÚ
2013**

**BACKUP DE SERVIDORES USANDO REDES IP, REDES SAN Y LOS
BENEFICIOS DE LA DEDUPLICACIÓN DE DATOS PARA ENTORNOS WAN**

DEDICATORIA

El presente trabajo está dedicado a mis padres
Martin y Dina, a mis hermanos, Yanina.
Y quienes me apoyaron y contribuyeron
en seguir el camino del aprendizaje,
a todos ellos muchas gracias

SUMARIO

El presente trabajo consiste en brindar una solución de backup de un conjunto de servidores distribuidos en el data center principal haciendo el uso conveniente de las redes LAN, redes SAN (Storage Area Network) y una solución de backup de un conjunto de servidores distribuidos geográficamente distantes al data center principal a través de la Red WAN utilizando los beneficios de la tecnología de Deduplicación de datos en el origen. Al principio se da una breve introducción teórica de las redes SAN y del protocolo Fibra Canal usado en este tipo de redes, se hablará de las consideraciones del Backup y Recuperación, en este ítem se explicará cuáles son los requerimientos de RPO y RTO así como también los diferentes tipos de medios para almacenar las copias, descripción del proceso de backup y recuperación; culminando con la descripción de la tecnología de Deduplicación de datos y formas de Deduplicación en el origen y el destino.

Al final del trabajo se registran los detalles de las licencias del Software y Hardware de backup considerados, así como también un sizing de disco que se necesita para almacenar los backup de datos Deduplicados, y el alcance de los servicios profesionales de administración. Este trabajo es planteado para un entorno corporativo donde se tiene un conjunto de servidores con grandes volúmenes de datos (10TB, 20TB, etc.) donde están corriendo aplicaciones como base de datos, y un conjunto de servidores (100GB, 80GB, etc.), distribuidos remotamente que están conectados al data center principal a través de la red WAN.

INDICE GENERAL

INTRODUCCION

CAPITULO I. -----	3
PLANTEAMIENTO DE INGENIERÍA DEL PROBLEMA -----	3
1.1 Descripción del Problema -----	3
1.2 Objetivos del Trabajo -----	5
1.3 Evaluación del Problema -----	5
1.3.1 Evaluación del problema de los servidores distribuidos localmente: -----	8
1.3.2 Evaluación del problema de los servidores distribuidos Remotamente: -----	9
1.4 Limitaciones del Trabajo -----	11
1.5 Sintesis del Trabajo-----	11
CAPITULO II. -----	13
MARCO TEÓRICO CONCEPTUAL -----	13
2.1 Resumen ejecutivo -----	13
2.2 Posicionamiento de EMC -----	14
2.3 Consideraciones de Backup y Recuperación -----	16
2.3.1 Objetivo de Punto de Recuperación (RPO): -----	17
2.3.2 Objetivo de Tiempo de Recuperación (RTO): -----	17
2.3.3 Granularidad de Backup: -----	17
2.3.4 Restaurando desde un Backup Incremental: -----	18
2.3.5 Restaurando desde un Backup Acumulativo: -----	18
2.3.6 Métodos de backup: -----	18
2.3.7 Proceso y arquitectura de Backup: -----	19
2.3.8 Operación de Backup: -----	19
2.3.9 Operación de Recuperación: -----	20
2.4 Tecnologías y topologías de Backup/Recuperación -----	20
2.4.1 Topologías de Backup: -----	20
2.4.1.1 Backup Directamente conectado: -----	21
2.4.1.2 Backup basado en LAN: -----	21
2.4.1.3 Backup basado en SAN (LAN libre): -----	21
2.4.1.4 Backup Mixto:-----	21

2.5 Opciones tecnológicas de Backup	22
2.5.1 Backup a tape o Cinta:	22
2.5.2 Librería física de cinta:	22
2.5.3 Backup a Disco:	22
2.5.4 Comparación de Restauración Cinta vs Disco:	23
2.5.5 Librería de Cintas Virtual:	23
2.6 Desarrollo de las redes SAN	23
2.6.1 Definición SAN:	23
2.6.2 Entendiendo Fibra Canal (FC):	24
2.6.3 Evolución de las redes SAN:	24
2.6.4 Componentes de la red SAN:	24
2.6.4.1 Node Ports:	24
2.6.4.2 Cables ópticos:	25
2.6.4.3 Conectores de Fibra:	25
2.6.4.4 Dispositivos de interconexión:	26
2.6.5 Opciones de interconectividad:	26
2.6.5.1 Punto a Punto:	26
2.6.5.2 FC-AL:	26
2.6.5.3 FC-SW:	26
2.6.6 Tipos de puertos:	27
2.7 FC Arquitectura y Administración	28
2.7.1 Características de la Arquitectura FC:	28
2.7.2 Pila del protocolo Fibra Canal:	28
2.7.3 Direcciones Fibra Canal:	29
2.7.4 World Wide Names:	29
2.7.5 Estructura y organización de FC Data:	30
2.8 EMC Avamar	30
2.8.1 EMC Avamar: Backup & Recuperación con Global 3D:	30
2.8.2 Software Avamar y Oficinas Remotas:	32
2.8.3 Software Avamar y Usuarios Finales:	33
2.8.4 Beneficios asociados a EMC Avamar:	33
2.9 EMC Networker	33
2.9.1 Administración y Seguridad:	34
2.9.2 Backup & Recuperación:	35
2.10 EMC Data Domain	40

CAPITULO III.	43
DESARROLLO DE LA SOLUCIÓN DE BACKUP PARA SERVIDORES EN EL DATA CENTER PRINCIPAL Y SERVIDORES DISTRIBUIDOS REMOTOS MEDIANTE EL USO CONVENIENTE DE LAS REDES SAN Y REDES LAN/WAN	43
3.1 Backup de servidores locales, Data Domain con Deduplicación en el destino	43
3.1.1 Análisis de datos:	43
3.1.2 Consideraciones del sizing Data Domain:	45
3.2 Backup de servidores remotos, Avamar con Deduplicación en el origen	49
3.2.1 Análisis de datos:	49
3.2.2 Consideraciones de sizing Avamar:	51
3.2.3 Diagrama topológico de backup con Avamar:	52
3.3 Solución EMC Networker	54
3.4 Solución EMC Data Domain	54
3.5 Solución EMC Avamar	54
3.6 Diagrama de Solución final	55
CAPITULO IV.	57
ALCANCE DE LOS SERVICIOS PROFESIONALES, SOPORTE Y MANTENIMIENTO	57
4.1 Servicios Profesionales	57
Alcance Avamar	57
Alcance Data Domain	57
Alcance Networker	57
Enfoque y Organización del Proyecto	58
Fase I – Planificación y Diseño	58
Fase II – Implementación	59
Fase III – Pruebas y Aceptación del Proyecto	59
Fase IV – Documentación	59
Entregables	60
Equipo de Trabajo	60
Gerente del proyecto	60
Arquitecto de soluciones	61
Especialista en implementación	61
Localización	61
Fuera de alcance	61
Supuestos	62
4.2 Servicio de Soporte y Mantenimiento Post-Venta	63
Descripción de la estructura de soporte local	63
4.3 Cursos de Capacitación	63

CAPITULO V. -----	64
COSTO DEL PROYECTO Y CONDICIONES COMERCIALES -----	64
5.1 Oferta Económica-----	64
5.2 Condiciones Comerciales-----	64
CONCLUSIONES Y RECOMENDACIONES -----	66
ANEXO A -----	68
EMC Data Protection Advisor -----	68
Solución EMC Data Protection Advisor -----	71
Licencias de Data Protection Advisor para la organización -----	71
ANEXO B -----	73
GLOSARIO DE TÉRMINOS -----	73
BIBLIOGRAFÍA -----	76

INTRODUCCION

La misión de la TI (Tecnología de la Información) siempre ha sido ayudar a las organizaciones a maximizar el valor de su información al menor costo total. Y ahora, el próximo desafío en almacenamiento y la administración del ciclo de vida de la información, permite que la TI cumpla cabalmente esta misión al aportar una comprensión renovada y profunda del valor de la información, de cómo clasificar la información según lo crítica que sea para el negocio y cómo construir la infraestructura de información necesaria para entregar este valor al negocio de manera simple, confiable y rentable.

Una tarea de las organizaciones es proteger esta información mediante una copia de seguridad llamado Backup y estar preparados en caso de pérdida de datos, de no ser así los efectos que implica la pérdida de la información pueden causar daños no menores, incluso irreversibles desde el punto de vista del corazón del negocio; esta tarea de protección de datos se convierte en un reto para tener un backup consistente y una rápida recuperación de la información que tiende a aumentar en toda empresa la cual podría ser expandido en múltiples lugares. Además, las organizaciones necesitan realizar backup con el menor costo y con el mínimo recurso.

Las organizaciones deben asegurar la información correcta, en el lugar correcto y en el tiempo correcto, evaluando las tecnologías de backups, recuperación y retenciones requeridas para los datos y las aplicaciones; este es un paso esencial para asegurar una implementación de solución de backup y una recuperación exitosa. La solución debe facilitar una fácil recuperación de las copias y archivos como requerimiento del negocio.

La solución planteada que integra estos requerimientos que puede ser cubierto por una gama de proveedores en el mercado. En función a la gama de soluciones se tomará a la empresa EMC como empresa proveedora a cubrir esta solución de backup, que brinda una solución integradora permitiendo a las organizaciones poder administrar los procesos de backups de manera sencilla y centralizada, además de contar con una solución que permita hacer frente al crecimiento de los negocios y regulaciones exigidas para el almacenamiento de la información.

A lo largo de los capítulos se explicará la solución de backup desde la perspectiva de desarrollo y arquitectura de backup, haciendo uso de la arquitectura implementada por la organización como una red LAN, red WAN y red SAN; asociado en función a sus requerimientos de tareas de backup (políticas de backup) y crecimiento del tamaño de la información constante esperada del 10% por año para el caso de los servidores locales y 50% para el caso de los servidores remotos. En tal sentido esta solución estará cubierta por 3 años sin necesidad de renovar las licencias de Software o Hardware por los próximos 3 años.

Cabe señalar que la red LAN Ethernet base 100Mbps o 1Gbps, la red SAN 4Gbps o 8Gbps son comúnmente desplegados en todo entorno corporativo donde se maneja grandes volúmenes de información así como también el despliegue de red WAN como modo de conexión y operación entre los servidores remotos y el data center principal. En este caso de estudio la organización cuenta con estas redes implementadas y en ejecución.

Agradezco la colaboración de mis compañeros de trabajo y de todas las fuentes de información.

CAPITULO I.

PLANTEAMIENTO DE INGENIERÍA DEL PROBLEMA

1.1 Descripción del Problema

En este ambiente de la tecnología de la información donde se vive una explosión del crecimiento de la información, las personas, los ecosistemas, las tecnologías e incluso los negocios avanzan a través de ciclos de vida, distintas fases de desarrollo que van desde el nacimiento hasta la adolescencia, madurez y ocaso.

Por ejemplo, el historial médico de cada persona es valioso en el momento en que su médico lo registra por primera vez. Meses o años más tarde esta información, podría llegar a tener una importancia crítica si esa persona se encontrara en una sala de emergencia o considerando un procedimiento de seguimiento para una condición existente. En los negocios, ocurre constantemente una fluctuación en el valor de la información y por lo general, esta se vincula a los ciclos operativos de la empresa, que pueden ser las transacciones diarias o por horas que generan ingresos, el procesamiento y distribución quincenal de las nóminas de pagos, el cierre de un trimestre financiero, el archivado de impuestos anuales o un acontecimiento repentino e inesperado, como una nueva amenaza competitiva, una auditoria, adquisición o desastre.

El objetivo de toda organización es proteger esta información mediante una copia de seguridad o backup, en tal sentido este trabajo cubre la necesidad de realizar backup de un grupo de servidores locales y de un conjunto de servidores distribuidos geográficamente distantes (remotos) del data center principal, planteando una solución con las redes comúnmente implementada en la organización y entendiendo los límites en recursos que posee dicha organización.

Para el caso de los servidores distribuidos localmente y remotamente se necesita realizar backup, en tal sentido las empresas tienen bien claro el valor de la información de su entorno y buscan conseguir siempre las copias en el menor tiempo posible, sin embargo en tal proceso se ha identificado inconvenientes comunes (el cual se detalla) los problemas o limitaciones principales en ambos ambientes:

Problemas o limitantes de servidores distribuidos localmente:

Los servidores a respaldar contienen grandes volúmenes de datos (2TB, 10TB o más) con aplicaciones críticas o transaccionales como base de datos (SQL, ORACLE, etc.), servidores File Server que al momento de realizar las tareas de backup consumen muchos recursos de hardware, recursos de la red LAN así como también los recursos en el servidor de backup por lo tanto hacen que los procesos de backup sean lentos.

Los tiempos estimados (ventanas de backup) para respaldar estos servidores de grandes volúmenes de datos que varias veces sobrepasan las 24 horas realizando este proceso, esto ocurre porque el ancho de banda disponible en la red LAN del cliente (10Mbps, 100Mbps, 1Gps), no es suficiente; esto se hace insostenible porque esos servidores deben estar con la máxima disponibilidad a fin de poder atender la demanda de procesos realizados en horarios de oficina o de acceso.

Problemas o limitantes de servidores distribuidos Remotamente:

Los servidores a respaldar no contienen grandes volúmenes de datos pero se encuentran distribuidos geográficamente distantes del data center principal y están conectados a través de la red WAN, una de las formas tradicionales de proteger esta información es realizar backup en cada uno de estos lugares remotos en dispositivos de backup como tape, disco extraíble, CD-Rom, etc.; luego trasladar físicamente estos dispositivos por medios terrestres, aéreos, etc. hacia el data center principal. Esto implica el despliegue de toda una tarea logística y de coordinación comprometiendo recursos del personal autorizado y lo más riesgoso es exponer esta información que esté sujeto a daños físicos irreversibles en el trayecto del transporte físico o también exponer esta información a ser adulterada o corrupta.

Por otro lado realizar backup tradicional directo desde cada sitio a través de la red WAN es una tarea complicada, básicamente por el recurso limitado del ancho de banda disponible que resultaría en algunos casos ventanas de backup mayores de 24 horas así como también un gran consumo de recursos de la red WAN que deja poca opción a otros procesos como alertas o reportes transaccionales, correo electrónico, acceso a internet,

VoIP, etc. una solución inmediata es incrementar el ancho de banda WAN, pero a veces resulta inconveniente desde el punto de vista económico porque implica el pago adicional a un proveedor que brinda este tipo de servicio.

Uno de los principales objetivos de toda organización es tener todos los backup remotos centralizados y administrados desde el data center principal para luego poder direccionar estos backup a un sitio alternativo y contar con una solución de recuperación de desastre.

1.2 Objetivos del Trabajo

Tabla N° 1.1. Descripción de los Objetivos

Objetivos	Descripción de los objetivos
1	Proporcionar una solución que permita a la organización contar con un respaldo de toda la información productiva en su ambiente local, dentro de unas ventanas de backups establecidas y de manera automática con el propósito de liberar al servidor que está siendo respaldado y que maneja la información crítica estar la mayor cantidad de tiempo con los mayores recursos disponibles atendiendo los procesos que le demandan.
2	Proporcionar una solución que permita a la organización contar con un respaldo como en el punto anterior pero de su ambiente remoto a través de los recursos limitados en la red WAN, permitiendo almacenar los backups a través de la red WAN de manera centralizada, segura y confiable.
3	Proporcionar una solución que permita a la organización poder replicar las copias de los servidores remotos a un sitio de contingencia, para luego contar con las copias cuando el sitio principal está incapacitado debido a un desastre ocurrido; administrar los procesos de copias de manera sencilla y centralizada para ambos ambientes.
4	Adicionalmente se presenta un escenario de software de monitoreo que permite alcanzar a la organización tener los reportes, análisis y capacidad de planificación de todas las cosas que está pasando en su entorno de backup, así como también el análisis y las conclusiones para futuros proyectos que implique la protección de datos de un conjunto de servidores.

1.3 Evaluación del Problema

Para brindar una solución acertada se requiere una especial atención a los puntos adversos que suceden en el ambiente de backup de la organización, para lo cual se necesita recolectar información del cliente mediante reuniones, llamadas de conferencia o en caso contrario el cruce de e-mails, queda implícito que mientras se logre recolectar una mayor información se podrá brindar una solución más fina.

Siendo este caso el de proteger la información de servidores locales y remotos se procederá analizar estos dos ambientes por separado con el fin de tomar por un instante el papel de un

“francotirador” y proporcionar solución por cada punto adverso en el ambiente de backup para luego consolidarlo en una sola solución centralizada; en tal sentido se detalla el proceso del análisis:

Situación actual de la organización:

La organización cuenta con unos servidores conectados a la red Ethernet de 100Mbps y otros a la red Ethernet de 1Gbps una red SAN de 4Gbps un switch Fibra Canal de 32 puertos en la cual tienen disponible 20 puertos.

La organización cuenta con una distribución de servidores remotos interconectados a través de un enlace WAN de 6Mbps, la cual debido a una latencia en el momento del transporte de datos puede reducir este ancho de banda hasta un 80% de eficiencia, además se debe tener en cuenta que el enlace sirve para otros propósitos críticos, así como también reportes de falla de equipos, correo electrónico, entorno web, VoIP, etc.

Por temas de disponibilidad y continuidad del negocio la organización cuenta con accesorios como cables fibra óptica Multimodo, puertos de switch FC, cables de red LAN, conectores, tarjetas HBAs, tarjetas de red NIC, etc. en el caso que se presente una falla el mismo personal de la organización pueda realizar ese trabajo de reemplazar las partes que han fallado.

La organización cuenta con una librería Robótica HP ESL E-Series 630e y están usando Networker Power Edition como software de backup instalado en Windows Server 2008; donde las políticas de backup son de lunes a viernes backup Incremental y los fin de semanas full backup.

- Retención Full Backups semanales : 4 semanas
- Retención Backups Incrementales diarios : 7 días

Las retenciones que actualmente desarrolla la organización son de 7 días para las copias Incrementales, y de 4 semanas de retención para las copias Full.

La organización realiza backup de los servidores locales en San Isidro y transporta los tapes físicos al sitio de contingencia ubicado en Monterrico, y tienen vigencia de mantenimiento.

La organización posee una distribución de servidores con diferentes tamaños de datos tal como se adjunta más adelante. En la siguiente figura 1.1 se muestra la situación actual.

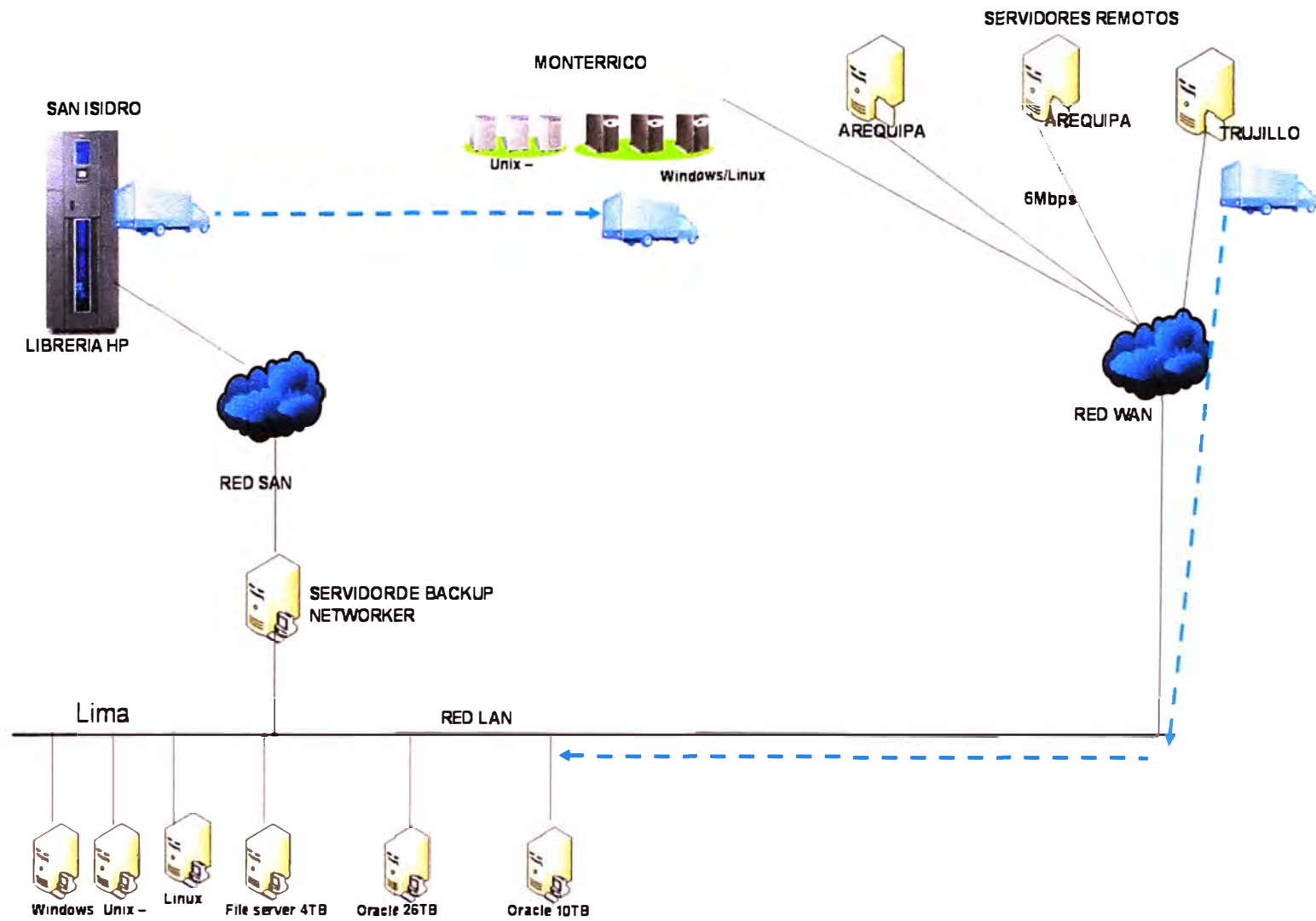


Figura 1.1. Esquema de la Situación actual de la organización

1.3.1 Evaluación del problema de los servidores distribuidos localmente:

Tabla N° 1.2. Distribución de servidores locales con grandes volúmenes de datos

CPU	Virtua l	Modelo	Tipo de Conexi ón	Espaci o usado en GB	DB	Tipo de Backup	Networker Agent
32	NO	SUN (sparc)	1Gbps	26000	ORACL E 10i	RMAN	V. 7.2.1.Build.311
32	NO	SUN (sparc)	1Gbps	10000	ORACL E 10i	RMAN	
4	NO	HP SUPERDOM E	100Mbps	540	ORACL E	RMAN	V. LNMs_2004.Build.273
7	NO	HP SUPERDOM E	1Gbps	950	ORACL E	RMAN	V. LNMs_2004.Build.273
8	NO	HP SUPERDOM E	1 Gbps	1000	ORACL E 10i	RMAN	V. 6.1.4.Build.562
6	NO	HP RX8640 (Windows 2008)	1 Gbps	4800			V. 7.1.3.Build.404
2	NO	HP GS320	100Mbps	1574	ORACL E	RMAN	V. 7.2.1.Build.311
4	NO	Proliant DL 580 G2	100 Mbps	1100	ORACL E	EXPOR T	V. 7.1.3.Build.404
4	NO	Proliant DL 580 G2	100 Mbps	100	ORACL E	EXPOR T	V. 7.1.3.Build.404
4	NO	Proliant DL 580 G2	100 Mbps	1100	ORACL E	EXPOR T	V. 6.1.4.Build.562
4	NO	Proliant DL 580 G2	100 Mbps	100	ORACL E	EXPOR T	V. 6.1.4.Build.562
8	NO	HP SUPERDOM E	1GB	1900	ORACL E	RMAN	V. LNMs_2004.Build.273
20	NO	HP SUPERDOM E	1Gbps	2500	ORACL E	ALTER A TAPE	V. 6.1.4.Build.562
17	NO	HP SUPERDOM E	1Gbps	1000	ORACL E	RMAN	V. LNMs_2004.Build.273
2	NO	Proliant DL 380 (G1)	100mbps	400			V. 7.1.3.Build.404
2	NO	HP DS20	100 Mbps	1000			V. 7.2.1.Build.311

Los servidores a respaldar contienen grandes volúmenes de datos (10TB, 20TB o más) con aplicaciones críticas, por la cantidad de información que se tiene resulta inconveniente que el flujo de datos de backup de estos servidores pasen directamente por la red LAN, no solamente porque compromete los recursos de la red LAN si no también los recursos del servidor de backup que se encarga de procesar los datos que recibe de estos servidores que están siendo respaldados y envía esta data a los dispositivos de backup (tape físico, disco); entonces para evitar estos inconvenientes se recomienda que los servidores que manejan grandes volúmenes de datos se deba instalar y configurar como un Storage Node (se

detallará más adelante), conectarlos a la red SAN y luego puedan realizar los respaldos directamente a los dispositivos de backup a través de la red SAN y enviar por la red LAN sólo los reportes o metadata al servidor de backup, esto permitirá liberar recursos de la red LAN, liberar recursos del servidor de backup y aumentar la velocidad de backup de los demás servidores.

Estos de servidores de grandes volúmenes de datos tienen aplicaciones como base de datos Oracle, SQL, que desarrollan respaldo mediante el agente Networker cliente y el agente Networker para base de datos, y todo este transporte de datos respaldo es a través de la red LAN hacia el servidor Networker y desde este último servidor transmite los datos hacia los dispositivos de backup a través de la red SAN.

Como en el ítem anterior estos servidores de grandes volúmenes, la transmisión de datos se debe direccionar para que este flujo pase por la red de alta velocidad SAN y no por la red LAN, para cumplir este objetivo se necesita unas licencias adicionales llamadas Networker Storage Node para los respectivos sistemas operativos (Windows, Linux, UNIX), en las cuales se está corriendo las aplicaciones de base de datos especificados.

1.3.2 Evaluación del problema de los servidores distribuidos Remotamente:

Tabla N° 1.3. Distribución de servidores remotos.

Categoría	Sistema Operativo	CPU	Virtual	Modelo	Tipo de conexión	Espacio usado en GB	Networker	Localización
Producción	Microsoft Windows	2	NO	Proliant ML 570	100Mbps	100	V. 6.1.3 Build.428	AREQUIPA
Producción	Microsoft Windows	4	NO	Proliant DL 380	100Mbps	80	V. 6.1.3 Build.428	AREQUIPA
Producción	Microsoft Windows	2	NO	Proliant ML 570	100Mbps	48	V. 6.1.3 Build.428	TRUJILLO
Producción	Microsoft Windows	1	NO	Proliant DL 360	100Mbps	60	V. 6.1.3 Build.428	TRUJILLO
Producción	Microsoft Windows	4	NO	Proliant DL 380	100Mbps	80	V. 6.1.3 Build.428	TRUJILLO

Los servidores a respaldar se encuentran distribuidos geográficamente distantes del data center principal y están conectados a través de la red WAN con un enlace de 6Mbps para el sitio de Trujillo y Arequipa.

Para el caso de los servidores distribuidos en Trujillo el tamaño total de la data actual es de 188GB, adicionalmente el cliente está creciendo en 50% anual para lo cual se espera que el tamaño total dentro de los 3 años será de aproximadamente 423GB considerando los cambios, considerando la latencia y la disponibilidad del enlace hasta 4Mbps se estaría contemplando una ventana de backup full de aproximadamente 29 horas, entonces

claramente se observa que realizando un backup tradicional no es la mejor opción. Considerando el tamaño de la data que es aproximadamente 423GB, existe una solución de backup llamado Avamar que tiene tecnología de Deduplicación de datos en el origen (más adelante se explica en detalle como es el proceso de Deduplicación de datos), la cual detecta e identifica bloques únicos de datos a los efectos de reemplazar los datos duplicados por referencias, por lo tanto al momento de realizar backup sólo enviaría por la red WAN aquellos bloques únicos de datos.

Para el caso de los servidores distribuidos en Arequipa el tamaño total de la data actual es de 180GB adicionalmente el cliente está creciendo en 50% anual para lo cual se espera que el tamaño total dentro de los 3 años será aproximadamente de 405GB considerando los cambios, el análisis para este caso es similar al anterior, en donde se usa la solución de Avamar con tecnología de Deduplicación para realizar backup de estos ambientes remotos. El proceso es instalar un agente Avamar (cliente Avamar) en cada servidor remoto, los datos de backup viajarán a través de la red WAN y serán almacenados en el nodo Avamar (hardware de disco que más adelante se explica al detalle) ubicado en el sitio local de Lima en San Isidro, adicionalmente la primera copia de seguridad completa de archivos es la que requiere más tiempo que las siguientes copias de seguridad por ser estos bloques únicos de información que requieren ser enviados a través de la red WAN.

Avamar permite realizar backup full todos los días, con una reducción de ventanas de backup 10 veces y reducción del consumo de ancho de banda de la red WAN de hasta un 98% considerando las retenciones especificadas por la organización, permitiendo el normal desarrollo de otros procesos como alertas, reportes transaccionales, correo electrónico, acceso a internet, VoIP, etc.

Por otro lado para garantizar la integridad y la seguridad al momento de realizar backup a través de la red WAN el software Avamar tiene tres niveles de encriptación. El primer nivel de encriptación envía la data desde el cliente Avamar remoto en formato comprimido y Deduplicado hacia el nodo Avamar; el segundo nivel de encriptación usa 128 bits AES (Advanced Encryption Standard) Encryption; el tercer nivel de encriptación usa 256 bits AES Encryption.

Una vez que la data Deduplicada es almacenada en el nodo Avamar ubicado en San Isidro podrá ser replicado vía IP hacia el otro nodo Avamar ubicado en Monterrico en formato Deduplicado, es decir los dos nodos Avamar “conversan” y sólo replica los bloques únicos de datos y sólo envía referencias cuando detecta algún dato duplicado.

1.4 Limitaciones del Trabajo

El presente trabajo, responde a una solución de cada ítem detallado en las descripciones de la evaluación del problema, entonces toda solución o implementación no mencionada queda explícitamente fuera del presente trabajo.

- Los siguientes análisis quedan explícitamente fuera del alcance:
- Análisis de performance de la red LAN, SAN y WAN.
- Rediseño de la red SAN, LAN.
- Análisis de performance a nivel de hardware del servidor de backup.
- Análisis de performance de los sistemas operativos Windows, Linux o Unix.
- Iniciativas que requieran análisis y programas de aplicaciones como base de datos Oracle, SQL, Exchange, etc.
- Detalles específicos de instalación de Avamar, Networker o Data Domain tales como pasos de instalación de agentes, configuraciones de Base de Datos Oracle RMAN, configuraciones Base de Datos SQL Server.
- Desarrollar soluciones de backup de dispositivos NAS (Network Attached Storage) usando el protocolo NDMP (Network Data Management Protocol).
- Desarrollar soluciones de High Available (Cluster de Sistemas Operativos).
- Desarrollar soluciones de clonado de cintas, Archiving de datos, Snapshot basado en software o Storage, eventualmente se puede mencionar en calidad informativa.
- Cualquier actividad o análisis de respaldo o recuperación más allá del alcance especificado.

1.5 Síntesis del Trabajo

A lo largo del capítulo II se expone el marco teórico y conceptual base para el desarrollo del presente trabajo. Como primer punto se da un resumen ejecutivo de la corporación EMC dedicada 100% a la solución de almacenamiento en red automatizado.

Como segundo punto se da a conocer conceptos acerca de las consideraciones de Backup y Recuperación, así como también el desarrollo de las redes SAN y el protocolo Fibra Canal.

Como tercer punto abarca las descripciones del software Avamar y cómo se desarrolla la tecnología de Deduplicación, Networker como software de backup y cómo Avamar se integra a este software de backup con el objetivo de administrar los backup locales y remotos desde una misma consola de administración, también se hace una descripción de Data Domain como solución de almacenamiento de backup con la tecnología de Deduplicación en el destino.

En el capítulo III se presenta el diagrama explicativo para el respaldo de los servidores que realizan backup directamente por la red SAN así como también los que realizan backup por la red LAN, los servidores remotos cuyo flujo de datos será a través de la red WAN y almacenada en el nodo Avamar para luego esta data pueda ser replicada aun lugar de contingencia.

En el capítulo IV se detalla el alcance de los servicios profesionales que desarrolla EMC para la implementación de las soluciones de backup, el tiempo de soporte y mantenimiento como el detalle de los contactos para un caso de soporte técnico.

En el capítulo V se detalla el alcance del costo de la solución del proyecto y las condiciones comerciales.

Como parte final de este trabajo en calidad de anexo se presenta un escenario de software de monitoreo que permite alcanzar a una organización tener los reportes, análisis y capacidad de planificación de todas las cosas que están pasando en su entorno de backup, así como también las conclusiones y recomendaciones.

CAPITULO II.

MARCO TEÓRICO CONCEPTUAL

2.1 Resumen ejecutivo

EMC es la única empresa en el mundo que está dedicada 100 por ciento al almacenamiento en red automatizado y a convertir el almacenamiento en la infraestructura de TI clave para obtener más y más valor de la información, de una manera simple y rentable. De hecho, la visión de EMC es crear una empresa que cuente con lo último en administración del ciclo de vida de dicha información para ayudar al cliente a obtener el máximo valor de su información al costo total más bajo en cada punto del ciclo de vida de la misma. Esta visión se apoya en el historial de innovación de EMC durante las últimas décadas respecto de la constante expansión de las capacidades del almacenamiento, al introducir más y más herramientas y automatización que reduzcan la complejidad, aminoren los costos y aumenten la productividad del personal de TI.

En los últimos 10 años, el 100 por ciento de la experiencia técnica e inversión de EMC se han concentrado en el almacenamiento en red automatizado, que es la base de la administración del ciclo de vida de la información. El almacenamiento en red automatizado proporciona a los clientes la única infraestructura integrada y automatizada, esencial para crear capas de almacenamiento en niveles y múltiples niveles de servicio. Cada día esta base se fortalece a medida que EMC agrega nuevas funcionalidades para administrar la información y aumenta la inteligencia, automatización e integración de la infraestructura de almacenamiento.

Actualmente, nadie puede igualar a EMC en el alcance de sus plataformas de almacenamiento en red en niveles, la amplia conectividad para SAN, NAS y CAS, la administración integrada de almacenamiento y las ofertas de software de administración inteligente de la información, o los servicios y soluciones comprobadas de EMC que pueden ayudar a los clientes en cada etapa del camino, desde la clasificación de datos y mapeo de aplicaciones hasta la implementación y el soporte.

La experiencia de la infraestructura de almacenamiento, el software de administración de almacenamiento, servicios comprobados y 10.000 profesionales dedicados al almacenamiento en las áreas de ventas y servicios; esto es lo que ubica a EMC muy por encima de cualquier otro proveedor en el camino hacia el cumplimiento de la promesa de la administración del ciclo de vida de la información.

EMC terminó el 2010 como líder del mercado a nivel mundial comandando con un 64.2% con una facturación anual cercana a los \$1.7 billones. EMC invierte más del 10% de sus ingresos en Investigación y Desarrollo, lo que le permitirá seguir subiendo en este ranking en el futuro.

En Hardware – Network Storage (SAN+NAS), EMC tiene un porcentaje de participación de mercado del 64%, siendo su competidor más cercano del 16.5%.

En Software – Storage Software, EMC tiene un porcentaje de participación del mercado del doble que el segundo competidor, con un total de 28%.

2.2 Posicionamiento de EMC

EMC ha avanzado en tres fases desde principios de los años 90. EMC ha avanzado de

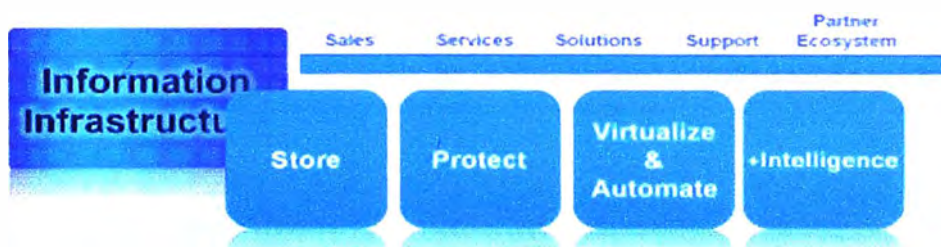


Figura 2.1. Posicionamiento EMC

ser una empresa de almacenamiento de información hasta convertirse en una empresa de administración de la información, donde se posiciona hoy en día, como una empresa de soluciones para la implementación de una infraestructura de información confiable.

EMC está posicionada como líder en soluciones de información globalmente.

EMC se dedica 100% al diseñar soluciones de Infraestructura de la información en base a su largo conocimiento de las tecnologías del almacenamiento. Otras compañías dedican sólo 2% de sus recursos.

Globalmente, EMC está posicionada en más de 80 países con más de 47,000 empleados en el mundo.

EMC es una empresa dedicada exclusivamente a la gestión de la información. Durante los últimos años ha adquirido más de 30 compañías de tecnología, todas líderes en su género, entre las que cabe recalcar:

- Data Domain
- VMware
- Avamar
- RSA Security
- Smarts
- Legato
- Data General

La inversión en adquisiciones de EMC es superior a los 10,500 millones de dólares.

EMC ha agregado nuevas capacidades para proteger y asegurar la información de nuevas formas.

Ha desarrollado una amplia gama de capacidades para ayudar a las empresas a optimizar su entorno.

Tabla N° 2.1. Top 5 Vendors, Worldwide Storage Software Revenue, for the First Quarter of 2011 (Revenues are in Millions)

Vendor	1Q11 Revenue	1Q11 Market Share	1Q10 Revenue	1Q10 Market Share	1Q11/1Q10 Revenue Growth
EMC	\$828.9	23.8%	\$695.6	22.4%	19.2%
Symantec	\$559.0	16.0%	\$528.4	17.0%	5.8%
IBM	\$501.6	14.4%	\$455.2	14.7%	10.2%
NetApp	\$293.2	8.4%	\$295.7	9.5%	-0.8%
Hitachi	\$156.87	4.5%	\$93.4	3.0%	69.0%
All Vendors	\$3,486	100%	\$3,103	100%	12.4%

Para ayudar a optimizar su infraestructura, EMC ha invertido considerablemente en virtualización.

Para ayudar a optimizar la manera en que administra la información, EMC ha desarrollado un gran conjunto de nuevas herramientas.

Para ayudar a optimizar los niveles de servicio que se ofrecen, EMC ha adquirido tecnologías innovadoras.

Finalmente, para ayudar a los clientes a aprovechar la información que ya posee, EMC ha invertido en poderosas herramientas para ayudarle a hacerlo.

Al juntar todas las piezas se podrá observar la forma en la que EMC está construyendo la infraestructura de información.

Con más de 14.000 equipos vendidos Data Domain es líder mundial en Deduplicación en destino, integrándose con todas las plataformas de respaldo Enterprise del mercado, como Legato Networker, Symantec NetBackup, entre otros, incluyendo ambientes open, mainframe e iSeries.

EMC Consultor posee la flexibilidad necesaria para considerar las cuatro áreas de servicio para todas las aplicaciones de negocios o categorías de datos no estructurados que se requiera.

EMC Consultor trabaja para verificar qué tipos de nivel de servicio son importantes para el negocio y continuidad del mismo. Por ejemplo, cuando se trata de Servicios de Almacenamiento de información primario, ¿interesa garantizar el performance, la disponibilidad u otros servicios?

Una vez que el cliente y EMC han acordado cuales son los niveles de servicio más importantes para el negocio, EMC reúne la información de negocio en grupos o clases de información con requerimientos similares y coordina con el departamento de TI para especificar la forma en que satisfará los requerimientos de nivel de servicio determinando sus características técnicas.

2.3 Consideraciones de Backup y Recuperación

Algunas decisiones importantes que necesitan consideración antes de implementar una solución de Backup/Recuperación son mostrados:

- El Objetivo de Punto de Recuperación (RPO)
- El Objetivo de Tiempo de Recuperación (RTO)
- El tipo de media a ser usado (disco o tape físico)
- Donde y cuando la operación de restauración ocurre, especialmente si un host alternativo es usado para recibir la data restaurada.
- Cuando desarrollar un backup.
- La granularidad de los backup, Full, Incremental o acumulativo.
- Cuanto tiempo mantener el backup; por ejemplo, algunos copias necesitan estar retenidos por 4 años, otros sólo por un mes.
- Es necesario o no hacer una copia del backup.
- Localización local o remota, plataformas heterogéneas.
- Número de archivos y tamaño.

2.3.1 Objetivo de Punto de Recuperación (RPO):

Es el máximo periodo tolerable en la cual la data podría perderse desde un servicio de TI debido a un incidente mayor. En la figura 2.2 se interpreta el concepto

2.3.2 Objetivo de Tiempo de Recuperación (RTO):

Es la duración de tiempo y un nivel de servicio dentro del cual un proceso de negocio debería ser restaurado después de un desastre con el fin de evitar una inaceptable consecuencia asociado con la ruptura en una continuidad del negocio. Ver figura 2.2.



Figura 2.2. Esquema de RPO y RTO

2.3.3 Granularidad de Backup:

Granularidad de Backup depende de la necesidad del negocio y RPO/RTO requeridos. Basados en granularidad, las copias pueden ser categorizados como Full, acumulativo, e incremental. Muchas organizaciones usan una combinación de esos tres tipos de copias para sus requerimientos de backup y recuperación

Backup Completo (Full backup): Es un backup de un volumen completo de la data en producción en cierto punto del tiempo. Un copia de un full backup es creado por copia de la data en la producción de volúmenes a un dispositivo de repositorio secundario.

Backup Incremental: Es la copia de la data que ha sido cambiado desde el último Full backup o Backup Incremental, cualquiera que ha ocurrido recientemente. Esto es mucho más rápido (porque el volumen de la data respaldada es restringido a los datos cambiados), pero esto toma un mayor tiempo al momento de desarrollar una recuperación.

Backup Acumulativo (backup diferencial): Copia la data que ha sido cambiado desde el último Full backup. Este método toma mayor tiempo que un Backup Incremental pero es más rápido restaurar.

Backup Sintético (backup constructivo): Es usualmente creado de un reciente Full Backup y todos los incrementales backup desarrollados después de un full backup. Un backup sintético permite una copia de Full backup ser creado offline sin interrumpir los procesos I/O en el volumen de producción. Esto también libera los recursos de la red en los procesos de backup, haciendo ellos disponibles para otros procesos.

2.3.4 Restaurando desde un Backup Incremental:

El proceso de restauración desde un incremental backup se requiere del último Full backup y de todas las copias incrementales disponibles hasta el punto de restauración.

2.3.5 Restaurando desde un Backup Acumulativo:

El proceso de restauración desde un Backup Acumulativo se requiere del último Full backup y del más reciente Backup Acumulativo.

2.3.6 Métodos de backup:

“Hot backup” y “Cold backup” son dos métodos desarrollados para copias basados en el estado de la aplicación mientras se desarrolla el backup. En un Hot backup, la aplicación esta activa y corriendo, con usuarios accediendo a sus datos. En un Cold backup, la aplicación no está activa durante el proceso de backup.

Un backup consistente de base de datos puede ser echo usando “Cold backup”. La desventaja de un “Cold backup” es que la base de datos es inaccesible para los usuarios durante el proceso de backup. “Hot backup” es usado en situaciones donde no es posible bajar los servicios de la base de datos. Esto es facilitado por “database backup agents” que puede desarrollar un backup mientras la base de datos esta activa. La desventaja asociado con “Hot backup” es que los agentes usualmente afectan el total performance de la aplicación.

Un método de copia point-in-time (PIT) es desarrollado en ambientes donde el impacto de tiempo de inactividad de un cold backup o la performance resultando de un Hot backup es inaceptable. Una copia point-in-time consume sólo una fracción del espacio del storage de disco y puede ser creado muy rápido. En este método de backup la base de datos es parada o congelada momentáneamente mientras una copia PIT es creado. La copia PIT es montado en un servidor secundario y el backup ocurre en un el primer servidor. Para asegurar consistentemente los atributos y propiedades adicionados al archivo, tal como permisos, propietarios, y otras metadata, también necesitan ser copiados.

Un ambiente de “recuperación de desastres”, bare-metal recuperación (BMR) se refiere a un backup en cual toda la metadata, información del sistema, y configuraciones de aplicaciones que son propiamente copiados para una recuperación total del sistema. BMR recupera primero la base del sistema, antes de empezar a recuperar los archivos de datos. Algunas tecnologías de BMR pueden recuperar un dentro de un servidor con similares características de hardware.

2.3.7 Proceso y arquitectura de Backup:

Un sistema de backup utiliza la arquitectura cliente/servidor con un servidor de backup y varios clientes de backup. El servidor de backup administra las operaciones de backup y mantiene el catálogo de backup, que contiene información sobre el proceso de backup y metadatos de backup. El servidor de backup depende de los clientes de backup para recopilar los datos a ser copiados. Los clientes de backup puede ser local en el servidor o que puede residir en otro servidor, supuestamente para realizar copias de seguridad de los datos visibles a ese servidor. El servidor de backup recibe los metadatos de backup de los clientes de backup para llevar a cabo sus actividades.

El storage node es responsable de escribir los datos a un dispositivo de backup (en un ambiente de backup, un storage node es un host que controla los dispositivos de backup). Típicamente, el storage node es integrado con el servidor de backup y ambos están conviviendo en misma plataforma física. Un dispositivo de backup es directamente conectado a la plataforma del storage node. Algunas arquitecturas de backup refiere el storage node como el media server porque esto conecta a los dispositivos de backup. Storage node juega un rol importante en un planeamiento de backup porque ellos pueden ser usados para consolidar los servidores de backup.

2.3.8 Operación de Backup:

Cuando un proceso de backup es iniciado, una significativa comunicación de red toma lugar entre diferentes componentes de una infraestructura de backup. El servidor de backup inicia el proceso de backup para diferentes clientes basado en un cronograma (schedule) de configuración para ellos. El servidor de backup coordina el proceso de backup con todos los componentes en una configuración de backup. El servidor de backup mantiene la información acerca de los clientes de backup a ser conectado y un storage node a ser usado en la operación de backup. El servidor de backup recupera la información de backup relacionados desde el backup del catálogo y, basado en esta información, ordena al storage node cargar la apropiada media de backup dentro del dispositivo de backup. Simultáneamente, este instruye a los clientes de backup para enviar la metadata al backup server, y la data, ser enviada a un apropiado storage node. En recibimiento a este requerimiento, el backup cliente envía la metadata al backup server. El backup server escribe esta metadata en su catálogo. El backup cliente también envía la data al storage node y el storage node escribe la data a un dispositivo de almacenamiento. El storage node, también envía la metadata al servidor de backup para mantener esta actualizada acerca de

qué media está siendo usada en el proceso de backup. El backup server continuamente actualiza el catálogo con esta información.

Después toda la data es copiada, el storage node cierra la conexión al dispositivo de backup. El backup server escribe el estatus de backup completado al catálogo.

2.3.9 Operación de Recuperación:

Un proceso de recuperación debería ser manualmente inicializado. Esta aplicación de recuperación es accesible sólo por el administrador. Bajo el recibimiento de un requerimiento de recuperación, un administrador abre la aplicación de recuperación para observar la lista de clientes que han sido respaldados. Mientras la selección del cliente por el cual un requerimiento de recuperación ha sido hecho, el administrador también necesita identificar el cliente que recibirá la data restaurada. La data puede ser restaurada en el mismo cliente por quien el requerimiento ha sido hecho o algún otro cliente. El administrador entonces selecciona la data para ser restaurada en un específico punto en el tiempo para la cual la data tiene que ser restaurada basado en un RPO (Objetivo de Punto de Recuperación).

El administrador primero selecciona la data a ser restaurada e inicia el proceso de restauración de la data que fue perdida o corrupta con el objetivo de volver a un punto inicial relativo.

El backup server, usando un apropiado storage node, entonces identifica un backup media que necesita ser montado en el dispositivo de backup. La data es entonces leída y enviada al cliente que ha sido identificado para recibir la data restaurada. Algunas restauraciones son satisfactoriamente completadas por recuperar sólo el requerimiento de la data en producción. Por ejemplo, el proceso de recuperación de una hoja de cálculo es completado cuando el archivo específico es restaurado. Esto asegura una consistente aplicación para restaurar la data. En estos casos, el RTO es extendido debido al adicional paso en el proceso de restauración de la data que fue perdida o corrupta.

2.4 Tecnologías y topologías de Backup/Recuperación

2.4.1 Topologías de Backup:

Tres topologías básicas son usadas en un entorno de backup:

Backup directamente conectado.

Backup basado en LAN.

Backup basado en SAN.

Backup Mixto.

2.4.1.1 Backup Directamente conectado:

En un “Direct-AttachedBackup”, un dispositivo de backup es directamente agregado al cliente. Solamente la metadata es enviada al servidor de backup a través de la red LAN. En este ejemplo, el cliente también actúa como un storage node que escribe la data en el dispositivo de backup.

2.4.1.2 Backup basado en LAN:

En backup basado en LAN, todos los servidores están conectados a la red LAN y la data a ser respaldado es transferida desde el backup cliente (fuente), al dispositivo de backup (destino) sobre la red LAN, la cual podría afectar la performance de la red LAN. El flujo a través de la red LAN también afecta la performance de todos los sistemas conectados al mismo segmento como el servidor de backup. Los recursos de red son severamente reducidos cuando múltiples clientes acceden y comparten el mismo “tape libraryunit (TLU)”. Este impacto se puede ser minimizado por adoptar un número de medidas, tal como configurar redes separadas para backup e instalando dedicados storages node para algunos servidores de aplicaciones.

2.4.1.3 Backup basado en SAN (LAN libre):

El backup basado en SAN es conocido como LAN libre. La topología de backup basado en SAN es la más apropiada solución cuando un dispositivo de backup necesita ser compartido entre los clientes. En este caso el dispositivo de backup y clientes son agregados a la red SAN.

En este ejemplo, los clientes leen la data desde el servidor de aplicación en la SAN y escriben al dispositivo de backup agregado. El tráfico de data de backup es restringido a la SAN, y la metadata es transportada sobre la LAN. Sin embargo, el volumen de la metadata es insignificante cuando es comparada con la data de producción porque sólo envía referencias o registro de información para el catálogo.

La performance de LAN no es degradada en esta configuración. Por removimiento del cuello de botella de la red LAN, la SAN mejora las copias de seguridad o la performance, porque se libera la red LAN del tráfico de backup.

2.4.1.4 Backup Mixto:

Una topología mixta usa ambas topologías de backup basado en LAN y backup basado en SAN en un mismo entorno. Esta topología podría ser implementada por muchas razones, incluyendo costo, localización del servidor, reducción en la sobrecarga administrativa, y consideraciones de performance.

2.5 Opciones tecnológicas de Backup

Un amplio rango de soluciones tecnológicas es actualmente disponible para la copia de seguridad en tape (cintas) y discos son los dos más comúnmente usado como un backup media. La tecnología de tape ha madurado hasta escalar las demandas empresariales, mientras que la backup a disco está emergiendo como una opción con disponibilidad de bajo costo en disco. “Virtual tape libraries” usa discos como backup emulando tapes, proporcionando las capacidades de backup y recuperación.

2.5.1 Backup a tape o Cinta:

Tapes o Cinta, una tecnología de bajo costo, son usados extensivamente para backup. Tape drives son usados para leer/escribir la data desde/a un “tape cartridge o tape”. Tape drives son referidos como secuencial, o lineal, dispositivo de acceso porque la data es escrita o leída secuencialmente.

Tape drive flujo o multi-flujo escribe la data desde múltiples flujos en un simple tape para mantener el drive ocupado. Múltiples flujos mejora la performance de la media, pero esto tiene asociado una desventaja. El backup de la data es intercalado porque la data desde múltiples flujos es escrita en el tape. Consecuentemente, el tiempo de recuperación de la data es incrementado.

2.5.2 Librería física de cinta:

La librería física de cinta proporciona vivienda y alimentación para un número de “tapes drives” y “tape cartridges”, junto con un brazo robótico o un mecanismo de selector. El software de backup tiene inteligencia para administrar al brazo robótico y todo el proceso de backup. “Tape drives” lee y escribe data desde y hacia un tape. Los tapes son ubicados en los “slots” cuando no está siendo usado por un tape drive. El brazo robótico es usado para mover las cintas dentro de la librería, tal como mover un tape drive dentro de un slot. Otro tipo de slot llamado “mail” o “import/export slot” es usado para adicionar o remover tapes desde la librería sin abrir la puerta de acceso porque abriendo la puerta de acceso causa a la librería ir offline (fuera de servicio). En adición, cada componente físico en una librería tiene una individual “element address” que es usado como un mecanismo de dirección para el movimiento de tapes dentro de la librería. El “tape drive” recibe la data de backup y almacena la data en su “buffer interno” (memoria interna). Esta data de backup es entonces escrito al tape en bloques.

2.5.3 Backup a Disco:

El sistema de backup a disco ofrece una fácil implementación, reducción de costo, y mejoramiento de la calidad de servicio. Realizando una copia en un “storage de disco”

ofrece una clara ventaja debido a su inherente acceso aleatorio y capacidades de protección RAID. En muchos ambientes de backup, el backup a disco es usado como “staging área” donde la data es copiada temporalmente antes de transferir esta información a tape más adelante. Esto mejora la performance del backup.

2.5.4 Comparación de Restauración Cinta vs Disco:

Un escenario de recuperación comparando tape (cinta) vs disco en un entorno de Microsoft Exchange que soporta 800 usuarios con 75MB en tamaño de mailbox y 60GB de base de datos. Una recuperación desde disco tomó 24 minutos comparado la recuperación desde tape (cinta), la cual tomó 108 minutos en el mismo ambiente.

2.5.5 Librería de Cintas Virtual:

Una “virtual tape library (VTL)” o librería virtual, tiene los mismos componentes como una librería robótica física excepto que la mayoría de los componentes son presentados como un recurso virtual. Para el software de backup, no hay diferencia entre una librería robótica física y una librería virtual. La librería virtual usa discos como la media para backup. El software de emulación tiene una base de datos como una lista de tapes virtuales, y cada tape virtual es asignada una porción de una LUN (logical unit number) en un disco. Un tape virtual puede expandirse en múltiples LUNs si es requerido. A diferencia de una librería física, la cual involucra demoras mecánicas, en una librería virtual esto es casi instantánea.

Los procesos de copia de seguridad y recuperación son secuenciales por naturaleza, pero ellos se benefician desde las características del acceso aleatorio del disco porque estos siempre están online y listos para ser usados, mejorando los tiempos de backup y recuperación.

2.6 Desarrollo de las redes SAN

2.6.1 Definición SAN:

Un “Storage Area Network” es una red dedicada de alta velocidad para acceso de datos por bloques. SAN permite una consolidación del storage de disco y permite compartir al storage de disco alrededor de múltiples servidores. SAN proporciona comunicación física de la infraestructura y permite una segura y robusta comunicación entre el host y el dispositivo del storage. FC SAN lleva la data, comandos y estatus de información entre servidores (host) y dispositivos de storage a través de los switches de Fibra Canal (redes llamadas también fabric).

2.6.2 Entendiendo Fibra Canal (FC):

La arquitectura FC forma la principal construcción de FC SAN como infraestructura. Fibra Canal es un tecnología de red de alta velocidad que corre en cables de fibra óptica (optical fiber cables) de grandes velocidades (preferido para front-end en la conectividad de la red SAN) y cable serial de cobre (preferido para back-end en la conectividad del disco). La tecnología FC fue creado para atender la demanda por incremento de la velocidad de transferencia de datos entre servidores, y concentración de dispositivos de storages.

Una alta velocidad de transmisión es una importante característica de la tecnología de red FC. La inicial implementación ofrecida de 100MB/s (equivalente a raw bit rate de 1GB/s es decir 1062,5Mb/s en Fibra Canal), la cual fue mayor que la velocidad de Ultra SCSI (20MB/s) comúnmente usado en entornos DAS. FC en modo full-duplex pudo mantener velocidades de 200MB/s. En comparación con Ultra SCSI, FC es un significativo líder en la redes de almacenamiento. La última implementación de 8Gb/s FC (Fibra Canal) ofrece velocidades de 320MB/s. La arquitectura FC es altamente escalable y teóricamente una simple red FC puede acomodar aproximadamente 15 millones de nodos.

Cuando se observa que en la infraestructura, la red SAN y LAN son mutuamente exclusivos, pero sirve para similares propósitos.

2.6.3 Evolución de las redes SAN:

En una temprana implementación, la SAN fue un simple grupo de hosts y la asociación de storages que estaba conectado a la red usando un “Hub” como un dispositivo de conectividad. Esta configuración de una SAN es conocido como un “Fibra Canal Arbitrated Loop (FC-AL)”. El uso de Hubs daba como resultado islas aisladas de FC-AL, porque los Hub proporcionan una conectividad y ancho de banda limitado. La inherente asociada limitación con los Hubs daba forma para la alta performance “FC switches”. La topología del “switch fabric” proporciona mejor conectividad y performance, la cual permite a la SAN ser altamente escalable. Esto mejora la accesibilidad para aplicaciones alrededor de la empresa. FC-AL ha sido abandonado por la red SAN debido a las limitaciones, pero todavía sirve como una interfaz de disk-drive.

2.6.4 Componentes de la red SAN:

2.6.4.1 Node Ports:

Un “end-to-end SAN” consiste de tres componentes básicos: servidores, infraestructura de red y storages. Estos componentes pueden ser desglosadas en un futuro dentro de los siguientes elementos: node ports, cables, dispositivos de interconexión, tal como FC switch o Hub, arreglo de storage y administración del software para SAN.

En Fibra Canal, los dispositivos tales como host, storages y librerías de cinta todos ellos son referidos como “nodos”. Cada nodo es una fuente o destino de información para uno o más nodos. Cada nodo requiere uno o más puertos para proporcionar una interfaz física para comunicarse con otros nodos. Estos puertos son componentes integrales de un HBA y los adaptadores de los storage “front-end”. Un puerto de operación en full-duplex el modo de transmisión de datos con una línea transmisor (Tx) y una línea receptor (Rx).

2.6.4.2 Cables ópticos:

Las implementaciones de SAN usan cables de fibra óptica u optical fiber cabling. Los cables de fibra óptica puede llevar la data en forma de luz. Hay dos tipos de cables de fibra óptica, multimodo y monomodo.

Multi-mode fiber (MMF) o cable de fibra multimodo lleva múltiples haces de luz proyectados simultáneamente en diferentes ángulos en el centro del cable. Basado en el ancho de banda, las fibras multimodo son clasificados como OM1 (62.5um), OM2 (50um) y láser optimizado OM3 (50um). En una transmisión en fibra multimodo (MMF), múltiples haces de luz viajando dentro del cable tienden a dispersarse y colisionar. Esta colisión debilita la potencia de la señal después de que esta viaja una cierta distancia, este proceso es conocido como dispersión modal. Un MMF de cable es frecuentemente usado a distancias de hasta 500 metros debido a una degradación de la señal por dispersión modal. Single-Mode fiber (SMF) o cable de fibra monomodo lleva un simple rayo de luz proyectado en el centro o núcleo del cable. Estos cables están disponibles en diámetros de 7-11 um; el más común de 9 microns. En una transmisión SMF, una simple señal viaja en una línea directa a través del centro del cable. El pequeño centro y la simple onda de luz limitan la dispersión modal. Casi todos los tipos de cable de fibra óptica, monomodo provee una mínima atenuación de la señal hasta una distancia 10km. Un cable de fibra monomodo es usado para largas distancias, limitado sólo por la potencia del láser en el transmisor y la sensibilidad del receptor.

2.6.4.3 Conectores de Fibra:

Los conectores “Standar Connector (SC)” y un “Lucent Connector (LC)” son dos conectores comúnmente usados en cables de fibra óptica. Un SC es usado para transmisión de datos de hasta 1Gb/s, mientras un LC es usado para velocidades de hasta 8Gb/s. Un “Straight Tip (ST)”, en los primeros días de desarrollo de FC, el cable de fibra óptica predominantemente usaba conector ST.

El Small Form-factor Pluggable (SFP) es un transceptor óptico usado en una comunicación óptica.

2.6.4.4 Dispositivos de interconexión:

“Hub, Switches, y Directores” son dispositivos de interconexión comúnmente usado en SAN. Hubs son usados como dispositivo de comunicación en una implementación FC-AL.

2.6.5 Opciones de interconectividad:

La arquitectura FC soporta tres opciones de interconectividad: Punto a punto, Circuito Arbitrario (FC-AL), y Fabric Connect (FC).

2.6.5.1 Punto a Punto:

Punto a punto es configuración simple de FC, dos dispositivos son conectados directamente a cada uno. Esta configuración provee una conexión dedicada para la transmisión de datos entre nodos. Por lo tanto, la configuración punto a punto ofrece una limitada conectividad, sólo dos dispositivos pueden comunicarse con cada uno al mismo tiempo. El estándar DAS (Direct Attached Storage), usa conectividad punto a punto.

2.6.5.2 FC-AL:

En la configuración FC-AL, los dispositivos son conectados en un circuito compartido (shared loop). FC-AL tiene las características de la topología “Token Ring” y la conexión física de la topología estrella. En FC-AL, cada dispositivo contiene con otro dispositivo para desarrollar operaciones de I/O. Los dispositivos en el circuito debería “arbitrar” para tener el control del circuito. En un tiempo dado, sólo un dispositivo puede desarrollar operaciones de I/O en el circuito.

Por lo tanto, una implementación FC-AL también podría usar Hubs mediante el cual el circuito arbitrario es físicamente conectado en una topología de estrella. La configuración FC-AL tiene las siguientes limitaciones en términos de escalabilidad:

FC-AL comparte el ancho de banda en un circuito. Solamente un dispositivo puede desarrollar operaciones de I/O al mismo tiempo. Porque cada dispositivo en un circuito tiene que esperar un requerimiento de cambio para un proceso de I/O, la velocidad de transmisión de datos es baja en una topología FC-AL.

FC-AL usa una dirección de 8-bit. Esto puede soportar hasta 127 dispositivos en un circuito.

Adicionando o removiendo un dispositivo resulta en una reiniciación del circuito, la cual causa una momentánea pausa en el tráfico del circuito.

2.6.5.3 FC-SW:

A diferencia de la configuración circuito, una “Fibra Canal Switched Fabric (FC-SW)” proporciona dispositivos interconectados, dedicado ancho de banda, y escalabilidad. La adición o removimiento de un dispositivo en una “Switched Fabric” es mínimamente

disruptivo; esto no afecta la continuidad de tráfico entre otros dispositivos. FC-SW es también referido como “Fabric Connect”. Una fabric es un espacio lógico en cual todos los nodos se comunican con otros nodos en una red. Este espacio virtual puede ser creado con un switch o una red de switches. Cada switch en una fabric contiene un único identificador de dominio, el cual es parte del esquema de direccionamiento de la fabric. En FC-SW, los nodos no comparten un circuito; en vez de eso, la data es transferida a través de un camino dedicado entre nodos. Cada puerto en una fabric tiene una única dirección Fibra Canal de 24bits para comunicación.

2.6.6 Tipos de puertos:

Los puertos están basados en construcciones de bloques en una red FC. Los puertos en la red pueden ser uno de los siguientes tipos:

- **N_port:** Un punto final en la fabric. Este puerto es también conocido como el “node port”. Típicamente, esto es como un host port (HBA) o puerto en un arreglo de storage que es conectado a un switch en una fabric.
- **NL_port:** Un nodo que soporta la topología Circuito Arbitrario. Este puerto también es conocido como “node loop port”.
- **E_port:** Un puerto FC que forma la conexión entre dos switches FC. Este puerto es también conocido como la “expansión port”. El E_port en un switch FC conecta el E_port de otro switch FC en la fabric a través de un link, la cual es llamado un “InterSwitch Link (ISL)”. ISL son usados para transferir la data del host al storage, tan bueno como la administración del tráfico desde un switch a otro. ISL es uno de los mecanismos escalables en la conectividad SAN.
- **F_port:** Un puerto en un switch que conecta un N_port. Esto es conocido como un “fabric port” y no puede participar en FC-AL.
- **FL_port:** Un puerto en la fabric que participa en FC-AL. Este puerto es conectado al NL_ports en un FC-AL loop. Un FL_port también conecta un circuito a un switch en un switch fabric. Como un resultado, todos NL_port en el circuito puede participar en un FC-SW. Esta configuración es referido como un “circuito público”. En contraste, un arbitrated circuit sin algún switch es referido como un “circuito privado”. Un circuito privado contiene nodos con NL_port, y no contiene un FL_port.
- **G_port:** Un genérico puerto que puede operar como un E_port o un F_port y determina sus funcionalidades automáticamente durante su inicialización.

2.7 FC Arquitectura y Administración

2.7.1 Características de la Arquitectura FC:

La arquitectura FC (Fibre Channel) representa la verdadera integración canal/red con la interconexión estándar de dispositivos. Conexiones en una SAN son realizados usando FC. Fibra Canal Protocol (FCP) es la implementación de serial SCSI-3 sobre FC network. En la arquitectura FCP, todos los dispositivos externos y remotos de storages son agregados a la SAN y aparece como un dispositivo local para la operación del sistema host. Las principales ventajas de FCP son los siguientes:

- La transmisión sostenida de ancho de banda sobre largas distancias.
- El soporte para un largo número de dispositivos sobre la red.
- Exhibiciones características del canal de transporte y proporciona velocidades de 8Gb/s

2.7.2 Pila del protocolo Fibra Canal:

FCP (Fibra Canal Protocol) define el protocolo de comunicación en 5 capas: FC-0 a través FC-4 (excepto la capa FC-3, la cual no está implementado). En el modelo de comunicación en capas, el par de capas en cada nodo habla con cada otro a través de protocolos definidos.

•**FC-4 Protocolo de Capa Superior:** FC-4 es la capa más alta de la pila de FCP. Esta capa define la interface de aplicación y la forma Upper Layer Protocols (ULPs) son mapeados a la capa más baja de las capas FC. El FC estándar define muchos protocolos que pueden operar en la capa FC-4. Algunos de los protocolos incluye SCSI, HIPPI Framing Protocol, Enterprise Storage Connectivity (ESCON), ATM, e IP.

•**FC-2 Capa de Transporte:** el FC-2 es la capa de transporte contiene la carga útil, dirección de la fuente, el puerto de destino, y el enlace de control de información. La capa FC-2 provee la dirección Fibra Canal, estructura, y organización de la data (frames, sequences, e intercambios). Esto también define los servicios de la fabric, clases de servicio, control de flujo, y direccionamiento.

•**FC-1 Protocolo de Transmisión:** Esta capa define el protocolo de transmisión que incluye la regla de codificación y descodificación serial, caracteres especiales utilizados, y control de error. En el nodo de transmisión un carácter de 8bits es codificado a caracteres de transmisión de 10bit. Este carácter es entonces transmitido para el nodo receptor. En el nodo receptor, el carácter de 10bits es pasado a la capa FC-1, la cual decodifica los 10 bits en el carácter original de 8bits.

●**FC-0 Interfaz Físico:** FC-0 es la capa más baja en la pila FCP. Esta capa define la interface física, medida, y la transmisión de raw bits. La especificación FC-0 incluye cables, conectores y parámetros ópticos y eléctricos para una variedad de tasa de transferencia de datos. La transmisión FC puede usar ambos una media eléctrica y óptica.

2.7.3 Direcciones Fibra Canal:

Fibra Canal Addressing (dirección de Fibra Canal) es dinámicamente asignada durante el inicio de sesión en la fabric. Esto tiene un formato distinto dependiendo del tipo de puerto del nodo.

El formato de dirección depende sobre el tipo de nodo en a fabric: N_port, NL_port en un circuito público, o NL_Port en un circuito privado.

Dirección FC de un N_port: El primer campo contiene el domain ID del switch. Aunque este un campo de 8bits, hay sólo 239 direcciones disponibles para el dominio ID, desde que algunas de esas direcciones están con una consideración especial y reservado para la administración de servicios en la fabric. El área ID es usado para identificar un grupo de F_port. Un ejemplo de este grupo de F_ports sería una tarjeta en el switch con más que un puerto en este. El último campo identifica el F_Port dentro del grupo.

El máximo número posible de N_ports en un switch fabric es de este modo: (239 domains x256 areas x256 ports)= 15 663 104 Fibra Canal addresses.

Dirección FC de un NL Port: Si no existe una conexión de switch a un circuito arbitrario, todos los NL_port están en un circuito privado y los dos bits superiores en su dirección FC son ceros.

Si el circuito es agregado a la fabric y un NL_port soporta un inicio de sesión en la fabric, los dos bits superiores son asignados con un valor positivo llamado el identificador circuito por el switch. Este es llamado circuito público. El identificador de circuito es el mismo para todos los NL_Port en un circuito dado.

En un circuito público o privado, el ultimo byte de 24bits de la dirección del puerto refiere al “arbitrated loop physical address (AL_PA)”. Hay 127 permisibles AL_PA dentro de un circuito. En una fabric-attached o circuito público allí podría ser más que 126 NL_Ports disponible, con una dirección reservada para el FL_Port en el switch.

2.7.4 World Wide Names:

Cada dispositivo en el entorno FC es asignado un único identificador de 64bit llamado el “World Wide Name (WWN)”. El entorno Fibra Canal usa dos tipos de WWNs: World Wide Node Name (WWNM) y World Wide Port Name (WWPN). A diferencia de la dirección FC, el cual es asignado dinámicamente, un WWN es un nombre estático por cada

dispositivo en un FC network. WWNs es similar a la dirección Media Access Control (MAC) usado en IP networking. WWNs son quemados dentro del hardware o asignado a través del software. Muchas definiciones de configuración en una SAN usa WWN para identificar los dispositivos de storages y HBAs. El nombre del servidor en un entorno FC mantiene la asociación de WWNs con la creada dinámicamente dirección FC para los nodos.

2.7.5 Estructura y organización de FC Data:

En una red FC, el transporte de la data es análogo a una conversación entre dos personas, por lo cual un frame representa una palabra, una secuencia representa una sentencia, y un intercambio representa una conversación.

- **Operación de intercambio:** Una operación de intercambio o Exchange operation permite dos puertos N_ports identificar y administrar una serie de unidades de información. Esta unidad mapea a una secuencia. Las secuencias pueden ser ambas direccionales y bidireccionales dependiendo sobre la secuencia cambiada del tipo de data entre el iniciador y el objetivo.

- **Secuencia:** Una secuencia se refiere a un conjunto continuo de frames que son enviados de un puerto a otro. Una secuencia corresponde a una unidad de información, como definido por el ULP (upper layer protocol)

- **Frame:** Un frame es la unidad fundamental de la transferencia de la data en la capa 2. Cada frame contiene hasta 2 112bytes de carga útil.

2.8 EMC Avamar

2.8.1 EMC Avamar: Backup & Recuperación con Global 3D:

EMC presentó (en Julio del 2011) públicamente Avamar Data Store Generación 4 y el software Avamar V6.0, nuevas versiones de los productos de backup y recuperación con Deduplicación global basada en origen de Avamar que son líderes en la industria. Estas versiones ofrecen nuevas funciones de escalabilidad, soporte y simplicidad, y mejoras que permiten a los clientes aumentar la eficacia de los recursos de TI y disminuir el costo total de propiedad de backup y recuperación.

Esta nueva versión incluye:

- **Mayor escalabilidad de backup,** para afrontar los requerimientos de escalabilidad en entornos de gran tamaño y, en el caso de entornos de cualquier tamaño, disminuye el espacio de los data centers para las copias de seguridad, lo que permite ahorrar en costos de espacio físico, energía, refrigeración e inversión.

- **Soporte ampliado para aplicaciones y sistemas operativos del cliente**, para soportar más requerimientos en los entornos del cliente.
- **Nuevo soporte para idiomas internacionales**, a fin de facilitar la implementación en organizaciones globales y para clientes ubicados fuera de América.
- **Simplificación del mantenimiento**, la administración y la capacidad de utilización para ofrecer flexibilidad operativa, una mayor eficacia del sistema y mejores diagnósticos.

Con respecto a las tasas de Deduplicación, EMC Avamar, al deduplicar globalmente en el origen, puede alcanzar tasas de Deduplicación cercanas a 500:1 –según análisis realizados– comparando con los métodos tradicionales de respaldo diario. Este número es calculado basado en los datos almacenados diariamente teniendo en cuenta la metodología de respaldos full de Avamar. Cada respaldo realizado por EMC Avamar es un resguardo full que puede ser recuperado en un solo paso sin necesidad de recuperar una imagen full y luego aplicar los incrementos correspondientes.

EMC Avamar permite alcanzar en algunos casos tasas de 500:1 al evaluar los datos a nivel subfile segmentándolo en pequeños segmentos de longitud variable en el source (cliente). Solamente aquellos segmentos únicos globalmente son transferidos al repositorio central de los respaldos.

A continuación se presenta un ejemplo para clarificar la tasa de Deduplicación.

Se asume:

- 1.76 TB a respaldar.
- 100% archivos de Microsoft Office.
- Retención de 90 días
- Resguardo Full Diario
- No – Tape Compression

Nota: la tasa de data de-duplication y capacidad varía según el tipo de datos y las políticas de retención.

Modelo Avamar:

- Backup Full Inicial deduplica un 70%. Es decir que únicamente un 30% se almacenará en el dispositivo de respaldo de Avamar.
- Backup siguientes (full diarios) deduplican a una tasa cercana al 99.7% siendo entonces solamente un 0.03% será respaldado en Avamar.

– Full inicial = 1.76 TB x 30 percent = 528 GB

– Respaldos siguientes = 1.76 TB x .003 = 5.3 GB

Modelo tradicional:

- Backup full diario, sin compression= 1.76 TB
- Backup full diario siguientes= 1.76 TB

Avamar Full Diarios versus Backup Tradicional Full Diario:

- Avamar backup inicial = 528 GB
- Tradicional inicial full = 1.76 TB
- Avamar backup siguientes = 5.3 GB
- Tradicional fulls diarios = 1.76 TB

Resultado:

La reducción obtenida con Avamar es de aproximadamente 332($1.76 \text{ TB}/5.3 \text{ GB} = 332$).

Luego de 90 días, el resultado sería:

- Avamar: $528 \text{ GB} + 89 * 5.3 \text{ GB} = \sim 1 \text{ TB}$
- Tradicional: $1.76 \text{ TB} + 89 * 1.76 \text{ TB} = \sim 159 \text{ TB}$

En esta caso, la reducción obtenida por EMC Avamar es de 159 ($159 \text{ TB}/1 \text{ TB} = 159$).

2.8.2 Software Avamar y Oficinas Remotas:

EMC Avamar es la plataforma ideal para dar respuesta a la problemática de oficinas remotas. A través de la aplicación de la Deduplicación a nivel global, se permite reducir el uso de ancho de banda en los respaldos y recuperaciones. Es posible encriptar el tráfico y definir un % de uso del ancho de banda a la hora de respaldar y/o recuperar.

EMC Avamar nos permite implementar de tal manera de alcanzar distintos tipos de objetivos de respaldo y recuperación. En tal sentido podrán existir:

- Sitios que realizan sus operaciones de respaldo y recuperación a través de la WAN a un sitio central o regional.
- Sitios que realizan sus operaciones de respaldo y recuperación de manera local y replicando a un sitio central o regional.
- El sitio central o regional podrá a la vez estar replicado a otro sitio.

Todo el tráfico por la WAN se realiza de manera comprimida, deduplicada y encriptado. La Deduplicación es global y de tal manera se optimiza el uso del ancho de banda y el espacio de almacenamiento destinado a los respaldos.

En el caso de sitios con su propio nodo de backup, eventualmente, podrán realizar sus operaciones de respaldo y recuperación sin afectar al resto.

En el caso de una administración centralizada, podrán definirse distintos niveles de jerarquías administrativas para que –por ejemplo- un administrador del sitio 1 no pueda respaldar y/o recuperar datos del sitio 2.

En tal sentido podría generarse un esquema jerárquico a través del uso de dominios y clientes que permita obtener perfiles distintos por cada sitio (tengan o no hardware local para respaldo).

2.8.3 Software Avamar y Usuarios Finales:

EMC Avamar para Desktops y Laptops es un cliente especialmente desarrollado para optimizar el uso de la plataforma Avamar permitiendo realizar respaldos y recuperaciones de ambientes Desktop y Laptops a través de un cliente diseñado para usuarios finales.

2.8.4 Beneficios asociados a EMC Avamar:

A continuación se enumera los beneficios asociados a EMC Avamar:

- De-duplicación en el Origen— antes de transferir la data
- Reduce el impacto en la utilización del vínculo LAN/WAN
- Permite un respaldo rápido, en algunos casos 10 veces más rápido.
- Proporciona un Web-Based Management Centralizado
- Simplifica las tareas de recuperación.
- Elimina la necesidad de transporte de Cintas fuera de sitio
- La Data es deduplicada, comprimida y encriptado.
- Reduce el almacenamiento de respaldo hasta en 50 veces según el ambiente.
- Provee solución de replicación / DR en ambientes distribuidos.

2.9 EMC Networker

La familia Networker de EMC es la solución de backup y recuperación más flexible y rápida de la industria. Networker proporciona un rendimiento sin precedentes, así como también soporte para las últimas tecnologías de backup en disco y copia instantánea. Networker protege los datos críticos para el negocio al centralizar, simplificar y acelerar las operaciones de backup y recuperación en entornos heterogéneos. Basado en una arquitectura abierta y altamente escalable, reduce la sobre carga de administración al automatizar la protección de los activos de almacenamiento de información en entornos de almacenamiento de información con conexión directa (DAS, Direct Attached Storage), entornos de almacenamiento de información conectado en red (NAS, Network Attached Storage), y entornos de almacenamiento de información en red (SAN, Storage Area Network), desde los data centers empresariales más grandes hasta las sucursales secundarias más pequeñas. Networker ofrece una poderosa integración con las aplicaciones

a través de módulos opcionales para aplicaciones como Oracle; Microsoft Exchange y SQL Server; Documentum de EMC; IBM DB2, Informix y Lotus Notes/Domino; SAP R/3 en Oracle y Sybase.

Networker de EMC protege los datos al simplificar y centralizar las operaciones de backup y recuperación. Con la performance sin precedentes, Networker de EMC es la alternativa ideal para oficinas pequeñas hasta grandes data centers.

2.9.1 Administración y Seguridad:

Arquitectura de EMC Networker

Networker es una solución de software del tipo cliente-servidor (que incluye una interfaz web-enabled de administración) que provee la habilidad de gestionar las operaciones de backup & recuperación permitiendo la protección de los datos de la empresa de manera centralizada para las distintas plataformas soportadas (UNIX, Windows, VMWare, Linux; según matriz de compatibilidad EMC Information Protection Software Compatibility Guide).

A continuación se menciona los componentes más importantes de Networker y su función dentro del ambiente EMC Networker.

Console Server: Todos los servers Networker y clientes son administrados desde la consola central de EMC Networker. Esta consola permite realizar las tareas de reporting y monitoreo para todos los clientes y servers Networker.

Web Browser: El Console server es accedido a través de una interfaz gráfica que puede ser ejecutada desde cualquier equipo que tenga un web browser y Java Runtime Environment (JRE) 1.4.2 o superior. Múltiples usuarios pueden acceder de manera simultánea a la consola. Más adelante se detalla el manejo de usuarios desde la perspectiva de la seguridad.

Networker Server: Provee servicios de backup y recuperación para los clientes Networker que integran un datazone.

Datazone: Un datazone está compuesto por un único Networker Server y todos sus clientes. Es posible contar con varios datazones en función de los requerimientos de backup.

Storage Node: Un Networker Storage Node puede ser utilizado para mejorar el rendimiento realizando la liberación de carga del Networker Server de la data involucrada en procesos de backup & recuperación. Generalmente se utilizan para la realización de

respaldo SAN para reducir el impacto en el tráfico de red LAN utilizando dispositivos de respaldo VTL o Librerías Físicas conectadas a una SAN.

Networker Cliente: Equipo (servidor) cuya data debe ser resguardada. El Networker Server Console, el o los Networker Servers y los Storage Nodes existentes son además, clientes de Networker.

2.9.2 Backup & Recuperación:

EMC Networker tiene la capacidad de realizar respaldos por LAN, SAN, NAS utilizando las licencias de “Networker Client Connection”, Storage Node y NDMP según corresponda o la nueva modalidad de licenciamiento por capacidad. Más detalles se desarrollan en el manual de Administración de Networker.

En el caso que un cliente sea un storage node, tiene la responsabilidad de realizar el respaldo utilizando el dispositivo de respaldo conectado de forma directa.

La SAN (Storage Area Network) es valorada cuando uno quiere compartir una TLU. Durante el respaldo, los clientes leen la data desde la SAN y escriben a una unidad SAN-Attached. Los clientes pueden compartir un dispositivo de respaldo. Lo único que es enviado por la LAN es información de metadata. Los datos nunca salen del ambiente SAN. EMC Networker permite tener más de una librería/VLT en el Networker DataZone y realizar los respaldos a través de la SAN minimizando el tráfico en la LAN. De esta manera, por la LAN solamente será transferida la metadata de cada respaldo que constituirá el catálogo de Networker para lograr esta funcionalidad se utilizan Storage Nodes. Además los Storage Nodes permiten hacer una liberación de la carga del Backup Server incluso manteniendo el catálogo centralizado.

Backup & Recuperación – Respaldos a Disco & Cinta

Networker soporta respaldos tanto en disco como en cinta, ya sea esto como una librería robótica o como una librería virtual basada en disco.

Las características de cada uno, como las alternativas de staging y clonning que se encuentran detallados en el Manual de Administración de Networker.

Backup & Recuperación – Scheduling

Networker provee la capacidad de realizar respaldos utilizando el “scheduler de Networker” como de manera manual aunque son preferidos los primeros.

Además del manejo de Scheduler, EMC Networker tiene la capacidad de definir prioridades en los clientes.

Backup & Recuperación – Recuperación

Networker permite realizar recuperaciones a través de selección de archivo, job incluyendo la posibilidad de realizar una recuperación utilizando como destino un cliente distinto al que se le tomó el respaldo.

Al realizar las recuperaciones el administrador podrá elegir a partir de qué respaldo (fecha, grupo, equipo) iniciará la recuperación.

Backup & Recuperación – Cloning

El clonado permite generar volúmenes exactamente iguales para un resguardo “offside” (Fuera del sitio). El clonado puede realizarse a nivel volumen como save sets.

El clonado puede ejecutarse de manera automática como parte de una actividad de respaldo.

Es posible, además, establecer una política de retención distinta para el volumen clonado.

Se puede realizar un clonado a otro dispositivo de almacenamiento utilizando la funcionalidad “Directing Clones to a Special Storage Node” (funcionalidad de networker). También puede realizarse sobre la misma librería siempre y cuando exista al menos 1 drive de lectura y 1 drive de escritura.

En el Manual de Administración de Networker se detalla el manejo de clones.

Backup & Recuperación – Staging

Con respecto al manejo de los discos, Networker también soporta la capacidad de realizar “staging” es decir, guardar de manera temporal en disco para luego enviar a cinta. Una vez que la data ha sido escrita en cinta puede liberarse automáticamente el espacio temporal en disco.

Un respaldo puede ser “staged” de un disco a otro las veces que se requiera incluso generando una especie de cadena, es decir de disk 1, a disk 2, a disk 3 y finalmente a un dispositivo basado en cinta. Una vez en la cinta, el medio puede ser clonado.

El proceso de Staging puede ser llevado a cabo de la siguiente manera:

- Calendar-based: mantener el respaldo por 30 días en el dispositivo de staging antes de enviarlo al dispositivo siguiente.
- Event-based: cuando el espacio disponible en el “pool de staging” alcanza un determinado umbral.
- Administrator-based: El proceso de Staging no afecta la política de retención de los datos.

Mayor detalle podrá encontrarse en el manual de Administración de Networker.

Backup & Recuperación – Compresión

El cliente Networker tiene la capacidad de realizar compresión de los datos antes de ser enviados al dispositivo de respaldo (LAN/SAN). Es importante destacar que hay que analizar la naturaleza de los datos a los efectos de establecer cuan comprimibles pueda ser y también analizar las características del dispositivo de respaldo para entender si el mismo tiene capacidad de compresión o no y de esa manera determinar si es más conveniente hacerlo en el dispositivo de respaldo o en el cliente.

Backup & Recuperación – Deduplicación

Data de-duplication (3D) es una nueva tecnología que permite mejorar el uso de las inversiones en soluciones de respaldo y recuperación a través de la utilización de algoritmos especializados en la detección de información duplicada, comparación e indexación de la misma.

Para el caso de Deduplicación en origen, EMC Networker integra a EMC Avamar a los efectos de brindar una experiencia de administración gestionada desde EMC Networker GUI.

EMC Networker incorpora los algoritmos de Deduplicación en la licencia Networker Deduplication Nodes ideal para el respaldo de oficinas remotas y ambiente VMWare. El dispositivo de respaldo en este caso es el Avamar Grid manejado por Networker.

En el caso de Deduplicación en el destino, EMC Networker puede integrarse con EMC Data Domain a partir del uso de EMC Data Domain boost para ambientes de respaldo por LAN (CIFS/NFS).

DDBoost es una pieza de software que permite distribuir parte de los procesos de Deduplicación de Data Domain en el servidor de backup.

En tal sentido, DDBoost proporciona:

- Distribución del procesamiento de los segmentos, lo que significa que el backup server solamente envía datos únicos al Data Domain.
- Gestión en Replicación, permite que el backup server controle la replicación de Data Domain.
- Load Balancing y link failover, proporciona link aggregation para el balanceo de carga y link failover para tener alta disponibilidad.

DDBoost puede mejorar la performance a partir de la reducción de la cantidad de datos a ser transferidos al Data Domain lo que además reduce la carga de trabajo en el equipo Data Domain. Adicionalmente, se cuenta con funcionalidades de clonado que se implementan

desde el Networker Management Console que permiten manejar la réplica y la clonación de volúmenes almacenados en Data Domain.

Backup & Recuperación – Integración con Tecnologías de Snap&Clone

EMC Networker cuenta con la posibilidad de integrarse con los dispositivos de almacenamiento masivo EMC Symmetrix, Clariion entre otros según matriz de interoperabilidad ya sea a partir de la utilización de scripting que realice la ejecución de un snap/clone y su posterior montada en el equipo que corresponda o bien la integración directa con herramientas como Time Finder.

Backup & Recuperación – Módulos de Respaldo Online

A continuación se resume las funcionalidades de respaldo online para las aplicaciones críticas estos respaldos se realizan sin interrumpir el normal funcionamiento de las aplicaciones:

- SQL Server.
- SAP SQL.
- SYBASE.
- Aplicaciones de respaldo Snapshot.
- PowerSnap.
- SnapImage.
- Integración con VSS (volumen shado copy service) de Windows.
- Documentum.
- Informix.
- DB2.
- MEDITECH.
- Lotus.
- Oracle.
- Oracle SAP.
- Exchange.

En la siguiente página (figura 2.3) se muestra el despliegue de EMC Networker.

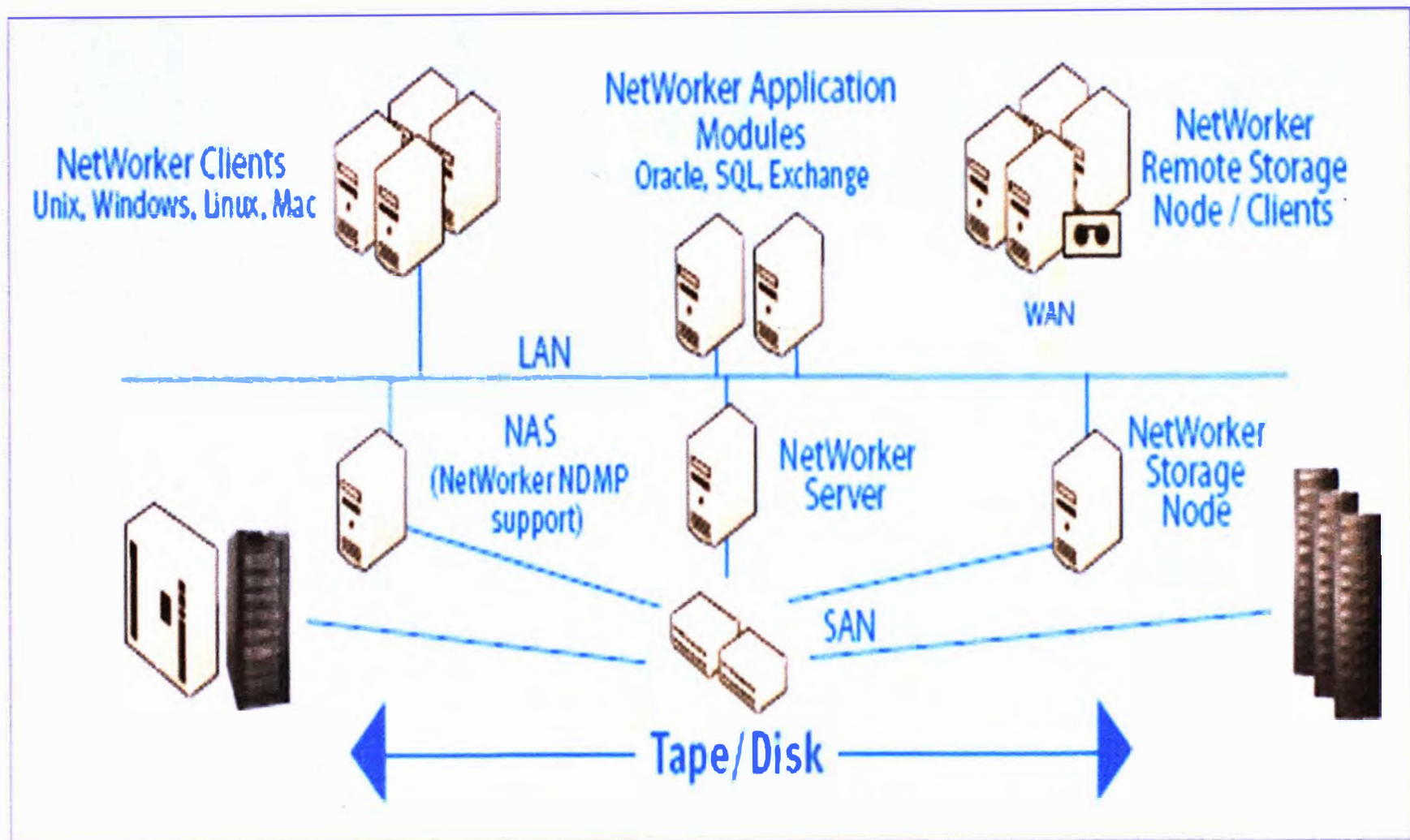


Figura 2.3. Despliegue de EMC NetWorker

2.10 EMC Data Domain

EMC Data Domain es una familia de productos diseñados para mejorar los tiempos de copia de seguridad y recuperación, disponibilidad y mejorar los niveles de servicios sobre los sistemas basados en soluciones de Cintas.

EMC Data Domain utiliza técnicas de Deduplicación en línea que permite optimizar drásticamente el ambiente de respaldo.

No es necesario cambiar el software existente de backup, hardware o procesos de operación ya que EMC Data Domain se encuentra homologado por los principales proveedores de estas tecnologías.

Diseñados para simplificar el ambiente, los equipos EMC Data Domain pueden ser instalados en horas y pueden ser integrados con el ambiente existente de backup. En tal sentido, EMC Data Domain puede ser integrado con EMC NetWorker, Symantec NetBackup, TSM, entre otros.

La técnica de Deduplicación empleada se basa en el uso de bloques de longitud variable, lo que permite alcanzar tasas de Deduplicación muy elevadas dado que existe una flexibilidad adicional a partir del uso de bloques de longitud variable para poder realizar las comparaciones involucradas en el proceso de Deduplicación.

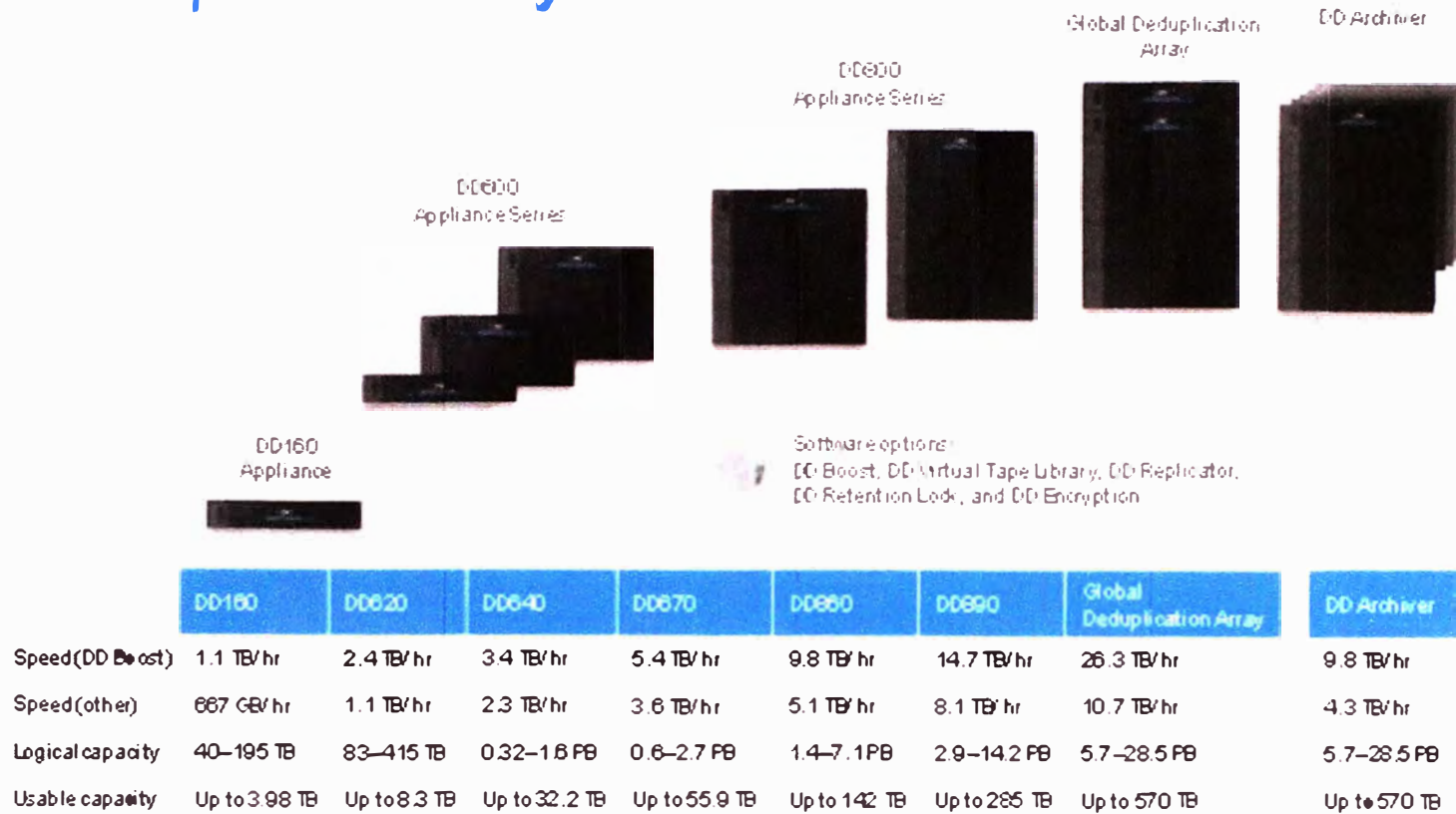
Con respecto al empleo de la Deduplicación, la misma se realiza a partir del uso de la potencia de los procesadores embebidos en los equipos EMC Data Domain. En las últimas dos décadas, el rendimiento de los CPUs se ha incrementado en más de 2.000.000x mientras que en el mismo período el rendimiento de los discos ha crecido en 11x según estudios realizados por Seagate. (*Seagate Technology Paper, Economies of Capacity and Speed, May 2004).

Es importante destacar que los equipos EMC Data Domain son capaces de replicar bajo ambiente IP a los efectos de brindar una solución de respaldo y recuperación de fin a fin.

También es importante destacar que los equipos EMC Data Domain son capaces de verificar que los backups guardados en el equipo estén consistentes y si existe un archivo dañado Data Domain pueda recrear esos bits dañados de tal forma que siempre mantenga consistente la información al momento de realizar una restauración; así como también realizar procesos de mantenimiento como el borrado de los backups que han expirado de tal forma no sobrescribir la información al momento de guardar los backups.

En la página (figura 2.4) se muestra la familia EMC Data Domain.

Industry's Most Scalable Inline Deduplication Systems



© Copyright 2011 EMC Corporation. All rights reserved.

Figura 2.4. Família de EMC DataDomain

EMC Data Domain permite obtener los siguientes beneficios:

- La Deduplicación en línea reduce drásticamente el espacio requerido para respaldos.
- La arquitectura de EMC Data Domain se basa en el uso de sus propios CPUs y la utilización de tecnología SATA en tal sentido, se reduce el costo asociado a infraestructura FC.
- Variedad en la Conectividad: NFS, CIFS, NDMP, o FC/VTL.
- Mejora en los tiempos de Recuperación: Los archivos desde los discos pueden ser recuperados en segundos o minutos vs horas que realiza la recuperación desde las cintas.
- Mejora en tiempos de Backup: Completar las copias durante las ventanas asignadas puede ser complejo. EMC Data Domain puede mejorar los tiempos de backup, dependiendo de la configuración y el ambiente de respaldo.
- Performance: Múltiples data streams pueden ejecutarse de forma simultánea mejorando la performance.
- Capacidades de disaster recuperación utilizando WAN.
- Infraestructura del tipo “plug & play” (conecta y funciona); integración con plataforma de backup existente.
- No hay reemplazos ni nuevos agentes para implementar.
- Posibilita la compra de menor cantidad de tapes. Esto reduce el costo de administración y mantenimiento de cintas.
- Brinda la posibilidad de adquisición de librerías más chicas y/o de extender el uso de la librería actual.
- Reducción de los costos asociados al ancho de banda requerido para la replicación dado que la replicación se realiza de manera deduplicada y en formato encriptado.

CAPITULO III.

DESARROLLO DE LA SOLUCIÓN DE BACKUP PARA SERVIDORES EN EL DATA CENTER PRINCIPAL Y SERVIDORES DISTRIBUIDOS REMOTOS MEDIANTE EL USO CONVENIENTE DE LAS REDES SAN Y REDES LAN/WAN

3.1 Backup de servidores locales, Data Domain con Deduplicación en el destino

3.1.1 Análisis de datos:

De acuerdo al análisis desarrollado anteriormente en la sección 1.3.1 “Evaluación del problema de los servidores distribuidos localmente” los servidores clientes Networker con grandes volúmenes de datos (26TB, 10TB, etc.) están localmente conectados a la red LAN y están siendo respaldados por el software de backup Networker, además están homologados por la matriz de compatibilidad de Networker, como lo descrito en el análisis del punto 1.3.1 se puede liberar carga de datos en la red LAN y liberar recursos del servidor Networker instalando un Storage Node todos en tres servidores (26TB, 10TB y 4.8TB) clientes Networker y conectarlos a la red SAN para luego puedan realizar backup a través de la red SAN y enviar la metadata por la red LAN a efectos de nutrir el catálogo.

Como el sistema de backup a disco ofrece un mejoramiento en la calidad del servicio y ventajas de performance y con tecnología de Deduplicación, se desarrollará copias incrementales con 7 días de retención y las copias full con retención de 4 semanas al Data Domain y los backup históricos (mayor a 4 semanas de retención) a la librería de cintas.

Actualmente desde la topología de backup se puede observar el despliegue de la solución desde una perspectiva más general para poder analizar el total de la distribución en la red de Backup en segmentos y de este modo analizar estos segmentos proponiendo por cada uno de ellos soluciones de backup para luego integrarlos en una solución de hardware y software basado en las soluciones que proporciona EMC.

En la figura 3.1 se ilustra la información recogida de la organización que representa la situación actual.

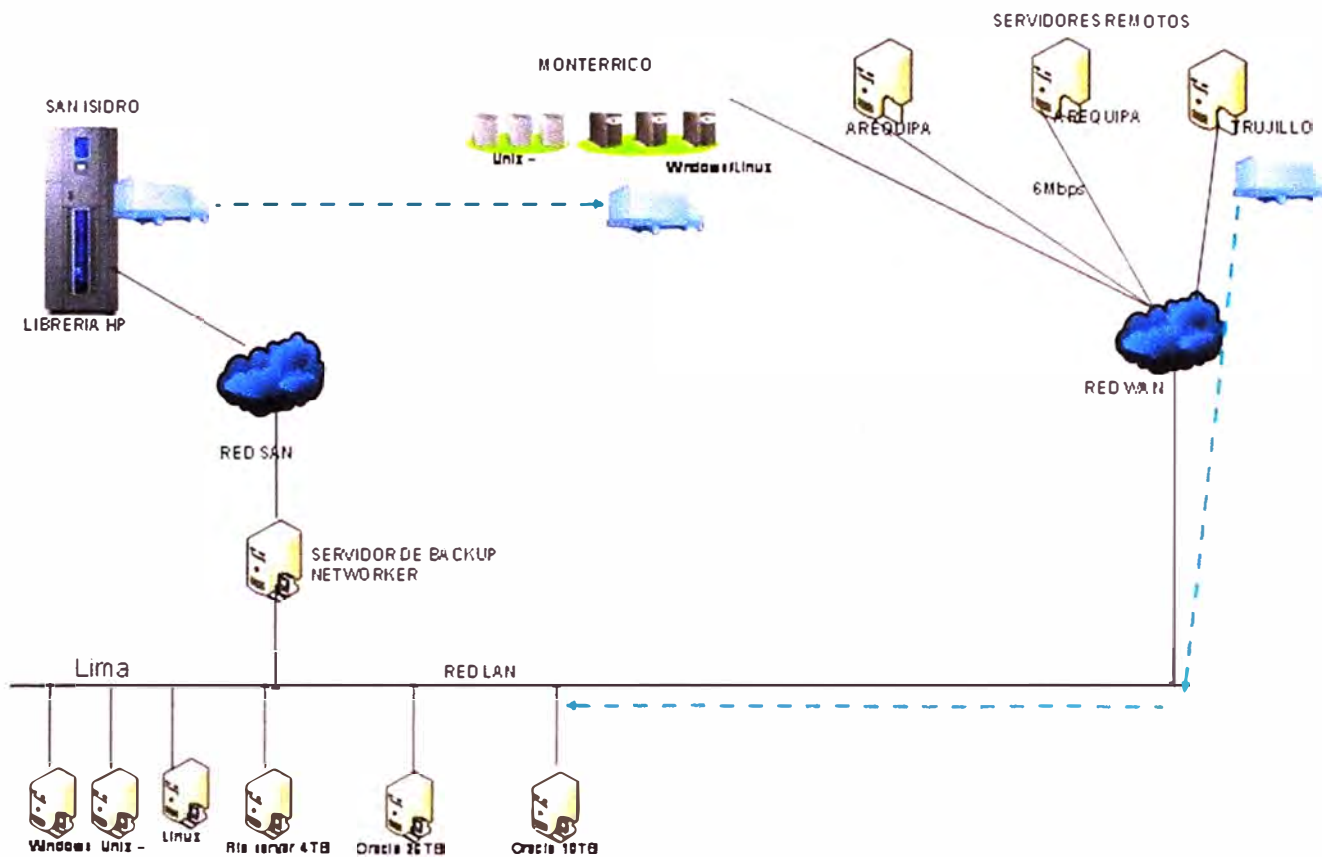


Figura 3.1: Esquema de la Situación actual de la organización

3.1.2 Consideraciones del sizing Data Domain:

Se ha tenido las siguientes consideraciones en función de la tabla 1.2 donde muestra una distribución de servidores en cuanto al dimensionamiento de la solución final:

Organización:

- Herramienta de respaldo: Netwoker Power Edition
- Tipo de respaldo: FULL+ incrementales diarios
- Capacidades a respaldar: Se considera cada tipo de datos
- Tipo de respaldo: Respaldo tradicional sin Deduplicacion

a) Imágenes y Video:

- 6.2TB full semanal
- 310GB de Incremental
- Retención full 4 semanas
- Retención Incremental 7 días

b) Base de Datos Oracle.

- 55TB full
- 2.75TB de Incremental
- Retención full 4 semanas
- Retención Incremental 7 días

c) Base de Datos SQL.

- 496GB full
- 24.8GB de Incremental
- Retención full 4 semanas
- Retención Incremental 7 días

d) EXCHANGE.

- 340GB full
- 17GB de Incremental
- Retención full 4 semanas
- Retención Incremental 7 días

e) FILES SYSTEM.

- 12TB full
- 600GB de Incremental
- Retención full 4 semanas
- Retención Incremental 7 días

Análisis de Capacidad:

A los efectos de calcular el espacio requerido para una solución sin Deduplicación, se presenta:

- Respaldos Fulls ,sin considerar Deduplicación se estima:

4 copias de 6.2TB (Imágenes, video) =	24.8TB
4 copias de 55TB (DB Oracle) =	220TB
4 copias de 496GB (DB SQL) =	1.984TB
4 copias de 340GB (EXCHANGE) =	1.36TB
4 copias de 12TB (FILES SYSTEM) =	48TB
Total requerido:	294.784TB

- Respaldos Incrementales, sin considerar Deduplicación se estima:

7 copias de 310GB (Imágenes, video) =	2.17TB
7 copias de 2.75TB (DB Oracle) =	19.25TB
7 copias de 24.8GB (DB SQL) =	0.1736TB
7 copias de 17GB (EXCHANGE) =	0.119TB
7 copias de 600GB (FILE SYSTEM) =	4.2TB
Total requerido:	25.913TB

En tal sentido, asumiendo las 4 versiones Fulls y 7 versiones Incrementales a respaldar por cada tipo de datos, se requiere aproximadamente 320.7TB netos para respaldar las capacidades detalladas por las retenciones especificadas. En este cálculo no se ha considerado compresión en el momento de la copia de seguridad.

En el caso de la solución de EMC basada en Data Domain, el sizing contempla una tasa de reducción del espacio requerido, se consideran los ratios obtenidos en ambientes similares con Data Domain.

A partir de estos parámetros, se ha calculado que un equipo Data Domain con un total sumado aproximado de 97TB alcanza para respaldar los volúmenes detallados y retenerlo según se ha especificado y permitirá operar durante 3 años a una tasa de crecimiento esperada de 10% sin necesidad de upgrade de discos o licenciamiento de software adicional. En la figura 3.2 se observa el resultado del sizing de Data Domain.

Results

SOLUCION DE BACKUP

Data Domain System

BACKUP DISK DATA DOMAIN

Notes

Total Capacity	96860.06 GB
Replication Target Total	96860.06 GB
Based on Backup Window	5067.86 GB/hr
Predicted Reduction	2.24:1 (55.32%)

Figura 3.2: Resultado del sizing DataDomain

Con respecto al rendimiento o a la capacidad de procesamiento, es importante mencionar que el rendimiento de un sistema de respaldos es sensible a los flujos de datos a través de los distintos componentes, tales como la red, los recursos disponibles en los clientes a ser respaldados, configuración de NICs, HBAs, switches, parámetros de sistema operativo, CPU, rendimiento del subsistema de almacenamiento, las tasas de transferencia de las unidades de cinta, etc.

- Los servidores clientes Networker homologados de grandes volúmenes de datos (26TB, 10TB y 4.8TB) son ahora storage node con base de datos Oracle conectados a la red SAN, y el flujo de datos de backup viajan a través de la SAN y la metadata viaja a través de la red LAN, esto permite claramente mejorar la performance del servidor Networker, la red LAN y reducir los tiempos de backup.
- Las copias de seguridad son almacenados en Data Domain en formato deduplicado y se obtiene los respaldos en línea según las retenciones especificadas 7 días las copias incrementales y 4 semanas los backup Full.
- Los backup con mayores retenciones o históricos (mayor a 4 semanas de retención) serán enviados a la librería robótica basado en cintas, el caso de los storages node podrán enviar las copias directamente a través de la SAN, para luego poder transportar las cintas al sitio de Monterrico por temas de recuperación de desastres.
- Ahora se podrá tener el backup del catálogo Networker server en el Data Domain según las retenciones especificadas y reducir considerablemente el RTO en el servidor Networker.

En la figura 3.3 se ilustra el diagrama topológico de la solución de Backup con Data Domain:

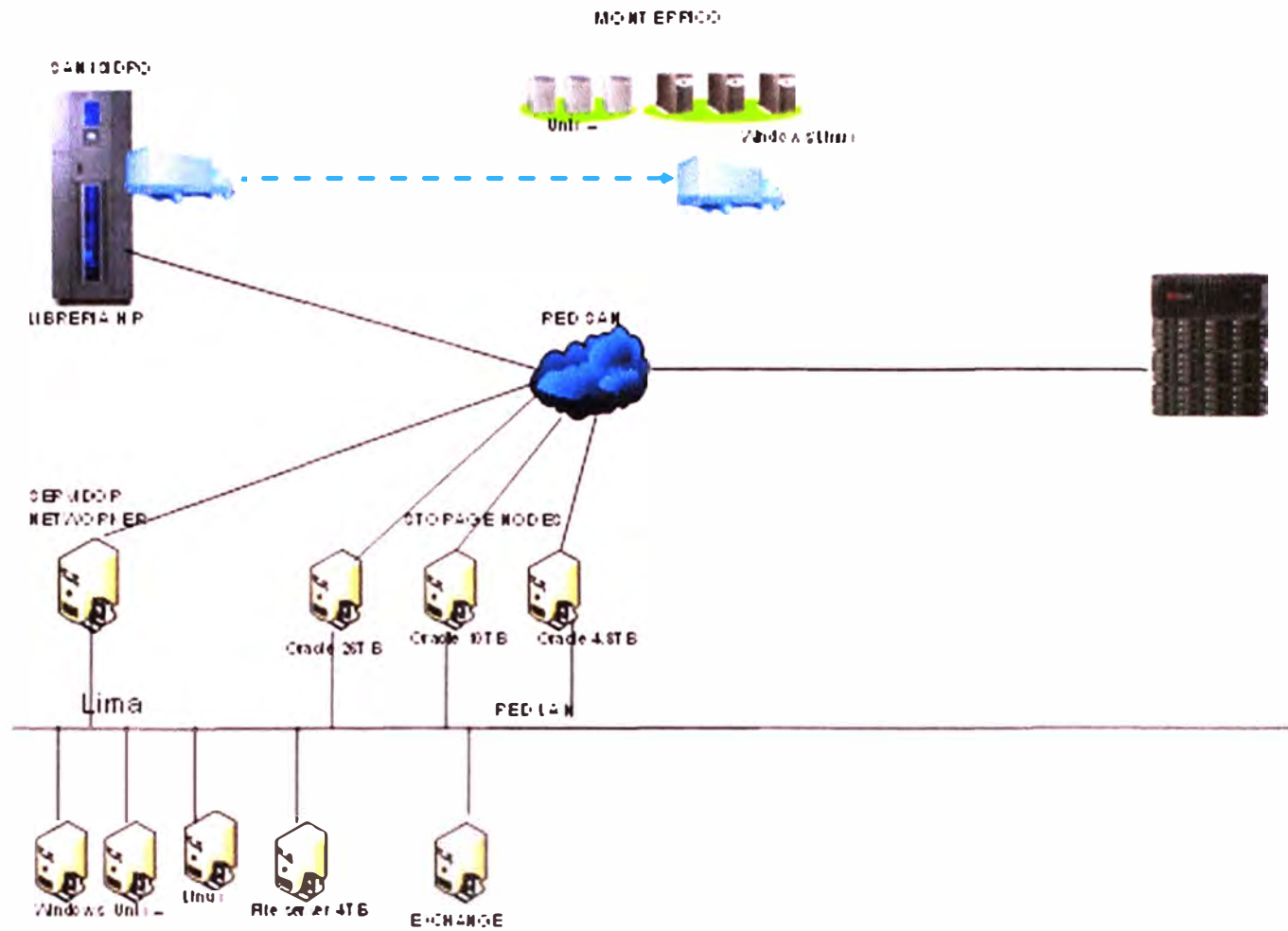


Figura 3.3: Esquema de solución con Data Domain

3.2 Backup de servidores remotos, Avamar con Deduplicación en el origen

3.2.1 Análisis de datos:

De acuerdo al análisis desarrollado anteriormente en la sección 1.3.2 “Evaluación del problema de los servidores distribuidos remotamente” los servidores clientes Networker distribuidos remotamente y homologados por la matriz de compatibilidad de Networker, están remotamente conectados al data center principal a través de la red WAN con un enlace de 6Mbps para el sitio de Trujillo y Arequipa por lo que estos servidores distribuidos están siendo respaldados manualmente en disco por una persona responsable para luego transportar la media o disco al data center principal ubicado en Lima a través de un contrato de terceros que no tiene nada que ver con backups y por consecuencia no sabe lo importante que es esta información para la organización; como lo descrito en el análisis del punto 1.3.2 con Avamar puede realizar backup de estos servidores remotos directamente desde los servidores al data center principal a través de la red WAN si temor de saturar la red WAN de manera segura, confiable y centralizada, esto evita el transporte de la media desde Trujillo o Arequipa hacia Lima y evita estar susceptibles a los inconvenientes que esto implica.

Avamar además cuenta con protección de su propia data es decir puede proteger los propios backup realizados, hace un escaneo de la información en caso de que algunos bits se encuentren dañados este puede reconstruir automáticamente estos bits dañados de tal forma de obtener siempre los backup consistentes a diferencia de una solución tradicional en la cual el software de backup puede terminar exitosamente pero luego que se realiza el traslado del disco nadie puede garantizar al 100% que la información no sufrió algún daño o alteración de la información.

Avamar permite enviar alertas SNMP al administrador de backup en caso de falla de disco o algún error o alteración de la información respaldada, así como también la protección a nivel de su propio catálogo es decir copias del catálogo.

Como el sistema de backup basado en Avamar con Deduplicación de datos en el origen se puede integrar a Networker y manejar las copias de seguridad desde una sola consola de administración permitiendo notables mejoras en la calidad del servicio y ventajas con tecnología de Deduplicación, se desarrollará copias full todos los días con las retenciones de 7 días de retención para los backup diarios y 4 semanas de retención para las copias semanales y 12 meses de retención para las copias mensuales.

En la figura 3.4 se ilustra la topología actual del cliente.

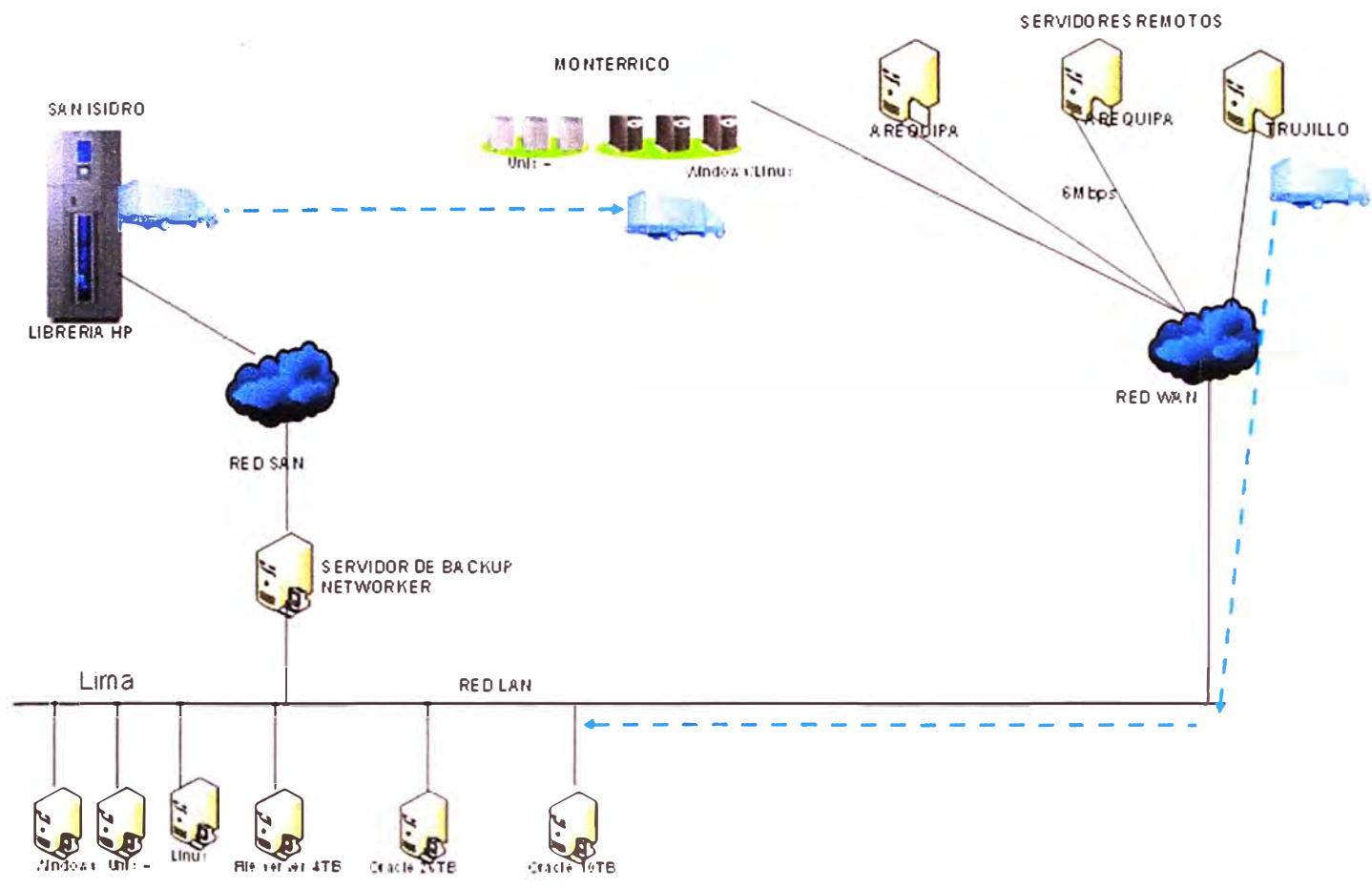


Figura 3.4: Topología actual del cliente

3.2.2 Consideraciones de sizing Avamar:

Se ha tenido las siguientes consideraciones en función a la tabla 1.3 donde muestra una distribución de servidores en cuanto al dimensionamiento de la solución final:

Se presenta el sizing en base a la información relevada en la organización

Arequipa:

- 180GB Full
- Retención diaria 7 días
- Retención semanal 4 semanas
- Retención mensual 12 meses
- Crecimiento anual 50%

Trujillo:

- 188GB Full
- Retención diaria 7 días
- Retención semanal 4 semanas
- Retención mensual 12 meses
- Crecimiento anual 50%

En tal sentido, realizando Full copias diarios, por cada tipo de datos, se requiere aproximadamente de 17TB netos para respaldar las capacidades detalladas por las retenciones especificadas; y tomando en consideración el crecimiento de la data anual de 50% durante tres años se necesitará aproximadamente 58TB netos para respaldar las capacidades detalladas con las retenciones especificadas sin Deduplicación.

En el caso de nuestra solución de EMC basada en Avamar, el sizing contempla una tasa de reducción del espacio requerido basado en el tipo de datos a ser respaldado y la cantidad de versiones a retener en línea, se consideran los ratios obtenidos en ambientes similares con Avamar.

A partir de estos parámetros, se ha calculado un equipo Avamar con un total sumado aproximado de 2TB que alcanza para respaldar los volúmenes detallados y retenerlo según se ha especificado y permitirá operar durante 3 años a una tasa de crecimiento de 50% sin necesidad de upgrade de discos. En la figura 3.5 se muestra el resultado del sizing Avamar.

Results	
SOLUCION DE BACKUP	Avamar
<u>BACKUP DISK AVAMAR</u>	
	Notes
Total Capacity	2087.42 GB
Replication Target Total	2087.42 GB
Based on Backup Window	2.74 GB/hr
Predicted Reduction	96.42% (27.96:1)

Figura 3.5: Resultado del sizing Avamar

3.2.3 Diagrama topológico de backup con Avamar:

- Una vez comprobada la homologación de los servidores remotos por Avamar, se instala agentes Avamar en cada servidor remoto que necesita ser respaldado y se registra en el Avamar single node.
- Antes de enviar los backup por la red WAN los agentes Avamar conversan con el catálogo y sólo se envían los datos únicos y los duplicados serán reemplazados por referencias.
- Los servidores clientes distribuidos remotamente ahora son agentes Avamar, y el flujo de datos de backup viajan a través de la WAN de manera deduplicada y encriptado, guardando estos datos deduplicados en Avamar y la metadata viaja a través de la red LAN y se guarda en el catálogo del Networker.
- Luego los backup podrán ser replicados en forma encriptada y deduplicada vía IP hacia el Avamar single Node ubicado en Monterrico de tal forma que sólo los datos que no se repiten sean replicados y los datos duplicados sólo enviar referencias.
- Todo este proceso se controla desde una misma consola de administración de Networker y permite claramente tener las copias centralizadas.
- Las copias son almacenados en Avamar en formato deduplicado y permite obtener los respaldos en línea según las retenciones especificadas, 7 días las copias diarias, 4 semanas los backup semanales y 12 meses los backup mensuales.
- Ahora se podrá tener las copias de los lugares remotos sin necesidad del transporte físico de la data.

En la siguiente página (figura 3.6) se muestra el esquema de solución con EMC Avamar (solución backup con Deduplicación en el origen):

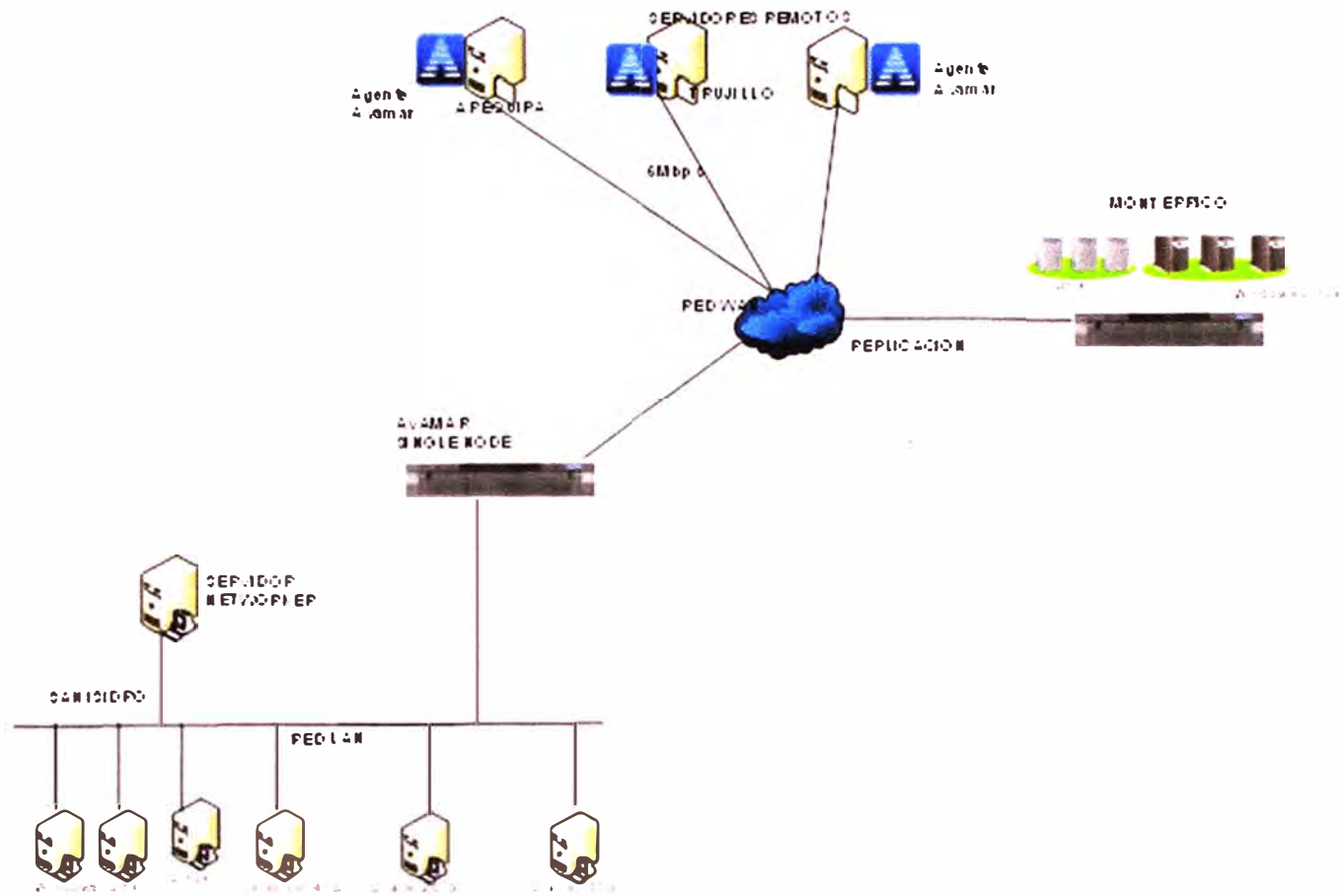


Figura 3.6.: Esquema de solución con EMC Avamar

3.3 Solución EMC Networker

En función a la información desarrollada en los puntos 3.1 y 3.2 se está considerando servidores storages nodes que estarán conectados a la SAN y compartirán los drives de la librería robótica, los backup al Data Domain y la integración de los agentes remotos Avamar con Networker para todo el ambiente de backup en la organización; por tal motivo se ha considerado obtener licencias de los siguientes puntos:

- 3 storages Nodes for Unix.
- 3 Dynamic Drive Sharing.
- Licensing bundle for Data Domain (activate, 10TB DBO and 10TB VTL)
- Licensing Networker for Data Domain VTL 87TB.
- Networker client for data Deduplication cantidad 3.

3.4 Solución EMC Data Domain

Se propone contar una solución de Data Domain DD860 que permitirá hacer upgrade en función al crecimiento de la data en la organización a un ritmo de 10% anual durante 3 años sin necesidad de hacer upgrade.

En tal sentido, se propone el siguiente escenario:

- Un EMC Data Domain DD860
- 9 Shelves de Expansión de 15TB
- Capacidad RAW:135TB
- TB NETO: Aproximadamente 99TB Protección RAID 6 + Spare Globales
- Dos placas HBA Dual Port 8Gbps
- Una NIC quad-port 1Gbps
- Licencia de VTL Option
- Un RACK de 40U

3.5 Solución EMC Avamar

Dada la flexibilidad de Avamar, se presenta un escenario de EMC Avamar que permite alcanzar a organización, tener las copias centralizados para todos los ambientes Remotos en función de las necesidades de respaldo y recuperación.

Esta solución permitirá cubrir las necesidades de crecimiento al 50% anual por los próximos 3 años en la organización sin necesidad de realizar upgrade.

En función al requerimiento de configuración a continuación se detalla los componentes asociados a dicho requerimiento.

La solución propuesta contempla dos single nodes Avamar cada uno detallado como:

- 2 (dos) nodos EMC Avamar de 2.6TB
- Licencia de 2TB Avamar capacity
- Licencia de 2TB Avamar replicación.
- No requiere licencias de clientes:
 - File System
 - Plug-ins de Aplicaciones (Oracle, Exchange, SQL, etc.)

Para el respaldo de los files server de los lugares remotos se propone un single node Avamar de 2.6TB en San Isidro que permitirá tener los respaldos de los servidores remotos y un single node Avamar de 2.6TB en Monterrico que permitirá tener la data que será replicada por el node Avamar de San Isidro y que servirá de contingencia.

3.6 Diagrama de Solución final

- Los backup de todos los servidores estará manejado por el administrador de backup la cual convenientemente puede asignar permisos para algunos usuarios como por ejemplo tener acceso a una restauración.
- Respecto al acceso de los grandes volúmenes de base de datos estará definido por el administrador de base de datos, que adicionalmente configura el servidor cliente mediante las aplicaciones de base de datos queda implícito mencionar que los permisos de las bases de datos lo determina el DBA (Data Base Administrator).
- Los servidores de grandes volúmenes de información realizan backup directamente por la red SAN y almacenan la información en el Data Domain en formato deduplicado con esto se logra una mejora en los tiempos de backup, liberar los recursos de la LAN, y tener los backups en línea las copias almacenados en Data Domain.
- Luego se podrá realizar al almacenamiento de los backup fuera de sitio desde San Isidro a Monterrico.
- Avamar realiza backup de los servidores distribuidos remotamente, el flujo de datos de backup viajan a través de la WAN de manera deduplicada y encriptada, guardando estos datos en Avamar para luego replicarla; con esto se logra tener las copias online de los servidores remotos.

La siguiente página (figura 3.7) se representa a alto nivel la manera en que los equipos Data Domain y Avamar se implantarían en la infraestructura de la organización representando la solución final.

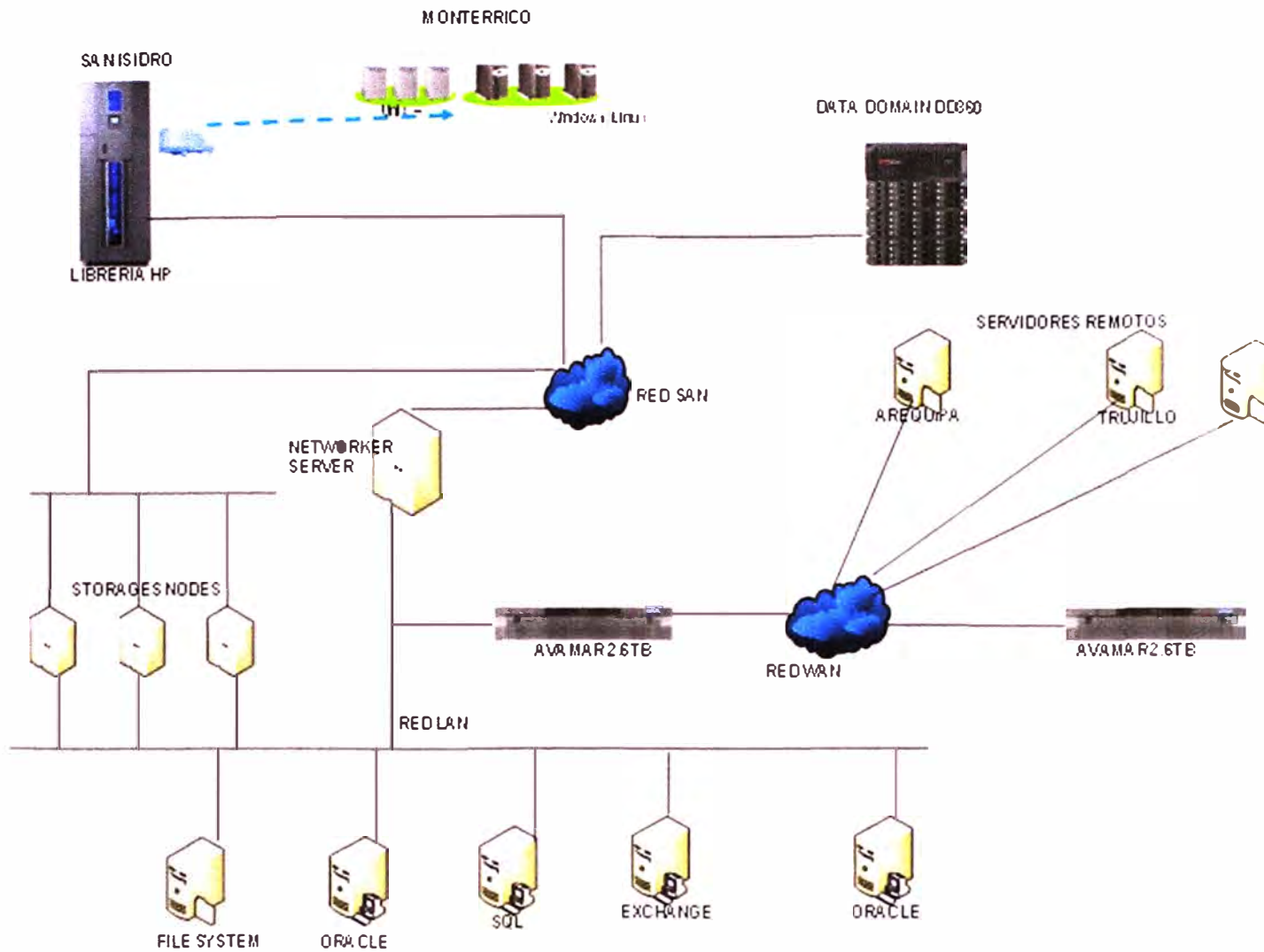


Figura 3.7: Diagrama de Solución Final

CAPITULO IV.
ALCANCE DE LOS SERVICIOS PROFESIONALES, SOPORTE Y
MANTENIMIENTO

4.1 Servicios Profesionales

Las secciones siguientes describen los servicios profesionales con mayor detalle e identifican las obligaciones de EMC, en adelante “Entregables”, las obligaciones de la organización y las responsabilidades mutuas necesarias para la finalización de los servicios dentro del cronograma y costos indicados adjuntos o los que se integran en la solución de backup.

Alcance Avamar

Como alcance general se prevé la instalación de dos equipos EMC Avamar, a saber:

- Instalación de Avamar en cada sitio de la organización y configuración de replicación.
- Instalación de 5 agentes Avamar.

El Avamar se instalará para realizar los backup de los servidores remotos. En tal sentido, cada servidor deberá estar homologado por la matriz de interoperabilidad de Avamar.

Alcance Data Domain

Como alcance general se prevé la instalación de un equipo EMC Data Domain, a saber:

- Instalación de EMC Data Domain DD860 e integración con Networker.
- Configuración de las licencias VTL para Data Domain.

En el caso se presentará como VTL al respectivo software de respaldo Networker. En tal sentido, deberá estar homologado por la matriz de interoperabilidad de Data Domain tanto el software de respaldo como los elementos de conectividad que existirán entre Data Domain y el Networker.

Alcance Networker

Como alcance general se prevé la instalación del software de backup Networker, a saber:

- Instalación y configuración de tres storages node.
- Configuración e integración con Avamar

- Configuración e integración con Data Domain

En este caso se presentará al servidor Networker una VTL Data Domain y un dispositivo de almacenamiento en disco para Avamar al respectivo software de respaldo Networker y se instalará los agentes (clients) descritos. En tal sentido, deberá estar homologado por la matriz de interoperabilidad de Networker.

Enfoque y Organización del Proyecto

EMC utilizará las siguientes fases para ejecutar este proyecto:

- Fase I – Planificación y Diseño
- Fase II – Implementación
- Fase III – Pruebas y Aceptación del Proyecto
- Fase IV – Documentación

Fase I – Planificación y Diseño

Durante esta etapa, EMC prestará los Servicios de Consultoría que incluyen trabajar en estrecha colaboración con el personal de la organización para planificar el proyecto, realizar las reuniones necesarias para recopilar la información relevante a fin de asegurar que satisfagan los requerimientos del negocio.

El equipo de proyecto de EMC:

Determina el nivel de personal requerido de la organización y de EMC para ejecutar el proyecto como se define en este trabajo.

Identifica los contactos de escalamiento en la organización.

Realiza una reunión de inicio con el equipo de la organización.

Informa los pre-requisitos para la instalación del hardware y/o software provistos por EMC, y planifica la implementación de los mismos, verificando que se cumplan todos los requerimientos para una correcta implementación de la solución.

Crea una agenda de programación para reuniones de seguimiento del proyecto.

Realiza entrevistas de acuerdo con un programa acordado en conjunto, a fin de reunir la información requerida.

Planifica y programa las tareas de instalación y configuración de los componentes de hardware y/o software provistos por EMC.

Participa (según se requiera) con la organización y los recursos asignados para desarrollar un plan de proyecto detallado con tareas y fechas acordadas en conjunto por EMC y la organización.

Determina el proceso y programa de implementación de la solución ofrecida por EMC.

Desarrolla el diseño de configuración de la solución ofrecida por EMC.

Proporciona a la organización la documentación de la configuración propuesta.

Propone a la organización el Plan de Pruebas y Aceptación del proyecto.

Fase II – Implementación

Durante esta etapa, EMC ejecutará los Servicios de Consultoría, tal como se describen a continuación.

Instalar y configurar los 3 storages nodes en un servidor físico corriendo sistema operativo homologado por EMC Networker y presentar al servidor Networker.

Instalar y configurar el Data Domain como VTL con todas sus expansiones.

Configuración y conexión del Data Domain hacia el servidor de Backup Networker.

Colocar en el rack propuesto como posición final del Data Domain.

Instalar y configurar el primer nodo Avamar para backup.

Instalar y configurar el segundo nodo Avamar para la sede de replicación.

Instalar 5 agentes Avamar distribuidos en los servidores remotos.

Configurar los dos nodos Avamar en integración al servidor EMC Networker para la administración de las copias desde la consola del EMC Networker.

Las políticas de respaldo a ser implementadas serán informadas por la organización. Eventualmente, EMC podrá realizar sugerencias en relación a best-practices pero no se contempla el diseño de dichas políticas de respaldo y retención.

Nota: Todos los equipos, sistemas operativos y versiones de aplicaciones deberán estar homologados por EMC.

Fase III – Pruebas y Aceptación del Proyecto

EMC llevará a cabo el Plan de Pruebas y Aceptación del proyecto previamente definido y aprobado por las partes en la Fase I, para asegurar que este se haya completado en forma satisfactoria. En caso que se verifique en forma adecuada el cumplimiento de las pruebas establecidas, se darán por aceptados los Servicios de Consultoría.

EMC realizará a posteriori, una revisión básica funcional de la solución implementada para familiarizar al equipo de trabajo de la organización con el nuevo ambiente, de manera de transmitir los conceptos básicos de la solución implementada.

Fase IV – Documentación

Durante esta etapa, EMC:

Proporcionará a la organización una Guía de Configuración que detalla la implementación y documenta los parámetros de configuración, como Arquitectura y políticas utilizadas.

Entregables

El servicio de Implementación brindado por EMC incluye los siguientes entregables:

- Avamar instalado según el alcance especificado.
- Data Domain instalado según el alcance especificado.
- Networker instalado según el alcance especificado.
- Guía de Configuración de EMC documentando la implementación realizada.

Equipo de Trabajo

EMC proporcionará un equipo de proyecto conformado por los siguientes conjuntos de habilidades técnicas y de administración.

Gerente del proyecto

EMC pondrá a disposición a tiempo parcial un Gerente de proyecto para proporcionar liderazgo al supervisar y administrar la ejecución de este esfuerzo en una estrategia de implementación en fases controladas que limitará los requerimientos de downtime e interrupciones de las aplicaciones.

Entre las responsabilidades del Gerente de proyecto de EMC se incluyen:

- Desarrollar un cronograma para el proyecto y asegurar que las programaciones se cumplan.
- Administrar las actividades diarias del equipo del proyecto. Resolver problemas con el Gerente de proyecto de la organización.
- Asegurar que el equipo de EMC, y Gerente de Proyecto de la organización tengan conocimiento del estado y avance del proyecto.
- Proporcionar un único punto de contacto entre la organización, EMC y los recursos de terceros durante el proyecto.
- Recopilar datos pertinentes y necesarios para que se implemente la infraestructura de EMC.
- Realizar la planificación de la evaluación de riesgos durante todo el ciclo de vida del proyecto.
- Verificar que se cumplan todos los requerimientos de hardware y software provistos por EMC.
- Administrar todos los recursos de EMC y la coordinación con el Gerente del Proyecto de la organización, los recursos de la organización y terceros asignados a las actividades relacionadas con el proyecto.

- Realizar reuniones según sea necesario para recopilar la información requerida y asegurar que el diseño de la configuración e implementación cumplan los requerimientos de negocios e infraestructura de TI de la organización.

Arquitecto de soluciones

EMC asignará a tiempo parcial un Arquitecto de Soluciones al proyecto. Las responsabilidades del Arquitecto de Soluciones de EMC serán:

- Verificación de la compatibilidad y del hardware soportado
- Recopilación de datos del proyecto a través de entrevistas y otros medios.
- Desarrollo del plan de pruebas.
- Creación y validación de las solicitudes de archivos de configuración.

Especialista en implementación

EMC proporcionará a tiempo parcial un especialista en implementación experimentado y calificado para la ejecución del proyecto. Las responsabilidades del Especialista en implementación serán:

- Instalar y configurar los diferentes productos de EMC.
- Ejecución del plan de pruebas.
- Proporcionar una revisión básica de los productos de EMC para demostrar que el producto fue instalado exitosamente.

Localización

Los Servicios serán efectuados principalmente en el data Center de la organización.

No obstante, EMC tendrá la opción de realizar distintas tareas como planificación y documentación en una ubicación seleccionada de EMC (el “sitio de EMC”).

Fuera de alcance

La modalidad comercial de la presente propuesta responde en un todo a la de alcance fijo, no estando incluidos servicios profesionales o elementos de Hardware y de Software que no se encuentren explícitamente detallados en dicho alcance.

Toda otra circunstancia no prevista en el alcance en el presente trabajo, será tratada mediante el procedimiento de Control de Cambios, y deberá ser formalizada a fin de poder instrumentar cambios al alcance del presente trabajo.

Se entiende al precio cotizado como fijo para el alcance detallado.

Las siguientes tareas quedan explícitamente fuera de alcance:

- Implementar HBE (Hosted Backup Enabler) o cualquier otra opción de backup integrado con una aplicación.

- Implementar EMC Data Domain en un ambiente mainframe o AS400.
- Instalar o configurar una o más TLU (Tape Library Unit) en la SAN.
- Implementar backup NDMP utilizando la EMC Data Domain como destino de los datos.
- Implementar backup NDMP utilizando el Avamar como destino de datos.
- Instalación de HBAs (Host Bus Adapter) en servidores. Esta instalación la realizará el fabricante del servidor o en su defecto la organización.
- Tendido y etiquetado de fibra óptica, cables de red, elementos de HW, tales como HBAs, necesarios para la conectividad de los elementos involucrados.
- Instalación de nuevos equipos resultantes de nuevas adquisiciones.
- Iniciativas que requieran capacidades no provistas por el/los consultor/es: Análisis y programación de aplicaciones, etc.
- Instalación, configuración y soporte de equipos eléctricos, de telecomunicaciones e infraestructura de cableado.
- Diseño y/o rediseño de políticas y procedimientos de backup. Eventualmente, EMC podrá realizar algún tipo de recomendación si detectase oportunidad de mejora.
- Cualquier actividad de administración y mantenimiento sobre la Infraestructura de Backup de la organización.
- Cualquier actividad de administración y mantenimiento sobre la Infraestructura de Servidores y Sistema Operativo de la organización.
- Instalación de cualquier producto de software adicional tanto EMC como NO-EMC.
- Migración de datos y/o cintas.
- Importación de Cintas.
- Cualquier tipo de scripting no explicitado en el presente trabajo.
- Cualquier otro Entregable no mencionado explícitamente en el presente.

Supuestos

Los servicios para esta implementación están sujetos a los siguientes supuestos:

- Aprobación del diseño propuesto, dentro del plazo fijado en el cronograma respectivo.
- Todo cambio en el alcance no contemplado por este trabajo deberá ser formalizado por el proceso de Control de Cambios. No se procederá a ejecutar cambio alguno en relación al alcance previsto en el presente trabajo, sin haber obtenido en forma previa el acuerdo de las partes.
- Los servidores, los equipos, HBAs y softwares de backup deberán estar certificados por la EMC Support Matrix (<https://elabnavigator.emc.com/do/navigator.jsp>).

- El precio y los descuentos especiales contenidos en este trabajo, o bien en la propuesta comercial, no obligan a EMC a ofrecer cotizaciones similares a la organización en contratos adicionales o futuros.

4.2 Servicio de Soporte y Mantenimiento Post-Venta

La garantía para el hardware y software es de 36 meses.

El soporte 7x24 para el servicio de hardware, la modalidad de este soporte es Premium.

El soporte 5x9 Remoto en horarios de oficina para el servicio de Software, la modalidad de este soporte es Enhanced.

Incluye acceso gratuito a nuevas versiones del software, que estén disponibles en el mercado durante el periodo de vigencia del mantenimiento.

El acceso al soporte telefónico será en idioma español a través del número:

0-800-52328.

Opcionalmente el acceso al soporte telefónico en idioma Inglés a través del número

0-800-52327.

También se podrá acceder al mismo a través de EMC Powerlink, nuestra página web destinada para tal fin. La misma será brindada por EMC en forma remota (desde USA). Adicionalmente EMC Perú cuenta con personal de Customer Service certificado en las tecnologías propuestas.

Descripción de la estructura de soporte local

- ◆ **Gerente de Área:** Eduardo Unger
- ◆ **Regional Technical Support:** Emilio Sero
- ◆ **Customer Eng. Sr:** Daniel San Roque
Oscar Stanich
- ◆ **Regional Software Support:** Claudio Feld
Gabriel Saroka

4.3 Cursos de Capacitación

La transferencia de conocimientos estará cubierta por los siguientes aspectos:

Capacitación para 1 persona para un curso oficial de Data Domain administration.

Capacitación para 1 persona para un curso oficial de Administración Avamar.

CAPITULO V. COSTO DEL PROYECTO Y CONDICIONES COMERCIALES

5.1 Oferta Económica

Se adjunta la propuesta por la solución de backup con la finalidad de optimizar y mejorar el desarrollo de backup en la organización además de los inconvenientes existentes.

VALOR DE LOS COMPONENTES DE DATA DOMAIN:

Los valores detallados incluyen los servicios profesionales para la instalación y la configuración de integración con el Networker.

US\$ 407,235.00 + IGV

VALOR DE LOS COMPONENTES DE NETWORKER:

Los valores detallados incluyen los servicios profesionales para la instalación y la configuración de integración con el Data Domain y el Avamar

US\$ 108,689.00 + IGV

VALOR DE LOS COMPONENTES DE AVAMAR:

Los valores detallados incluyen los servicios profesionales para la instalación y la configuración de integración con el Networker.

US\$ 35,930.00 + IGV

VALOR TOTAL ESPECIAL DE LA SOLUCION US\$ 551,854.00 + IGV

Los valores previamente detallados incluyen los servicios profesionales para la instalación y la configuración básica de los mismos.

5.2 Condiciones Comerciales

Precios

Los precios indicados están expresados en Dólares Estadounidenses y no incluyen el IGV.

Los precios citados precedentemente corresponden a equipamiento puesto en el domicilio del Cliente, e incluyen la puesta en marcha y prueba de funcionamiento del equipamiento según test y especificaciones técnicas y funcionales de fábrica.

En caso que las autoridades gubernamentales pertinentes incrementen las tasas y/o aranceles de importación, y/o introduzcan cualquier nuevo gravamen sobre la importación y/o comercialización del equipamiento propuesto, y el mismo tenga fuerza vinculante, EMC se reserva el derecho de modificar el precio cotizado sin previo aviso.

Plazo de Entrega

El plazo de entrega será dentro de los 45 días, a partir de la recepción de la orden de compra en EMC.

Lugar de Instalación

El lugar de instalación (Sitio) del Cliente, deberá cumplir con los requerimientos técnicos de ambiente y alimentación requeridos por EMC para el equipamiento cotizado en la presente.

La preparación y/o adecuación del lugar de instalación, incluyendo instalación eléctrica de alimentación, cableado de cualquier tipo, acondicionamiento de aire/humedad, etc. correrá por exclusiva cuenta y cargo del Cliente, limitándose en este aspecto la responsabilidad de EMC a la verificación pre-entrega del cumplimiento del sitio de los requerimientos de ambiente y alimentación.

Soporte y Mantenimiento

La presente propuesta económica está incluyendo el soporte y mantenimiento de la solución como se detalla a continuación:

Soporte “Premium” para el hardware 24x7x365 durante los siguientes 36 meses con 4 horas de respuesta como máximo frente a una falla disco duro o cualquier componente de hardware ofertado.

Soporte “Enhanced” para el software 5x9 durante los siguientes 36 meses con tiempo de respuesta “nex business day” (al siguiente día laboral) de tiempo de respuesta como máximo.

Esto permite en general actualización del software ofertado y actualización del micro-código o upgrade del hardware propuesto a efectos obtener mejoras en el desarrollo o aprovechar las nuevas características del software de backup, dicha actualización sin costo estará disponible mientras dure el periodo de mantenimiento.

Validez de la Oferta

La presente propuesta posee una validez de oferta de treinta (30) días, contados a partir de la fecha del encabezado.

Forma de Pago

La organización deberá cancelar el valor del equipamiento adquirido, en modalidad Compraventa-Contado, a favor de EMC, de la siguiente forma: 100% a los 30 días de la fecha de factura.

ANEXO A
EMC Data Protection Advisor

CONCLUSIONES Y RECOMENDACIONES

1. En el presente trabajo se puede observar que se ha cubierto los problemas previamente definidos en el ambiente de backup con las tecnologías nuevas como por ejemplo desarrollar backup a disco con tecnología de Deduplicación en origen (Avamar), en el destino (Data Domain) y tecnologías existentes presentes en la organización, lo que se realizó es adquirir nuevas licencias Storages Node de Networker para poder instalar en los 3 servidores que tienen grandes volúmenes de datos (26TB, 10TB y 4.8TB), además estos servidores ya eran clientes Networker respaldando aplicaciones base de datos Oracle (por lo tanto homologados por la matriz de compatibilidad Networker), para luego pasar a ser Storages Nodes y poder tener acceso directo a la red SAN y a través esta red pasar el flujo de datos de backup.
2. En este trabajo se ha presentado la tecnología de Deduplicación en el origen con Avamar, la cual posee beneficios como reducir el impacto sobre la red WAN, permite un respaldo rápido, replica la data en formato deduplicado y esto eliminó la necesidad de transporte de cintas fuera de sitio.
3. El almacenamiento en Avamar Single Node posee una solución basado en discos con la protección RAID1 y posee procesador multicore QuadCore 2.0GHz y memoria RAM 24GB por lo que en el desarrollo de los backup remotos sólo consumirá recursos del cliente Avamar y del Avamar storage node y no del servidor Networker, el servidor Networker se encargará de almacenar la metadata.
4. En este trabajo se ha presentado la tecnología de Deduplicación en el destino con Data Domain, posee beneficios como la Deduplicación en línea (la data es deduplicada antes de su almacenamiento), los archivos desde los discos pueden ser recuperados en segundos o minutos, variedad en la conectividad NFS, CIFS, NDMP o FC/VTL en simultaneo dando opción la organización poder respaldar soluciones NAS.
5. Data Domain además tiene la capacidad de replicar las copias en formato deduplicados de forma similar que Avamar a través de la red WAN es decir transmitir datos únicos.

6. Pensar en replicar la data y reducir, incluso eliminar la necesidad de transportar las cintas fuera de sitio.
7. El análisis desarrollado en el presente trabajo se ha logrado gracias a las experiencias de diferentes organizaciones/clientes que tienen grandes volúmenes de datos y que constantemente siguen creciendo y en muchos casos tienen similares necesidades con algunas particularidades.
8. A manera de recomendación se puede reprogramar algunos backup Fulls para que se realice los días de semana y no todos ellos el fin de semana, esto permite reducir el cuello de botella y otros recursos implicados.
9. También se observa en una organización que posee grandes volúmenes de datos, tienen muchos archivos (millones de archivos) en varios servidores donde se respalda cada uno de ellos, para aligerar la carga de backup se puede contar con soluciones NAS (Network Attached Storage) que permite tener estos millones de archivos en un solo servidor con permisos y autenticaciones en la cual sólo los servidores con los permisos y autorización asignada pueden acceder a estos archivos y por lo tanto sólo acceder a estos archivos en vez de llevar todos estos archivos (millones de archivos) en el disco de cada servidor.
10. Desarrollar soluciones de Archiving de Emails y File server para archivos que no son accedidos o modificados por un tiempo determinado y enviarlos a un repositorio con Deduplicación con discos de gran capacidad, con la intención de liberar espacio los discos que poseen alta velocidad.
11. Desarrollar un plan periódico de “backup infrastructure assessment” y sacar reportes de capacity planning así como también ver los reportes del comportamiento del ambiente de backup y rendimiento en la red, la solución basada en “Data Protection Advisor” que puede sacar estos reportes se analiza en este trabajo en calidad de anexo.

DPA es una solución de reportes y análisis para el software de Backup de datos y permite a los usuarios una visibilidad completa dentro de todas las cosas que están pasando en su entorno de backup, tomar las decisiones correctas antes de una eventual falla o tomar las decisiones correctas más rápida frente a una falla reciente.

DPA consiste en un motor de colección, análisis y reportes. En tal sentido, DPA permite coleccionar información de performance, estado de la o las plataformas de respaldo, luego permite presentar, analizar, correlacionar, sacar promedios, desvíos y otras operaciones para facilitar el monitoreo y administración de la plataforma de respaldo de una manera flexible y fácil de personalizar. En la figura A.1 se muestra la consola de reportes DPA.

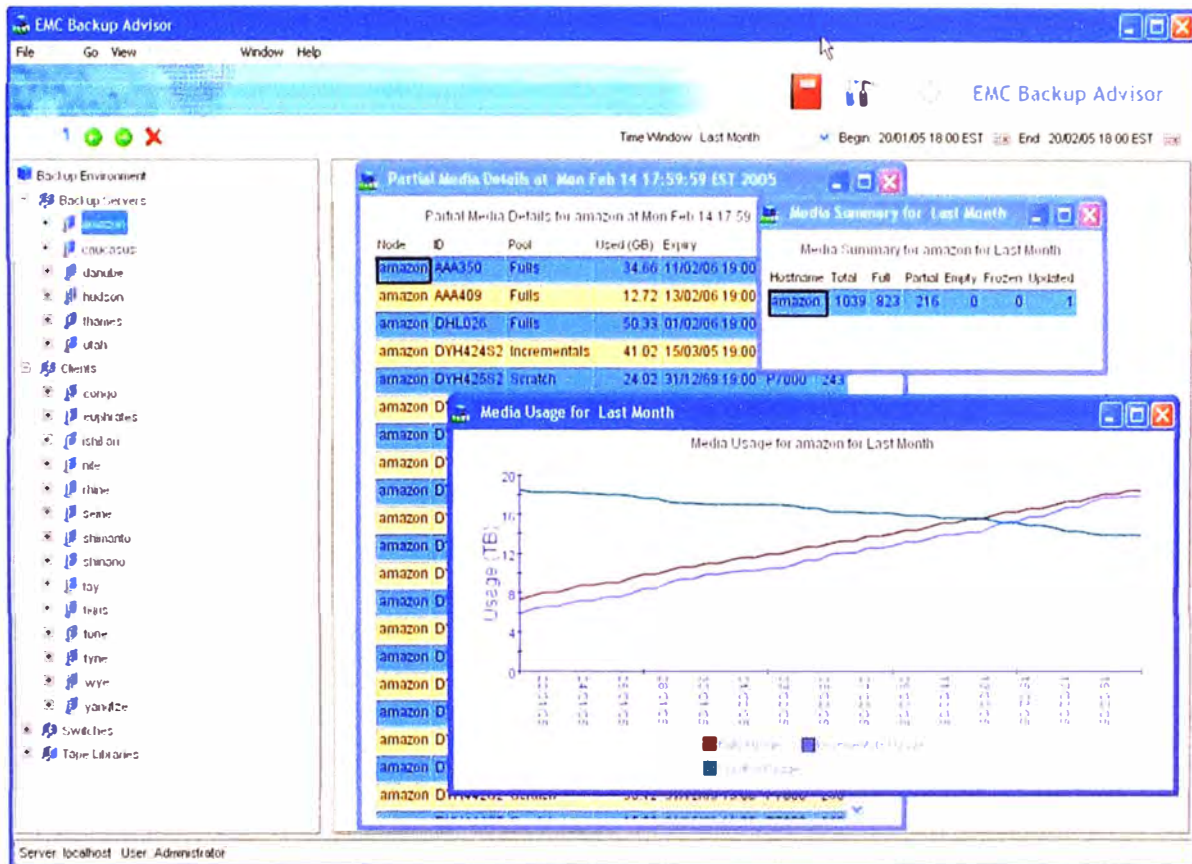


Figura A.1. Consola de reportes DPA.

DPA cuenta con más de 400 reportes predefinidos y permite además, la capacidad de diseñar reportes personalizados para alcanzar distintos niveles de información. DPA organiza estos reportes en cuatro grandes dominios de información:

- Planificación de Capacidad
- Administración de Cambios
- Devolución de Cargo
- Rendimiento
- Organización de Tareas

En la siguiente figura A.2 se muestra la administración de Chargeback.

Chargeback Overview for Configuration for Last Week

ChargeBack Summary

Num. Backups	Backup Size (GB)	Num. Restores	Restore Size (GB)	Data Retained (GB)	Num. Tapes	Total Cost
147	2 048	34	2	3 74	3	<u>£301 73</u>

Figura: A.2. Administración de Chargeback

EMC DPA es capaz de coleccionar, analizar y correlacionar eventos de plataformas de respaldo EMC (Networker, Avamar) como de 3ras partes como TSM, Netbackup entre otros.

EMC DPA permite realizar el monitoreo centralizado de distintas plataformas de respaldo y recuperación en simultáneo permitiendo:

Monitorear con alertas en tiempo real, monitorea identifica potenciales problemas en la plataforma antes que escalen.

Desarrollo de inconvenientes (Troubleshooting) con los datos correlacionados.

Identifica la raíz del problema (root cause) en las herramientas de backup.

Reportes de rendimiento incluyendo información del host, dispositivo de respaldo y dispositivos intermedios.

Realizar capacidad de planificación (Capacity planning) para tomar decisiones correctas frente al crecimiento esperado de la información en un determinado tiempo.

Ganar visibilidad en requerimientos futuros (Cintas, discos)

Flexibilidad 100% customizable para alcanzar los requerimientos de presentación de la información.

Desde la perspectiva financiera, DPA cuenta con un módulo incluido llamado "Chargeback" que provee la capacidad de realizar análisis de costos financieros de la plataforma de respaldo. Los costos son calculados para cada cliente de respaldo y puede ser direccionado a una unidad de negocios responsable por ese o por esos clientes.

Permitiendo obtener reportes como el siguiente:

En el caso de Netbackup, se pueden generar un reporte que permita determinar un monto a cargar a un centro de costos en función del número de respaldos, el tamaño del respaldo (GB protegidos y GB restaurados).

Solución EMC Data Protection Advisor

Dada la flexibilidad de Data Protection Advisor, se presenta un escenario de DPA que permite alcanzar a la organización tener los reportes, análisis y capacidad de planificación de todas las cosas que está pasando en su entorno de backup desde una sola consola de administración, para luego tomar las decisiones correctas antes de una eventual falla o tomar las decisiones correctas más rápida frente a una falla reciente.

En función al requerimiento, a continuación se detalla los componentes asociados a dicho requerimiento:

Licencias de Data Protection Advisor para la organización

La solución propuesta está compuesto por:

- 1 (Uno) DPA server license para windows.
- 8 paquetes (20 clients) Data Protection Advisor 160 clientes DPA en total.
- Licencia free para dispositivos de backup (librería, switches, VTL)

La configuración descrita soporta las aplicaciones de software de backup, base de datos, soporta aplicaciones para monitoreo de la red LAN, red SAN y todos los dispositivos de backup que están homologados en la matriz de compatibilidad del DPA.

El DPA recolecta información desde el backup servers, clientes, base de datos, Network switches, SAN switches, librerías robóticas físicas, almacenamiento en disco y librerías virtuales, homologados por la matriz de compatibilidad del DPA.

DPA correlaciona la información de cada área y proporciona la habilidad de desarrollar reportes avanzados.

Todos los reportes pueden ser almacenados en una base de datos propietaria del DPA como iAnywhere o alguna proporcionada por la organización esta base de datos puede ser SQL en Windows, Oracle en Windows y todas aquellas bases de datos homologados por la matriz de compatibilidad del DPA.

También todos los reportes como soluciones de entornos NAS y replicación de Storage de discos, siempre que estos dispositivos estén homologados por la matriz de compatibilidad del Data Protection Advisor.

En la siguiente figura A.3 se ilustra el diagrama de conexión DPA con integración de la solución basada en Networker, Data Domain y Avamar.

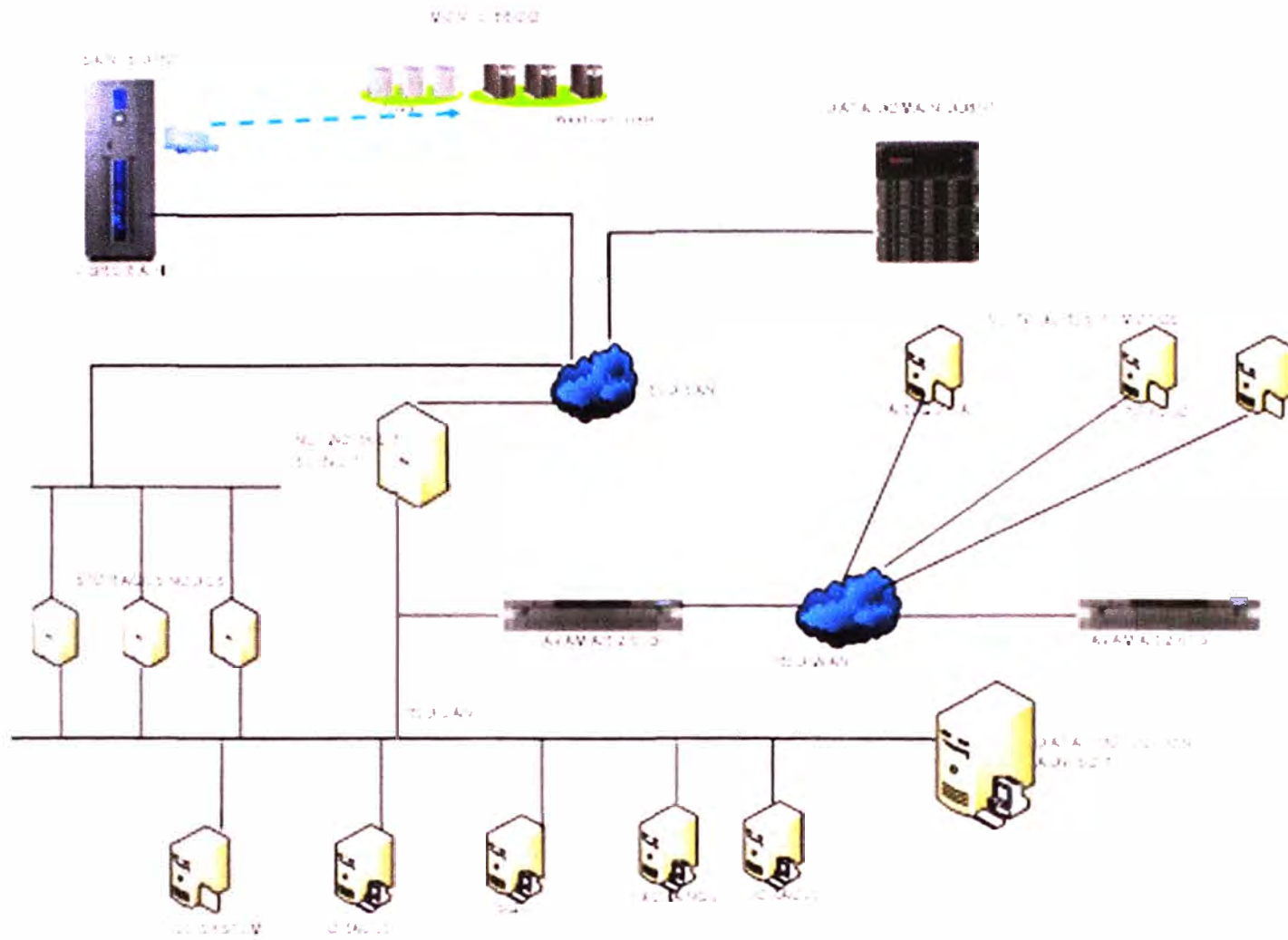


Figura: A.3: Diagrama de conexión DPA

ANEXO B
GLOSARIO DE TÉRMINOS

AL PA: Arbitrated Loop Physical Address.

BACKUP: Copia de Respaldo de una información.

BACKUP CLIENT: Un servidor que será o está siendo respaldado por un determinado software de backup.

BMR: Bare Metal Restore solución de backup para recuperación de desastres.

CATALOGO: Base de datos de un software de backup

CIFS: Common Internet File System.

DAS: Direct Attached Storage.

DATA SETS: Archivos o datos que se van a respaldar.

DBA: Data Base Administrator.

DEDUPLICACION: Proceso de detectar e identificar bloques únicos de datos a los efectos de reemplazar datos duplicados por referencias.

DEDUPLICACION GLOBAL: Proceso de deduplicación entre todos los data sets

DOWNTIME: Tiempo en la cual los procesos se detienen.

DPA: Data Protection Advisor.

EMC: Empresa dedicada a la solución de almacenamiento.

FABRIC: Opción de interconectividad en una arquitectura Fibra Canal.

FC: Fibra Canal.

FRAME: Trama o paquete de datos.

FULL BACKUP: Backup completo de todos los archivos.

GUI: Grafical User Interface.

HBA: Host Bus Adapter.

I/O: Input/Output define la cantidad de data entrante y saliente en un servidor.

ISL: InterSwitch Link protocolo que permite la interconexión entre dos switch.

LAN: Local Area Network.

LC: Lucent Connector.

LUN: Logical Unit Number.

MEDIA: Dispositivo de almacenamiento tape, disco, etc.

METADATA: Información para el registro del catálogo en un software de backup.

MMF: Multi Mode Fiber.

NAS: Network Attached Storage.

NDMP: Network Data Management Protocol.

NETBACKUP: Software de Backup.

NFS: Network File System.

NIC: Network Interface Controller.

OFFSITE: Lugar alternativo a un data center.

ORACLE: Base de datos.

RAID: Protección en un arreglo de discos.

RESTORE: Recuperación de la información desde un backup.

RPO: Recovey Point Objective.

RTO: Recuperación Time Objective.

SAN: Storage Area Network.

SC: Standar Connector.

SCSI: Small Computer System Interface.

SFP: Small Form-Factor Pluggable.

SMF: Single Mode Fiber.

SNAPSHOT: Copia en disco basada en punteros.

SIZING: Cálculo en el tamaño de almacenamiento.

SQL: Base de datos propietaria de Windows.

STAGING: Copia temporal a disco llamado también Staging Area.

STORAGE DE DISCO: Almacenamiento basado en disco.

SWITCH FABRIC: Switch para redes SAN que permite escalabilidad.

TAPE: Dispositivo de almacenamiento basado en cinta magnética.

TAPE DRIVE: Dispositivo que permite escribir/leer en la cinta magnética.

THROUGHPUT: Velocidad de transmisión de datos en una determinada red.

TI: Tecnología de la Información.

TLU: Tape Library Unit.

TROUBLESHOOTING: Método de resolución de problema basado en el descarte de posibilidades hasta encontrar el punto de falla.

TSM: Tivoli Storage Manager.

ULP: Upper Layer Protocol.

UPGRADE: Acción que implica la compra componentes o de un dispositivo basado en su crecimiento.

WAN: Wide Area Network.

BIBLIOGRAFÍA

- [1] Alan F Benner, “Redes sobre Fibra Canal”, Mcgraw-Hill/Interamericana de España, Edición, 2002**
- [2] Alberto Prieto Espinosa, “Introducción a la informática”, Mcgraw-Hill/Interamericana de España, Edición, 2006**
- [3] Tom Clark Alfredo, “Designing Storage Area Networks”, Gecko Networks Colombia, Segunda Edición 2003**
- [4] W. Curtis Preston, “Using SANs and NAS”, Sebastopol, CA 95472, 2002**
- [5] Robert W. Kembel, “Fibra Canal a comprehensive introduction”, USA, Edición 2004**
- [6] EMC Education, “Information Storage and Management”, EMC Corporation Hopkinton, 2004**
- [7] EMC Education, “EMC Backup and Recuperación Technologies”, EMC Corporation Hopkinton, 2004**
- [8] Ulf Troppens “Storage Networks Explained”, Speichernetze, Alemania 2003**
- [9] <http://www.fibrechannel.com>**
- [10] <http://www.forrester.com>**
- [11] <http://www.infonetics.com>**
- [12] <http://www.redbooks.ibm.com>**
- [13] <http://www.gosignal.com/whitepapers/quantum1.pdf>**
- [14] <http://www.emc.com>**
- [15] <https://my.datadomain.com/US/en/documentation.jsp>**