




*Instituto Universitario Aeronautico*

Ingeniería de Sistemas



# Trabajo Final de Grado

PROCEDIMIENTO PARA LA SELECCIÓN DE  
INFRAESTRUCTURA DE BACKUP EN LAS ORGANIZACIONES

**Paris, María**

**DNI: 28.064.716**

**Whelan, Germán**

**DNI: 29.606.838**

**Año 2014**



## ÍNDICE GENERAL

<b>CAPÍTULO I – Introducción</b> .....	1
<b>1.1. Antecedentes</b> .....	1
<b>1.2. Situación Problemática</b> .....	2
<b>1.3. Problema</b> .....	3
<b>1.4. Objeto de Estudio y Campo de Acción</b> .....	3
<b>1.5. Objetivos</b> .....	4
1.5.1. Generales .....	4
1.5.2. Específicos .....	4
<b>1.6. Idea a Defender/Propuesta a Justificar/Solución a Comprobar</b> .	4
<b>1.7. Delimitación del Proyecto</b> .....	5
<b>1.8. Aporte Teórico</b> .....	5
<b>1.9. Aporte Practico</b> .....	6
<b>1.10. Métodos y Medios de Investigación</b> .....	6
<b>CAPÍTULO II – Marco Contextual</b> .....	8
<b>2.1. Conceptos Básicos</b> .....	8
2.1.1. Backup .....	8
2.1.2. Dispositivos de Almacenamiento Secundarios.....	9
2.1.2.1 Disco Duro .....	9
2.1.2.2 Cinta Magnética .....	12
2.1.3. Interfaces de Conexión de Dispositivos .....	14
2.4.2.1 Grupo ATA.....	14
2.4.2.2 Grupo SCSI .....	15
2.4.2.3 Resto de Interfaces.....	15
2.1.4. Dispositivos Lectura/Escrituras .....	16
2.1.4.1 Dispositivos Basados en Cintas .....	16
2.1.4.1 Dispositivos Basados en Discos .....	18
<b>2.2. Entorno del Objeto de Estudio</b> .....	19
<b>2.3. Relación Tesista y Objeto de Estudio</b> .....	19
<b>2.4. Análisis de los Problemas Observados</b> .....	20
2.4.1. Descripción de la situación Actual .....	22
2.4.2. Características, conexiones y configuraciones de los sitios actuales.....	22

2.4.2.1 Sitio Grande.....	22
2.4.2.2 Sitio Mediano.....	26
2.4.2.3 Sitio Pequeño.....	30
<b>CAPÍTULO III – Marco Teórico.....</b>	<b>34</b>
<b>3.1. Librería de Cintas Tradicionales.....</b>	<b>34</b>
3.1.1. Ventajas y Desventajas generales.....	35
3.1.1.1 Ventajas.....	35
3.1.1.2 Desventajas.....	35
<b>3.2. Librerías de Cintas Virtuales.....</b>	<b>36</b>
3.2.1. Ventajas y Desventajas generales.....	37
3.2.1.1 Ventajas.....	37
3.2.1.2 Desventajas.....	37
<b>3.3. Librería Física vs Librería Virtual.....</b>	<b>38</b>
<b>3.4. Deduplicación.....</b>	<b>38</b>
3.4.1. ¿Qué es la Deduplicación de datos?.....	38
3.4.2. Ratios de Deduplicación.....	40
3.4.3. Beneficios.....	42
3.4.4. Ventajas.....	43
3.4.5. Desventajas.....	45
3.4.6. Técnicas de Deduplicación.....	45
3.4.6.1 Por Nivel de Análisis (File o Sub-File).....	45
3.4.6.1.1 Nivel de Archivos.....	46
3.4.6.1.2 Nivel de Bloques.....	46
3.4.6.2 Por Algoritmo.....	47
3.4.6.2.1 Por Hashing.....	47
3.4.6.2.2 Por Diferencias Binarias.....	50
3.4.6.3 Con o sin conocimiento de Metadatos/Estructuras de Datos.....	51
3.4.6.3.1 Técnicas Content-Aware.....	51
3.4.6.3.2 Técnicas Agnósticas.....	53
3.4.6.4 Por Agente de Deduplicación.....	53
3.4.6.4.1 En Origen.....	53
3.4.6.4.2 En Destino.....	54
3.4.6.5 Por Instante de Deduplicación.....	54

3.4.6.5.1	Inline .....	54
3.4.6.5.2	Diferida.....	54
3.4.6.6	Por Tipo de Solución: Appliance o Software .....	54
3.4.7.	Ideas Erróneas de la Deduplicación .....	55
<b>3.5.</b>	<b>Matriz de Ponderación .....</b>	<b>56</b>
3.5.1.	¿Qué es una Matriz de Ponderación?.....	56
3.5.2.	¿Cómo se elabora una Matriz de Ponderación? .....	56
<b>3.6.</b>	<b>Estudio de Selección de Alternativas .....</b>	<b>57</b>
<b>3.7.</b>	<b>Propuesta de soluciones .....</b>	<b>67</b>
3.7.1.	Propuesta para Sitio Grande – Curitiba. ....	67
3.7.2.	Propuesta para Sitio Mediano – Córdoba. ....	69
3.7.3.	Propuesta para Sitio Pequeño – Buenos Aires. ....	71
<b>3.8.</b>	<b>Comparación de Solución Actual vs Propuesta .....</b>	<b>73</b>
3.8.1.	Comparación de Sitio Grande – Curitiba.....	74
3.8.2.	Comparación de Sitio Mediano – Córdoba.....	77
3.8.3.	Comparación de Sitio Pequeño – Buenos Aires .....	78
<b>Capítulo IV - Conclusión.....</b>		<b>81</b>
<b>Bibliografía .....</b>		<b>82</b>
<b>Anexos .....</b>		<b>84</b>
<b>Anexo A – Librerías de Cintas.....</b>		<b>84</b>
<b>Anexo B – Librerías de Cintas Virtuales.....</b>		<b>89</b>
<b>Anexo C – Estudio Selección Hardware .....</b>		<b>93</b>

## ÍNDICES TABLAS

2.1. Capacidad bruta vs. Capacidad real en los discos duros .....	10
2.2. Comparación de los diferentes tipos de RAID .....	12
2.3. Tipos de interfaces y sus tasas de transferencia máxima.....	16
2.4. Clasificación de los sitios según tamaño y complejidad.....	21
2.5. Descripción de los servidores de backup de Curitiba .....	23
2.6. Descripción de la librería de backup L700.....	24
2.7. Volumetría de datos resguardados en una semana – sitio grande .....	24
2.8. Volumetría de datos con 21 días de retención – sitio grande .....	24
2.9. Costo mantenimiento de licencias de hardware de Curitiba .....	25
2.10. Descripción del servidor de backup de Córdoba .....	27
2.11. Descripción de la librería de backup L180.....	28
2.12. Volumetría de datos resguardados en una semana – sitio mediano .....	28
2.13. Volumetría de datos con 21 días de retención – sitio mediano.....	28
2.14. Costo mantenimiento de licencias de hardware de Córdoba .....	29
2.15. Descripción del servidor de backup de Buenos Aires .....	31
2.16. Descripción de la librería de backup L40.....	31
2.17. Volumetría de datos resguardados en una semana – sitio pequeño.....	31
2.18. Volumetría de datos con 21 días de retención – sitio pequeño.....	32
2.19. Costo mantenimiento de licencias de hardware de Buenos Aires .....	32
3.1. Librerías Tradicionales vs. Librerías virtuales de cintas .....	38
3.2. Relación Ratio vs. Porcentaje de reducción .....	41
3.3. Tamaño de índice vs. Tamaño de repositorio (segmentos 8kb) .....	48
3.4. Tamaño de índice vs. Tamaño de repositorio (segmentos 16kb) .....	50
3.5. Escala Valorativa de Criterios .....	59
3.6. Matriz de Ponderación de Criterios .....	60
3.7. Escala Valorativa de Opciones.....	60
3.8. Matriz Final para la selección del mejor hardware.....	63
3.9. Matriz Final para la selección del mejor proveedor.....	65
3.10. Matriz Final para la selección del hardware para cada sitio.....	66
3.11. Descripción del nuevo servidor de Curitiba .....	68
3.12. Descripción de la VTL de Curitiba .....	69

3.13. Descripción de la librería de Curitiba.....	69
3.14. Descripción del nuevo servidor de Córdoba.....	71
3.15. Descripción de la librería de Córdoba.....	71
3.16. Descripción del nuevo servidor de Buenos Aires.....	72
3.17. Descripción de la nueva librería de Buenos Aires.....	73
3.18. Matriz final para la selección de la mejor infraestructura – sitio grande.....	75
3.19. Resumen de los problemas con sus soluciones y mejoras en sitio grande.....	76
3.20. Matriz final para la selección de la mejor infraestructura – sitio mediano.....	77
3.21. Resumen de los problemas con sus soluciones y mejoras en sitio mediano.....	78
3.22. Matriz final para la selección de la mejor infraestructura – sitio pequeño.....	79
3.23. Resumen de los problemas con sus soluciones y mejoras en sitio pequeño.....	80
C.1 Configuración librería Curitiba.....	93
C.2 Configuración librería Córdoba.....	95
C.3 Configuración librería Buenos Aires.....	96

## ÍNDICES FIGURAS

1.1. Esquema de las etapas de la metodología.....	7
2.1. Roadmap DLT.....	13
2.2. Roadmap LTO-Ultrium .....	14
2.3. Drive .....	17
2.4. Autoloader.....	17
2.5. Librería.....	18
2.6. DataDomain para backup a disco .....	18
2.7. Diseño infraestructura actual sitio de Curitiba – Sitio grande.....	23
2.8. Diseño infraestructura actual sitio de Córdoba – Sitio mediano.....	27
2.9. Diseño infraestructura actual sitio de Buenos Aires – Sitio pequeño ....	30
3.1. Deduplicación .....	39
3.2. Ejemplo de deduplicación .....	40
3.3. El Ratio de Reducción de Espacio .....	41
3.4. La curva de ahorro de espacio.....	42
3.5. Impacto en la performance por incremento del tamaño de índice .....	49
3.6. Tablas de comparación entre opciones de hardware según su impacto para cada criterio .....	62
3.7. Tablas de comparación entre opciones de proveedores según su impacto para cada criterio .....	64
3.8. Tablas de comparación entre opciones de tamaño según su impacto para cada criterio .....	66
3.9. Diseño de infraestructura propuesta para sitio de Curitiba – Sitio grande.....	68
3.10. Diseño de infraestructura propuesta para sitio de Córdoba – Sitio mediano .....	70
3.11. Diseño de infraestructura propuesta para sitio de Buenos Aires – Sitio pequeño .....	72
3.12. Tablas de comparación entre infraestructura actual vs. Propuesta – Sitio grande.....	74
3.13. Tablas de comparación entre infraestructura actual vs. Propuesta – Sitio mediano .....	77
3.14. Tablas de comparación entre infraestructura actual vs. Propuesta – Sitio pequeño .....	79
A.1 Librerías de cintas de Hewlett Packard .....	85
A.2 Librerías de cintas de Hewlett Packard (cont.) .....	86

A.3 Librerías de cintas de SUN/Oracle .....	87
A.4 Librerías de IBM.....	88
B.1 Librería de cintas virtuales de Hewlett Packard.....	89
B.2 Librería de cintas virtuales de SUN/Oracle – Línea Prime.....	90
B.3 Librerías de cintas virtuales de SUN/Oracle – Línea Plus .....	90
B.4 Librerías de cintas virtuales de EMC – Datadomain .....	91
B.5 Librerías de cintas virtuales de IBM .....	92
C.1 Dimensionamiento Datadomain .....	94



# *Capítulo I - Introducción*

---

## **CAPÍTULO I - INTRODUCCIÓN**

### **1.1 ANTECEDENTES**

El desarrollo de este **Trabajo Final de Grado**, tiene como finalidad principal la obtención de nuestro título de Ingenieros en Sistemas coronando de esta manera una etapa de mucho estudio y sacrificio a lo largo de estos años académicos.

La selección de nuestro tema, parte como una necesidad de negocio dentro de una empresa multinacional francesa dedicada a la fabricación y comercialización de automóviles situada en la ciudad de Córdoba en la cuál desarrollamos nuestra actividad laboral desde hace algunos años dentro del área informática, trabajando específicamente con la infraestructura de servidores y backup.

La empresa tiene la necesidad de realizar un estudio o investigación de nuevas tecnologías ofrecidas en el mercado relacionadas a la infraestructura de backup, que le permita mejorar o sustituir la infraestructura utilizada actualmente, teniendo en cuenta ciertas características que serán mencionadas y desarrolladas a lo largo de este proyecto de investigación.

La empresa ve a este proyecto como una forma de mejorar los rendimientos de backup/restore sin descuidar la seguridad en el resguardo de los datos y mejorar sobre todo el rendimiento Costo/Beneficio.

Además de esta necesidad de la empresa, la decisión de realizar esta investigación por nuestra parte, es el interés personal de conocer las nuevas tecnologías de hardware, técnicas y metodologías de backup que nos permita ofrecer mejores soluciones de resguardo de la información a cualquier tipo y tamaño de empresa.

Algunas razones por las que se escogió realizar este trabajo:

- Posibilidad de aplicar los conocimientos obtenidos a lo largo de esta investigación en la organización donde desempeñamos nuestra actividad laboral.
- Es parte de nuestro trabajo diario y es considerado un punto muy crítico en los objetivos de negocio de la empresa a nivel del área de informática.

- La importancia y el valor que tiene la información dentro de cualquier organización.

Como se mencionó anteriormente, se llevará adelante esta investigación tomando como base **la infraestructura de soluciones de backup** utilizadas por la empresa. La misma cuenta con 75 sitios distribuidos en 25 países a los largo del mundo, donde el tamaño, características y criticidades de cada uno de ellos son diversos.

Las soluciones de backup homologadas por la empresa que se están utilizando actualmente incluyen:

- Servidores de backup + robot STK (StorageTek) administrados por la aplicación Time Navigator de Atempo.
- Algunas excepciones utilizan la copia de disco a disco según se detalla:
  - o Backup de filer sobre filer (NetApp) vía SnapMirror (sitios pequeños)
  - o Servidores de backups + filer sobre filer vía SnapVault
- Aunque no es aplicado de igual manera en todos los sitios se realiza la externalización de cintas vía outsite.

## **1.2 SITUACIÓN PROBLEMÁTICA**

Los puntos frágiles relevados dentro de la organización son:

- Incremento exponencial año a año de la información.
- Alto costo de mantenimiento de la solución actual de backup.
- Ventanas de backups cada vez más reducidas.
- Presupuestos limitados.
- Arquitectura trabajando al 100% y sin posibilidades de ampliación.
- Puntos de recuperación (RPO) y tiempos de recuperación (RTO) más agresivos.
- Tiempo de retenciones cada vez mayores
- Saturación de la red.
- Calidad de servicio reducido
- EOL (End of Life) y EOSL (End of Support Life) del hardware actual.
- Seguridad de la información

### **1.3 PROBLEMA**

En función del aumento constante del volumen de datos que utilizan los sistemas de información de la empresa y su criticidad, se requiere disponer de una infraestructura de backup altamente confiable. El principal problema que se tiene es el fin de vida de los robots de cintas y servidores de backup y la amplia ventana de backups que esta provocando un déficit en la calidad del servicio.

Con la realización de este proyecto se pretende aportar soluciones que mejoren la infraestructura de backup de la organización sin tener en cuenta dentro de la investigación lo relativo al software utilizado.

### **1.4 OBJETO DE ESTUDIO Y CAMPO DE ACCIÓN**

El objeto que será tema de investigación es la infraestructura de backup en lo que respecta a hardware, dejando fuera de toda investigación el software de backup utilizado.

Se tendrá como parámetro de comparación el diseño actual utilizado por la empresa, lo que permitirá llegar a conclusiones más certeras sobre la factibilidad de implementación de las soluciones de backup propuestas.

Algunos conceptos a tener en cuenta que serán objeto de estudio:

- **Librerías físicas de cintas**: dispositivos robotizados que permiten la utilización de varias unidades de cintas magnéticas de una forma automatizada.
- **VTL (Virtual Tape Library)**: Son dispositivos que emulan las librerías físicas de cinta utilizando almacenamiento en disco.
- **Deduplicación**: es el proceso de identificación de datos redundantes en un volumen de información y su posterior sustitución por referencias a instancias únicas.

El campo de acción de este proyecto abarcará la investigación y el análisis de las nuevas tecnologías de hardware para la infraestructura de backup que están disponibles en el mercado y proveer soluciones que puedan ser utilizadas por la empresa.

Específicamente este trabajo se basará en el estudio de las librerías tradicionales de cinta y las VTL, conjuntamente con las técnicas de Deduplicación, cuya técnica relativamente nueva es ofrecida en la actualidad por la mayoría de los proveedores de tecnología de backup.

## **1.5 OBJETIVOS**

### **1.5.1 GENERAL**

Proponer diseños de soluciones de backup para diferentes tamaños de sitios, los cuales serán comparados con los diseños utilizados actualmente en la empresa, permitiendo de esta manera llegar a la conclusión de cuál es el que ofrece mayores ventajas a nivel operacional, económico, entre otros aspectos.

### **1.5.2 ESPECÍFICOS**

- Relevar nuevas tecnologías de infraestructura de backup
- Analizar métodos y técnicas de deduplicación
- Realizar un análisis general de las características del hardware y sus costos
- Definir métricas de comparación para la elección del hardware
- Diseñar nuevas soluciones de backup acorde al tamaño de cada sitio.
- Realizar una comparación entre la solución utilizada actualmente por la empresa versus la nueva solución propuesta.

## **1.6 IDEA A DEFENDER / PROPUESTA A JUSTIFICAR / SOLUCIÓN A COMPARAR**

Una vez realizada la investigación, se propondrán diseños de arquitecturas de backup. Éstas, permitirán hacer una comparación contra la solución utilizada actualmente en la empresa. Mediante esta comparación se intentará determinar ***que la nueva solución propuesta puede ser más adecuada o mejor que la solución utilizada por la organización en la actualidad.***

## **1.7 DELIMITACIÓN DEL PROYECTO**

El proyecto abarcará la investigación teórica, el análisis y diseño de nuevas soluciones de infraestructura de backup, quedando excluida la investigación y análisis del software de backup y la implementación de la solución propuesta.

## **1.8 APORTE TEÓRICO**

La realización de este trabajo final, permitirá:

- Adquirir conocimientos sobre las nuevas tecnologías de backup y los métodos y técnicas que éstas utilizan adecuándolas a las necesidades de un sitio de acuerdo a sus características.
- Facilitar a la empresa de nuevas ideas y tendencias para mejorar la solución de backup que está utilizando actualmente.
- Poder detectar las carencias y defectos de un sistema de backup existente.
- Realizar una comparación teórica entre dos soluciones de backup (el existente y el propuesto).

Una de las técnicas más novedosa que se está utilizando y que será desarrollada en este trabajo como uno de los aportes más importantes, es la ***Deduplicación de datos***.

Éstas técnicas son una serie de algoritmos y procedimientos computacionales que intentan localizar subconjuntos de datos repetidos (duplicados) dentro de un súper conjunto de datos, sustituirlos por referencias a la ocurrencia única de dicho subconjunto, reduciendo el espacio de almacenamiento necesario para alojar el conjunto de datos.

Las técnicas de deduplicación están ampliamente adoptadas dentro de las librerías virtuales de cintas (VTL - Virtual Tape Library) como también en otras tecnologías.

Es importante destacar que los conocimientos aplicados en la realización de éste proyecto servirán como base para nuevos y futuros proyectos independientemente del tipo

de organización con que se trabaje. Del mismo modo se podrá dar soluciones alternativas y soporte a implementaciones ya existentes.

## **1.9 APOORTE PRÁCTICO**

Al iniciar esta investigación como una necesidad de negocio de la empresa, el aporte que se le va a realizar a la compañía es muy importante, ya que este proyecto le permitirá determinar una mejor infraestructura de backup pudiendo actualizar toda la tecnología de hardware, como así también adicionar nuevas técnicas utilizadas.

Sumado al aporte brindado a la empresa cuestión de estudio, todas las organizaciones que dispongan de una infraestructura de backup, pueden hacer uso de esta investigación como una manera de pensar en una actualización de su tecnología y para aquellas que por motivos económicos, políticas internas, etc.; aún no disponen de una solución de backup, puede servir de punta pie inicial para el desarrollo futuro.

## **1.10 MÉTODOS Y MEDIOS DE INVESTIGACIÓN:**

Los métodos de investigación que se van a utilizar constan de una serie de actividades que van a tener como objetivo el diseño y administración de una solución de backup. Estas actividades son: la investigación, el análisis y la elaboración de propuestas para las nuevas soluciones.

Se utilizará como actividad de relevamiento encuestas a personal de la empresa, recopilación de información a través de Internet, contacto con diversos proveedores, observación directa, documentos y demás publicaciones de divulgación científica.

La metodología que se va a utilizar para llevar adelante este trabajo final, será a través de una serie de etapas que se realizarán de manera secuencial, teniendo la posibilidad de regresar al paso anterior en caso de requerirlo.

La figura 1.1 muestra el diagrama de 4 pasos en el proceso de investigación y diseño de una solución de backup.

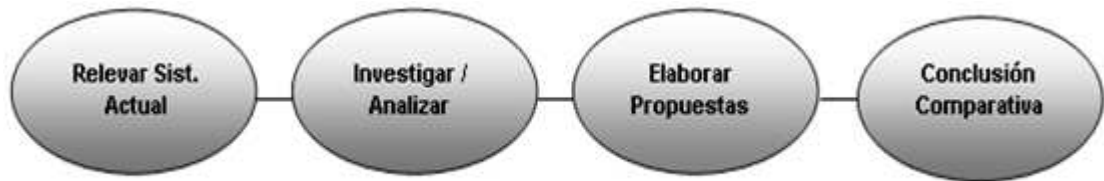


Fig. 1.1 - Esquema de las etapas de la Metodología.

En la fase de **Relevamiento de la situación Actual** se van a levantar todas las especificaciones de requerimientos y se definirá la situación actual de la solución de backup que será utilizada como punto de partida de la siguiente etapa.

En la fase de **Investigación**, se buscará información a través de Internet, de manuales, pedidos de cotización y documentación técnica a los proveedores y se analizará en forma detallada cada una de ellas.

En la fase de **Elaboración de Propuestas** se diseñarán alternativas de soluciones de backup que puedan adaptarse a las necesidades y requerimientos de la empresa.

La última etapa es la **Conclusión comparativa**, es la fase final del proceso y consiste en mostrar las conclusiones obtenidas en el proyecto.



## *Capítulo II – Marco Contextual*

---

## **CAPITULO II - MARCO CONTEXTUAL**

### **2.1 CONCEPTOS BÁSICOS**

Antes de introducirnos dentro del desarrollo de la investigación, es importante mencionar algunos conceptos básicos que serán muy utilizados a lo largo del trabajo y podrán ayudar al lector con un mínimo de conocimiento, seguir este documento sin problemas.

#### **2.1.1 BACKUP**

“Un backup en informática, es una **copia de seguridad** - o *el proceso de copia de seguridad* - con el fin de que estas copias adicionales puedan utilizarse para restaurar el original después de una eventual pérdida de datos. Fundamentalmente son útiles para dos cosas: Recuperarse de una catástrofe informática, o recuperar una pequeña cantidad de archivos que pueden haberse eliminado accidentalmente o corrompido.”<sup>1</sup>

En la actualidad existen dos técnicas para realizar el backup, las cuales optimizan el proceso de copia y recuperación de los datos. Ellas son: el backup Full + incremental y el mirror + diferencial.

#### **Full + Incremental**

Esta técnica, consta de una backup Full, también llamado completo, combinado con el backup incremental. La primera implica realizar un volcado completo de los datos al soporte de backup, mientras que la segunda solo guarda los datos que se han añadido o modificado desde el backup anterior. Para recuperar un sistema completo hará falta tener el último backup completo con todos los incrementales a partir de él, lo que hace que la restauración sea un poco más lenta.

#### **Mirror + Diferencial**

Esta otra técnica, muy similar a la anterior, es una combinación del backup completo y el diferencial. Consiste en crear una copia completa, llamada espejo, y a partir de ahí solo registrar los cambios respecto esa copia en los backups diferenciales, pero no respecto a

---

<sup>1</sup> Fuente Wikipedia

otros diferenciales. La ventaja respecto a la anterior es que para restaurar un sistema completo solo se necesita el último backup completo y el último diferencial.

## **2.1.2 DISPOSITIVOS DE ALMACENAMIENTO SECUNDARIOS**

Para guardar las copias de seguridad se necesita de dispositivos físicos que almacenen los datos. En la actualidad existe una gran variedad de dispositivos de almacenamiento secundario, pero para el fin de este apartado, se toman en cuenta específicamente los utilizados para tareas de backup a nivel empresarial. No se explica como es el funcionamiento de cada uno de los dispositivos, ya que solamente interesa conocer sus características en cuanto a tamaño, velocidad de acceso y riesgo de pérdida de información.

Los más utilizados a nivel empresarial son los discos duros y las cintas magnéticas, cuyas características se detallan a continuación.

### **2.1.2.1 Disco Duro**

Un disco duro o disco rígido es un dispositivo de almacenamiento de datos no volátil que emplea un sistema de grabación magnética para almacenar datos digitales. Es un dispositivo práctico ya que su acceso es muy rápido y permite leer y escribir datos a gran velocidad.

Las características que se deben tener en cuenta en un disco duro son:

**Tiempo medio de búsqueda:** tiempo medio que tarda el cabezal en situarse en la pista deseada; es la mitad del tiempo empleado por el cabezal en ir desde la pista más periférica hasta la más central del disco.

**Tiempo de lectura/escritura:** tiempo medio que tarda el disco en leer o escribir nueva información. Depende de la cantidad de información que se quiere leer o escribir, el tamaño de bloque, el número de cabezales, el tiempo por vuelta y la cantidad de sectores por pista.

**Latencia media:** tiempo medio que tarda el cabezal en situarse en el sector deseado; es la mitad del tiempo empleado en una rotación completa del disco.

**Tiempo medio de acceso:** tiempo medio que tarda el cabezal en situarse en la pista y el sector deseado; es la suma del Tiempo medio de búsqueda (situarse en la pista), Tiempo de lectura/escritura y la Latencia media (situarse en el sector).

**Velocidad de rotación:** revoluciones por minuto de los platos. A mayor velocidad de rotación, menor latencia media.

**Tasa de transferencia:** velocidad a la que puede transferir la información a la computadora una vez que el cabezal esté situado en la pista y sector correctos. Puede ser velocidad sostenida o de pico. Esta característica, esta ligada al tipo de conexión que tenga el disco duro. Los tipos de conexiones o interfaces son abordados en el siguiente punto.

En cuanto a la capacidad de almacenamiento, se pueden encontrar discos que superan el Terabyte. Algunos tamaños utilizados son mostrados en la tabla 2.1, donde se puede observar la capacidad indicada por los fabricantes en su formato comercial utilizando el sistema decimal y la capacidad real de los discos, calculada según el sistema binario, el cual es utilizado por el hardware y los sistemas operativos.

Capacidad Bruta/Comercial	Capacidad Real Gigabytes
160 GB	149,01 GB
250 GB	232,83 GB
320 GB	298,02 GB
500 GB	465,66 GB
1 TB	931,32 GB
1,5 TB	1396,98 GB

Tabla 2.1 - Capacidad Bruta vs. Capacidad Real en discos duros

Una desventaja que tienen los discos, se debe a que son dispositivos mecánicos y ciertamente delicados y cualquier fallo o golpe puede dañar el mismo provocando la pérdida de información. Para controlar estos fallos y aumentar la seguridad se puede optar por crear un **RAID** (Redundant Array of Independent Disks) con varios discos para poder recuperar toda la información si alguno de los discos fallara.

El método RAID es una combinación de dos o más discos duros para formar una única unidad lógica en la que se almacenan los datos de forma redundante. Para el sistema operativo aparenta ser un sólo disco duro lógico. Tiene ventajas como:

- Tolerancia a fallos (protege contra la pérdida de datos).

- Alta disponibilidad (aumenta el tiempo de funcionamiento del sistema).
- Mejora el rendimiento y la velocidad (permite el trabajo paralelo de varias unidades de discos).
- Aumenta la fiabilidad (empleando las técnicas de redundancia y bit de paridad).

Los tipos de RAID más destacados son:

#### RAID 0: Stripping

Es en realidad un falso RAID ya que solamente se trata de crear un array de 2 o más discos, pero sin ningún tipo de redundancia para evitar la pérdida de datos en caso de fallar algunos de los discos. Es más, si uno de los discos falla no se puede recuperar el array entero, de forma que, si solo fallara un disco perderíamos la información de todos los demás. Se usa únicamente para aumentar la tasa de transferencia de datos entre los discos.

#### RAID 1: Mirror

Se crea un array de varios discos y se usa un disco para contener una copia exacta de otro. Si por ejemplo, se tienen 2 discos de 100GB, al hacer RAID 1 de los dos discos se seguirá teniendo 100GB, pero en caso de que uno de los discos falle, se podrá recuperar toda la información desde el otro disco. Su seguridad es muy alta pero también su costo, ya que se dedica la mitad de los discos a la seguridad (para hacer el mirror) y se pierde la posibilidad de usarlos para almacenar datos.

#### RAID 5: Bit de Paridad

En este método se usa una parte de cada disco para almacenar información de paridad del conjunto total de discos, lo que permite recuperar toda la información ante cualquier fallo de uno de los discos mediante una operación de XOR. Tiene un 80% de aprovechamiento del tamaño total de los discos. Esta solución reduce el costo de RAID 1 y ofrece la misma tolerancia al fallo. Se necesita un mínimo de 3 discos para implementarlo.

En la tabla 2.2 se desarrolla un pequeño cuadro comparativo entre los tipos de RAID mencionados.

















Nivel RAID	Nº mín. discos	Descripción	 Ventajas  Inconvenientes
RAID 0 o Striping	2	Partición de Datos <b>sin redundancia</b> .	 Gran velocidad en las operaciones de Lectura/Escritura.  La velocidad de transferencia de datos aumenta en relación al número de discos que forman el conjunto.  <b>No hay protección de datos: si un disco falla, se pierde toda la información.</b>
RAID 1 o Mirroring	2	Se duplican todos los datos de una unidad en otra.	 Prestaciones muy altas.  Protección de datos muy alta.  <b>Costo duplicado de discos</b>
RAID 5	3	Acceso independiente con paridad distribuida.	 Muy altas prestaciones.  Muy altas protección de datos.  Soporta múltiples accesos de lectura y escritura simultaneas.  Aprovecha el 80% de la capacidad de los discos.  <b>Las prestaciones de escritura son inferiores a RAID 1 o RAID 0.</b>
Combinado RAID 0 + 1 o RAID 10	4	Combinación de RAID 0 y RAID 1.	 Prestaciones excelentes.  Protección de datos excelente (tolera fallos en varios discos).  <b>Altos costos por duplicación de disco y un mínimo de 4 unidades.</b>

Tabla 2.2 - Comparación de los diferentes tipos de RAID.

### 2.1.2.2 Cinta Magnética

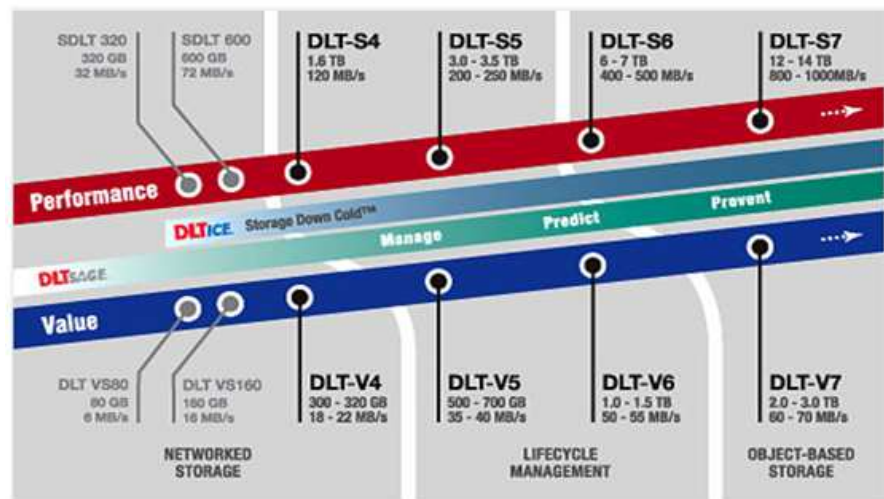
Es el medio de almacenamiento de datos mas utilizado a nivel empresarial de los últimos años. Aunque tienen una velocidad de acceso y de transferencia algo más baja que los discos duros, su capacidad es muy aceptable: actualmente alcanzan capacidades de 1.5 terabyte y hasta el doble si se utiliza la compresión. Otras ventajas que ofrecen es que son reutilizables, fácil de manejar y por sobre todas las cosas de bajo costo.

Existen varios formatos de cintas magnéticas para almacenar datos. Los dos mas utilizados y con fuerte rivalidad en el mercado actual son las cintas LTO (Linear Tape-Open) y las DLT (Digital Linear Tape), estando la primera un escalón por encima, ya que ofrece

una tasa de transferencia y una capacidad de almacenamiento superior. A continuación se presenta una pequeña reseña de cada uno de ellos.

### **DLT: Digital Linear Tape**

Tipo de cinta magnética desarrollado por la *Digital Equipment Corporation* y adquirida en 2002 por *Quantum*. Los modelos más utilizados son SDLT320 y SDLT600, que alcanzan hasta 300 GB de espacio nativo con tasas de transferencia de 36 MB/s, duplicando ambas características en el formato comprimido. El modelo más reciente es el DLT-S4 de 800 GB, fuerte rival de la LTO, aunque con una tasa de solo 60 MB/s.



All specifications assume 2:1 compression  
 Source: Quantum Corporation

Fig. 2.1 - Roadmap DLT

### **LTO: Linear Tape-Open**

Cinta magnética desarrollada originalmente a fines de 1990 por *HP*, *IBM* y *Seagate* como los estándares abiertos para las cintas magnéticas patentada que estaban disponibles en ese momento (Quantum y Sony). LTO se desarrolló como reemplazo para la DLT utilizando el mismo formato, facilitando a los proveedores de librerías de cintas para convertir sus productos en bibliotecas DLT/LTO.

Bajo el nombre LTO-Ultrium, actualmente tiene introducido en el mercado 5 modelos, del LTO1 al LTO5, siendo este último la mejor opción en cintas magnéticas con una capacidad de 1.5 terabyte de almacenamiento, el doble si se aplica la compresión, y una velocidad de transferencia de datos de 140MB/s.

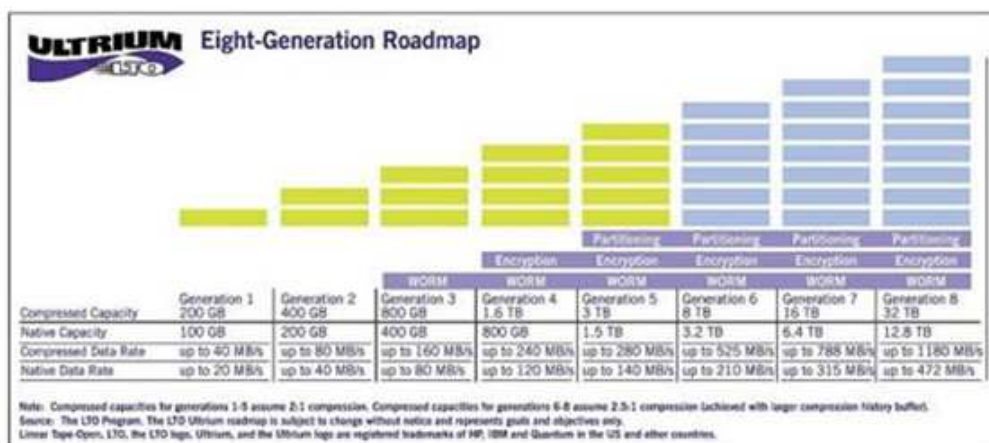


Fig. 2.2 - Roadmap LTO-Ultrium

### 2.1.3 INTERFACES DE CONEXIÓN DE DISPOSITIVOS

Para conectar dos dispositivos independientes se necesita una interfaz común y compatible que los comunique, tanto para los discos duros, servidores y dispositivos externos, como es el caso del hardware de backup.

Las diferentes interfaces, se pueden clasificar en 3 grupos: el grupo ATA, el grupo SCSI, y un tercer grupo donde se pueden incluir el USB y la fibra óptica.

#### 2.1.3.1 Grupo ATA

Sus siglas hacen referencia a **Advanced Technology Attachment** y es el bus de transferencia de datos más común para discos duros y para CD-ROM. Dentro de este grupo encontramos el PATA (paralelo) y el SATA (serie).

- **PATA: Parallel ATA.** Fue el más usado en el pasado pero actualmente está cada vez más en desuso por su baja velocidad de transferencia y su limitada capacidad de almacenamiento. Se está quedando prácticamente obsoleto a causa de los nuevos discos SATA. Dentro de este grupo se encuentran a los buses IDE, ATAPI, Ultra ATA y Ultra DMA.

- **SATA: Serial ATA.** Son la tecnología usada en estos momentos en la mayoría de discos duros. Es el doble de rápido que PATA, alcanzando unos 600MB/s, y más estrecho. Permiten desconexión en caliente de los dispositivos (**hot swap**: extracción sin necesidad de apagar o reiniciar).



### **2.1.3.2 Grupo SCSI**

**Small Computer System Interface.** Comúnmente usado para conectar dispositivos externos como librerías de backup, arrays de discos, etc., aunque también existen discos duros con conectividad SCSI, sobretodo a nivel de servidor. Orientado a los sistemas profesionales.

- **SCSI:** *SCSI paralelo.* Alcanza velocidades mucho mayores que la mayoría de los ATA. Actualmente existe el Ultra-640 SCSI, con velocidades de 640MB/s.

- **SAS:** *Serial Attached SCSI.* Es la evolución en serie del SCSI y permite velocidades algo mayores, 750MB/s. También admite la conexión de más dispositivos simultáneos y es compatible con dispositivos SATA.

### **2.1.3.3 Resto de interfaces**

Existen otras interfaces muy utilizadas y que no son incluidas dentro de los grupos anteriores. A continuación se mencionan y detallan algunas características de ellas.

- **USB:** *Universal Serial Bus.* El actual estándar de facto para la conexión de dispositivos, sobretodo para dispositivos de tamaño reducido y de uso doméstico. Aunque es muy práctico y casi todas las computadoras llevan un puerto para conectar USB, su velocidad es inferior al resto de dispositivos vistos, alcanzando un máximo de 60 MB/s en su versión 2.0.

- **Fibra óptica:** El futuro de la transferencia de datos. De momento es el bus con más velocidad disponible en el mercado, superando los 1250 MB/s.

En la tabla 2.3 se muestran diferentes tipos de interfaces y sus tasas de transferencia máximas.

Tipo de interfaz	Grupo	Tasa de transferencia máxima (MB/s)
ATA-7 ó Ultra ATA/133	ATA	133
ATA-8 ó Ultra ATA/166	ATA	166
SATA I	ATA	150
SATA II	ATA	300
SATA III	ATA	600
Ultra-160 SCSI	SCSI	160
Ultra-320 SCSI	SCSI	320
Ultra-640 SCSI	SCSI	640
SAS 300	SCSI	375
SAS 600	SCSI	750
USB 2.0	Otras	60
FC 2 Gbps	Otras	250
FC 4 Gbps	Otras	500
FC 8 Gbps	Otras	1000
FC 10 Gbps	Otras	1250

Tabla 2.3 - Tipos de interfaces y sus tasas de transferencia máxima

## **2.1.4 DISPOSITIVOS DE LECTURA / ESCRITURA**

Para la transferencia de datos de los servidores principales hacia o desde los dispositivos de almacenamiento secundarios (discos o cintas), a través de algún tipo de conexión mencionada, se necesita un dispositivo capaz de escribir y leer sobre esos dispositivos.

Existe una amplia variedad de dispositivos de lectura/escritura. Algunos se basan en cintas magnéticas y otros en discos duros. Algunos están conectados directamente al servidor de datos principal y otros al servidor de backup o almacenamiento secundario. Las interfaces de las conexiones más usadas son las Ultra-320 SCSI, así como la fibra óptica a 2 o 4 Gbps y en menor medida SAS a 3 Gbps.

### **2.1.4.1 Dispositivos basados en cinta**

#### **A- Drive**

Un drive de cinta es un aparato dedicado únicamente a leer y escribir en una cinta magnética. Es el componente más básico del sistema de backup basado en cintas que puede ir por sí solo, como se verá en este punto, pero que normalmente encontramos integrados en los otros tipos de dispositivos: los autoloaders y las librerías.

En un drive solamente cabe una cinta que se introduce manualmente cuando se necesitan copiar los datos. Es un sistema muy poco práctico si se tiene que copiar una cantidad considerable de datos y desde diversos servidores de origen.



Fig. 2.3 - Drive

## B- Autoloaders

Los autoloaders o cargadores de cintas, son el primer paso en la automatización de los backup. Están compuestos normalmente por un sólo drive y una serie de slots o compartimentos internos donde guardar las cintas, todo integrado en un solo dispositivo rackeable, es decir, que se puede montar en un rack junto con los servidores y demás dispositivos.

Con las cintas dentro, el autoloader es capaz de tomar automáticamente las que necesita de los compartimentos, ponerlas en el drive, y escribir o leer los datos que necesito y volverlas a dejar en su compartimento. De esta forma se pueden programar varios backup que copien los datos de servidores diferentes y a diferentes horas y dejar que el autoloader se encargue de montar las cintas que necesito. El usuario únicamente tiene que encargarse de cambiar las cintas de los compartimentos para que siempre haya disponible.



Fig. 2.4- Autoloader

## C- Librerías

Las librerías de backup son el paso definitivo en los sistemas de backup orientados a cintas magnéticas de carácter profesional. Funcionan básicamente igual que un autoloader pero son más grandes, con lo que permiten múltiples drives y muchos más slots para colocar cintas. Esto ofrece un gran espacio para almacenar una cantidad de datos mucho mayor, además de más autonomía para copiar datos sin necesidad de cambiar las cintas a menudo.



Fig. 2.5 - Librería

#### **2.1.4.2 Dispositivos basados en discos**

Actualmente se empiezan a desarrollar sistemas de backup basados en disco duro, que reemplazan o, sobretodo, complementan los sistemas basado en cinta. Al funcionar a través de disco duro, el proceso de copia y recuperación de datos es mucho más rápido que en los de cinta.

Estos dispositivos basados en disco suelen emular una librería a partir de un software integrado. Así se pueden organizar cintas virtuales de la misma manera que se haría con las reales, pero sin necesidad de cambiarlas, al menos hasta que haya espacio en los discos. Además, con la posibilidad de desconexión en caliente de las nuevas tecnologías de disco, se pueden insertar y extraer discos con la misma facilidad que las cintas en las librerías tradicionales.

Los dispositivos basados en disco, se suelen usar en las grandes empresas como sistema de backup de primer nivel, donde se guardan las copias más recientes y, al pasar el tiempo, se hace una copia a cinta de su contenido, para almacenar durante un periodo más largo.



Fig. 2.6 - DataDomain para backup a disco

## **2.2 ENTORNO DEL OBJETO DE ESTUDIO**

Como ya fue mencionado en el capítulo anterior, el objeto de estudio a lo largo del desarrollo de este trabajo será la **infraestructura de backup** utilizada por una empresa multinacional de origen francesa dedicada a la comercialización y fabricación de automóviles.

Esta automotriz cuenta actualmente con 75 plantas industriales distribuidas alrededor de 25 países en todo el mundo. Cada una de estas plantas tiene características particulares en relación de la producción y objetivos de venta, pero en lo que respecta a informática y tecnología deben cumplirse ciertas políticas definidas por la casa central situada en Francia.

Argentina junto con Brasil son los países de mayor importancia en Sudamérica contando con una gran infraestructura tecnológica, desde los cuales se lleva a cabo la administración de países como Chile, México, Venezuela y Colombia.

Este proyecto se enfoca en el análisis de la infraestructura de backup de Argentina, donde se cuenta con una importante infraestructura en la provincia de Córdoba y otra más pequeña en la provincia de Buenos Aires. También se analizara la infraestructura de Brasil ubicada en la ciudad de Curitiba que es la más grande y de mayor criticidad que tiene la empresa en la región de América.

## **2.3 RELACIÓN TESIS Y OBJETO DE ESTUDIO**

El desarrollo de este proyecto puede mencionarse que tiene origen a partir de dos justificaciones:

Nace dentro de la empresa como una necesidad de estudio de mejora o actualización de la infraestructura de backup que se esta utilizando en Francia como en el resto de los países. Se dividió en diferentes dominios de investigación (software, hardware, seguridad, entre otros) y responsables en todo el mundo, quedando solamente para América el proyecto de estudio del hardware.

Muchos factores a tener en cuenta a lo largo de todo el desarrollo de este trabajo, llevaron a la empresa a tomar la iniciativa para hacer esta investigación.

Los puntos más importantes son:

- Incremento exponencial año a año de la información.
- Alto costo del mantenimiento de la solución actual de backup.
- Ventanas de backup cada vez más reducidas.
- Presupuestos limitados.
- Arquitectura trabajando al 100% y sin posibilidades de ampliación.
- Puntos de recuperación (RPO) y tiempos de recuperación (RTO) más agresivos.
- Tiempo de retenciones cada vez mayores
- Saturación de la red.
- Calidad de servicio reducido.
- EOL (End of Life) y EOSL (End of Support Life) del hardware actual.
- Seguridad de la información

Por otro lado, siendo uno de los tesisistas el responsable para llevar adelante este estudio, se consideró de gran aporte realizar la presentación como Trabajo Final de Grado, el cual puede ayudar a otras empresas a conocer las diferentes tecnologías relativas a las soluciones de backup que están hoy disponibles en el mercado.

## **2.4 ANÁLISIS DE LOS PROBLEMAS OBSERVADOS**

En la actualidad y desde hace ya algunos años, la organización esta trabajando para centralizar la mayor cantidad de aplicaciones y datos de los países de Sudamérica entre Argentina y Brasil, impulsado por la competencia económica de ambos. Esto provoca que las infraestructuras disponibles en estos países vayan quedando reducidas y obsoletas si no se toman acciones inmediatamente para poder soportar la demanda que se tiene. En este aspecto, no solo la infraestructura de servidores es la única impactada, sino también la seguridad e integridad de la información. Lo que nos invita a hacer foco en la infraestructura de backup para un resguardo seguro y de alta calidad.

De acuerdo a lo expuesto en el párrafo anterior uno de los primeros problemas encontrados es el incremento exponencial de la información y el crecimiento que se tendrá con la llegada de nuevos sistemas y bases de datos.

Otro de los problemas encontrados es que el hardware utilizado, sobre todo en las librerías de backup, ha llegado a su fin de soporte de mantenimiento y su fin de vida en el año 2012. Debido a esto, los costos actuales de mantenimiento de licencias y de los insumos son demasiados elevados.

Los factores de performance y calidad se ven afectados, la infraestructura ya esta al máximo de su capacidad. Las ventanas de backup son cada vez menores, lo que hace imposible hacer el resguardo de la información fuera del horario de producción impactando así la calidad de servicio por la saturación en la red que esto provoca.

La mayoría de estos problemas mencionados conllevan a otros considerado por todas las organizaciones el más crítico, la seguridad de la información.

En el apartado 2.2 se mencionó que serán tres las infraestructuras de backup utilizadas actualmente a ser analizadas en este estudio. Ellas son la de Curitiba (Brasil), la de la provincia de Córdoba y Buenos Aires ambas de Argentina.

Antes de desarrollar las características que poseen cada uno de los sitios de estudio, se creo una tabla para clasificar los distintos escenarios por su tamaño y complejidad. Ésta permitirá definir donde se ubica cada infraestructura.

En la tabla 2.4 se puede observar las características tenidas en cuenta a la hora de clasificar el tamaño de un sitio específico. El volumen de datos a salvar, como la cantidad de filers y cantidad de servidores que son resguardados por el software Time Navigator definen el tamaño de la infraestructura de backup.

Clasificación de sitios según el tamaño y complejidad									
	Volumen de datos salvados			# Filers			# Servidores		
	< 2 TB	[2-10] TB	> 10 TB	1	[2-3]	4 ó +	[0-10]	[10-40]	40 ó +
Sitio Pequeño	✓			✓			✓		
Sitio Mediano		✓		✓	✓			✓	
Sitio Grande			✓		✓	✓			✓

Tabla 2.4 - Clasificación de sitios según el tamaño y complejidad

## **2.4.1 DESCRIPCIÓN SITUACIÓN ACTUAL**

Independientemente del tamaño de cada uno de los sitios, todos cumplen con ciertas especificaciones generales definidas por la casa matriz en Francia. A saber:

- Como hardware estándar en la infraestructura se está utilizando servidores de la marca Hewlett Packard y las librerías de cintas son las StorageTek de la empresa Sun/Oracle.
- El software para la administración de los backups se llama Time Navigator. Esta aplicación es propiedad de la empresa Atempo.
- La política de retención de los datos es de 21 días.
- Se realizan backup diarios incrementales de lunes a viernes y los fines de semana se realizan los backup full.
- Se realiza externalización de cintas de forma semanal que son resguardadas outside.

## **2.4.2 CARACTERÍSTICAS Y CONFIGURACIÓN ACTUAL DE LOS SITIOS**

### **2.4.2.1 Sitio Grande**

La primera infraestructura que se presenta está situada en Curitiba - Brasil y fue implementada en los primeros meses de 2005. En ella se realiza el resguardo de la información de los sistemas considerados más críticos para el funcionamiento del negocio de la empresa de toda América.

En la figura 2.7, se puede observar que la infraestructura cuenta con dos servidores de backup conectados a la librería de cintas magnéticas para poder soportar el resguardo de toda la información. Se debió adicionar un nuevo servidor de backup debido al importante crecimiento de los hosts clientes.



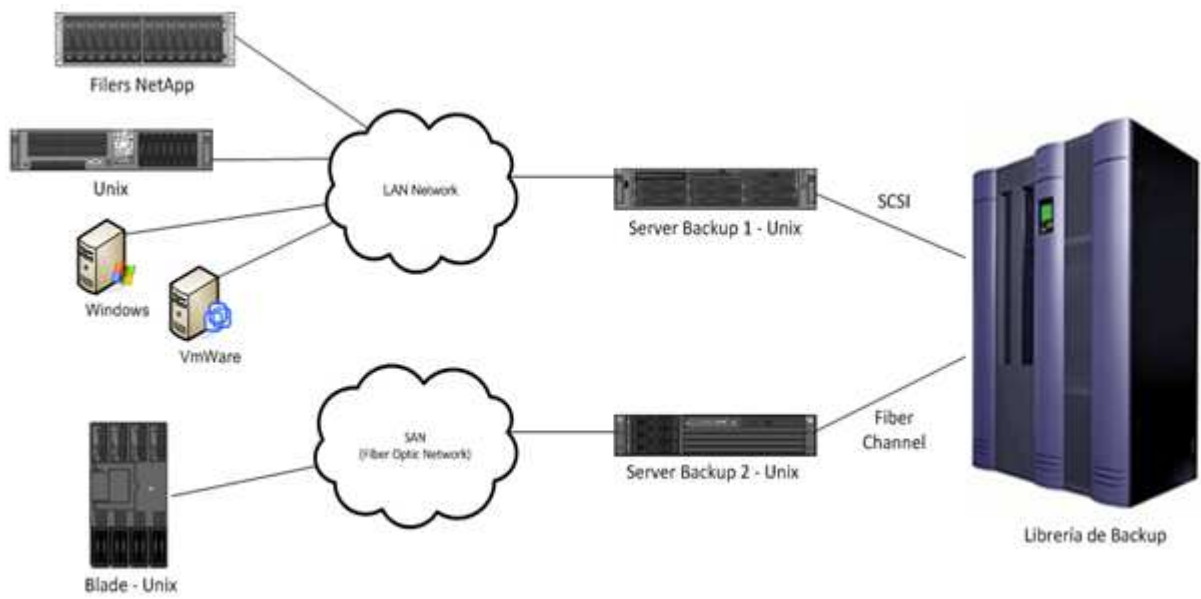


Fig. 2.7 - Diseño infraestructura actual sitio de Curitiba - Sitio Grande

En la tabla 2.5 se detallan las características de los servidores de backup.

	Servidor BKP #1	Servidor BKP #2
Modelo	HP DL380 G4	HP RP3440
Sist. Operativo	Linux RedHat Enterprise 2.4.2	HP-UX 11.23
Memoria	3 Gb RAM	6 Gb RAM
Disco	6 x 72 Gb discos	2 x 146 Gb discos
Conexión	SCSI	SCSI y FC
Unidad Rack	2U	2U
Consumo eléctrico	735 watts/hr	600 watts /hr
Hostname	tinacur	tinacas
Apl. Backup	Time Navigator 4.1	Time Navigator 4.1
# Clientes	57	12
# Catálogos	2	1
Nombre catalogo	curcat curburo	cursap

Tabla 2.5 - Descripción de los servidores de backup Curitiba

En la tabla 2.6 se detallan las características de la librería de backup.

Librería de Backup	
Modelo	StorageTek L700e Single Frame
# Drives	8 (LTO-2 SCSI) 4 (LTO-2 FC)
# Slots	678
# Mailslots	20
Capacidad Nativa	135,6 Terabytes
Picker / brazo	1 (SCSI)
Interfaz conexión	SCSI y FC
Panel Operacional	SI
Management WEB	SI
Height	182,9 cm
Depth	95,3 cm
Width	155,7 cm
Rackeable	No
Consumo eléctrico	210 watts/hr

Tabla 2.6 - Descripción de la librería de backup L700

El volumen total de datos resguardado en esta infraestructura en el plazo de una sola semana es de **67.6 Terabyte**.

En la tabla 2.7, se observa el volumen de datos y la cantidad de hosts clientes respaldados por cada catalogo.

Catalogo	# Clientes	Vol. resguardado x una semana	
		Full (Tb)	Incr. (Tb)
curcat	55 Servidores	6,8	0,9
curburo	2 Filers	5,2	0,6
cursap	12 servidores	52	2,1

Tabla 2.7 - Volumetría de datos resguardados en una semana - Sitio Grande

De acuerdo a la política de retención de datos fijada por casa central de 21 días, se calcula aproximadamente un volumen de **202.8 Terabyte** de datos guardados en cinta. La tabla 2.8 muestra a continuación los valores detallados.

Retención	Full (Tb)	Incr. (Tb)	Total (Tb)
Una semana	64	3,6	67,6
Tres semanas	192,8	10,8	202,8

Tabla 2.8 - Volumetría de datos con 21 días de retención - Sitio Grande

El software de backup utilizado se llama **Time Navigator** cuyo propietario es la empresa Atempo. Si bien este trabajo no tiene en cuenta el análisis a nivel de software, es válido mencionar los costos de licencias, ya que en gran parte se licencia el hardware utilizado. En la tabla 2.9 se detalla el costo de mantenimiento de las licencias necesarias para el hardware utilizado.

Licencia Atempo	Cantidad	€ Unid	€ Total
Servidor Backup HP-UX	1	€ 1.400,00	€ 1.400,00
Servidor Backup Linux	1	€ 650,00	€ 650,00
Drives LTO-2	12	€ 450,00	€ 5.400,00
Opción para Shared drives	12	€ 320,00	€ 3.840,00
			€ 11.290,00

Tabla 2.9 - Costo mantenimiento de licencias de hardware Curitiba<sup>2</sup>

Los problemas relevados en esta infraestructura puntual, se mencionan a continuación dividiéndolos en:

#### 1- Complejidad de la solución:

- a- Utiliza dos servidores de backup, conectados a una librería (vía SCSI y FC) para atender la cantidad actual de servidores clientes.
- b- Utilización de sistemas operativos diferentes en los servidores de backup.
- c- La cantidad de horas hombres utilizadas para la administración de la infraestructura es muy alta.

#### 2- Costos:

- a- Licencia para los dos servidores de backup adquiridas a Atempo
- b- Licencia para la utilización de 12 drives.
- c- Mantenimiento de los dos servidores de backup.
- d- Necesidad de adquirir más cintas, ya que se utiliza gran cantidad por el volumen actual de datos.

#### 3- Performance:

- a- La información mas critica de la empresa tiene un tiempo de backup de 26 horas y el tiempo de restore es de 5 días. Tiempos inaceptables para el negocio.

<sup>2</sup> Costo de mantenimiento para el año 2013

- b- El backup es realizado por la misma LAN de producción a 1000 MB/s. Esto afecta visiblemente el ancho de banda en los horarios productivos.

#### 4- Hardware:

- a- Importante cantidad de incidentes mensuales producidos por el hardware de la librería.
- b- Actualmente existen dos drives offline por problemas.
- c- La falta de mantenimiento y limpieza de los drives provocan que las cintas queden presas y se requiera intervención manual para liberarlos.

#### 5- Seguridad:

- a- Las cintas son muy viejas y no confiables. Propensas a dañarse, provocando la pérdida de los datos.
- b- Alto porcentajes de los Jobs de backup cancelados.
- c- Imposibilidad de cumplir con los 21 días de retención de la información según las políticas de la compañía.

#### 6- Mantenimiento:

- a- Librería L700 sin mantenimiento.
- b- Servidor de backup numero 1 llego a fin de vida y esta fuera de soporte y mantenimiento.

#### 7- Escalabilidad:

- a- Solución trabajando al 100 %.

### **2.4.2.2 Sitio Mediano**

Como caso de estudio para los sitios de tamaño mediano, es la infraestructura situada en la provincia de Córdoba - Argentina. Ésta fue puesta en producción a fines del 2005 y se mantiene en la actualidad. En la figura 2.17 se muestra la infraestructura utilizada.

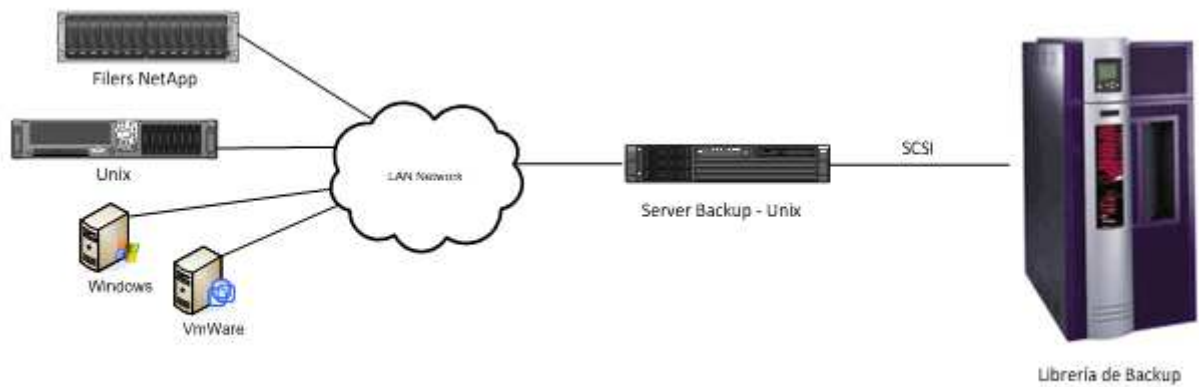


Fig. 2.8 - Diseño infraestructura actual sitio de Córdoba - Sitio Mediano

En la tabla 2.10 se detallan las características del servidor de backup.

	Servidor Backup
Modelo	HP RP3440
Sist. Operativo	HP-UX 11.11
Memoria	6 Gb RAM
Disco	2 x 300 Gb discos
Conexión	SCSI
Unidad Rack	2U
Consumo eléctrico	600 watts /hr
Hostname	tinacor
Apl. Backup	Time Navigator 4.1
# Clientes	38
# Catálogos	3
Nombre catalogo	curcat1 curburo1 curiao1

Tabla 2.10 - Descripción del servidor de backup Córdoba

En la tabla 2.11 se detallan las características de la librería de backup.

Librería de Backup	
Modelo	StorageTek L180
# Drives	6 LTO-2 SCSI
# Slots	174
# Mailslots	10
Capacidad Nativa	34,8 Terabytes (descomprimido)
Picker / brazo	1 - SCSI
Interfaz conexión	SCSI
Panel Operacional	SI
Management WEB	SI
Height	165,4 cm
Depth	125,2 cm
Width	71,9 cm
Rackeable	No
Consumo eléctrico	120 watts/hr

Tabla 2.11 - Descripción de la librería de backup L180

El volumen total de datos resguardado en esta infraestructura en el plazo de una sola semana es de **9 Terabyte**.

En la tabla 2.12, se observa el volumen de datos y la cantidad de hosts clientes respaldados por cada catalogo.

Catalogo	# Clientes	Vol. resguardado x una semana	
		Full (Tb)	Incr. (Tb)
corcat	36 Servidores	3,2	0,4
corburo	1 Filer	1,9	0,2
coriao	1 Filer	2,8	0,5

Tabla 2.12 - Volumetría de datos resguardados en una semana - Sitio Mediano

De acuerdo a la política de retención de datos, fijada por casa central de 21 días se calcula aproximadamente un volumen de **27 Terabyte** de datos guardados en cinta. La tabla 2.13 muestra a continuación los valores detallados.

Retención	Full (Tb)	Incr. (Tb)	Total
Una semana	7,9	1,1	9
Tres semanas	23,7	3,3	27

Tabla 2.13- Volumetría de datos con 21 días de retención - Sitio Mediano

En la tabla 2.14 se muestra el costo de mantenimiento a nivel de software de backup.

Licencia Atempo	Cantidad	€ Unid	€ Total
Servidor Backup HP-UX	1	€ 1.400,00	€ 1.400,00
Drives LTO-2	6	€ 450,00	€ 2.700,00
Opción para Shared drives	6	€ 320,00	€ 1.920,00
			€ 6.020,00

Tabla 2.14 - Costo mantenimiento de licencias de hardware Córdoba<sup>3</sup>

Los problemas relevados en esta infraestructura puntual, se mencionan a continuación dividiéndolos en:

1- Costos:

- a- Mantenimiento del servidor de backup.
- b- Licencia de Sistema Operativo (HP-UX) es muy elevada.
- c- Necesidad de adquirir más cintas, ya que se utiliza gran cantidad por el volumen actual de datos.

2- Performance:

- a- El backup es realizado por la misma LAN de producción a 100 MB/s. Esto afecta visiblemente el ancho de banda en los horarios productivos.

3- Hardware:

- a- Importante cantidad de incidentes mensuales producidos por el hardware de la librería.
- b- Actualmente existe un drive offline por problemas.
- c- La falta de mantenimiento y limpieza de los drives provocan que las cintas queden presas y se requiera intervención manual para liberarlos.

4- Seguridad:

- a- Las cintas son muy viejas y no confiables. Propensas a dañarse, provocando la pérdida de los datos.
- b- Alto porcentajes de los Jobs de backup cancelados.
- c- Imposibilidad de cumplir con los 21 días de retención de la información según las políticas de la compañía.

5- Mantenimiento:

- a- Librería L180 sin mantenimiento.

<sup>3</sup> Costo de mantenimiento para el año 2013

- b- Servidor de backup llego a su fin de vida y esta fuera de soporte y mantenimiento.

6- Escalabilidad:

- a- Solución trabajando al 100 %.

### 2.4.2.3 Sitio Pequeño

Para los sitios de tamaño pequeño, se toma como estudio la infraestructura situada en la provincia de Buenos Aires - Argentina. Ésta solución también fue puesta en producción a fines del 2005 y se mantiene en la actualidad. En la figura 2.9 se muestra la infraestructura utilizada.

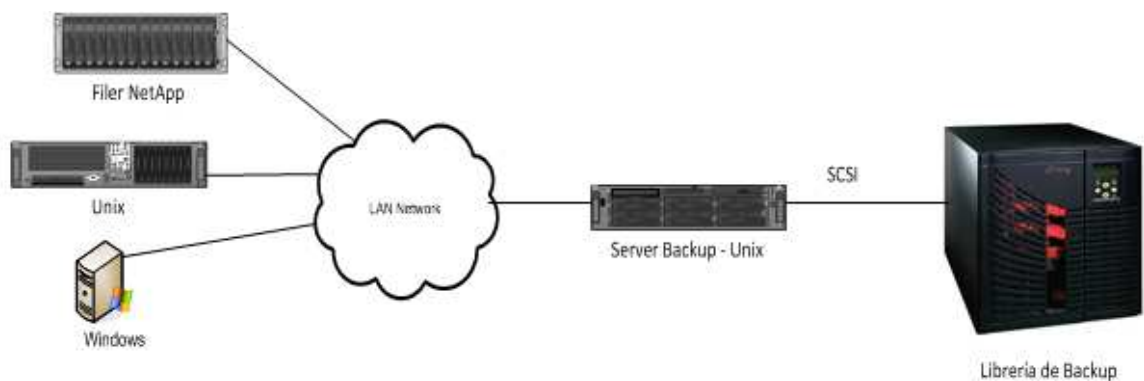


Fig. 2.9 - Diseño infraestructura actual sitio de Buenos Aires - Sitio Chico

En la tabla 2.15 se detallan las características del servidor de backup.



Servidor Backup	
Modelo	HP DL380 G4
Sist. Operativo	Linux RedHat Enterprise 2.4.2
Memoria	2 Gb RAM
Disco	2 x 72 Gb discos
Conexión	SCSI
Unidad Rack	2U
Consumo eléctrico	735 watts/hr
Hostname	tinabsa
Apl. Backup	Time Navigator 4.1
# Clientes	4
# Catálogos	1
Nombre catalogo	bsaburo1

Tabla 2.15 - Descripción del servidor de backup Bs. As.

En la tabla 2.16 se detallan las características de la librería de backup.

Librería de Backup	
Modelo	StorageTek L40
# Drives	1 LTO-2 SCSI
# Slots	24
# Mailslots	1
Capacidad Nativa	8 Terabytes (descomprimido)
Picker / brazo	1 - SCSI
Interfaz conexión	SCSI
Panel Operacional	SI
Management WEB	SI
Height	44,5 cm
Depth	75,6 cm
Width	48,3 cm
Rackeable	SI (10U)
Consumo eléctrico	180 watts/hr

Tabla 2.16 - Descripción de la librería de backup L40

En la tabla 2.17, se observa el volumen de datos y la cantidad de hosts clientes respaldados en esta infraestructura.

Catalogo	# Clientes	Vol. resguardado x una semana	
		Full (Tb)	Incr. (Tb)
bsaburo	1 filer + 3 servidores	1,8	0,3

Tabla 2.17 - Volumetría de datos resguardados en una semana - Sitio Chico

De acuerdo a la política de retención de datos, fijada por casa central de 21 días se calcula aproximadamente un volumen de **6.3 Terabyte** de datos guardados en cinta. La tabla 2.18 muestra a continuación los valores detallados.

Retención	Full (Tb)	Incr. (Tb)	Total
Una semana	1,8	0,3	2,1
Tres semanas	5,4	0,9	6,3

Tabla 2.18- Volumetría de datos con 21 días de retención - Sitio Chico

En la tabla 2.19 se muestra el costo de mantenimiento a nivel de software de backup.

Licencia Atempo	Cantidad	€ Unid	€ Total
Servidor Backup Linux	1	€ 650,00	€ 650,00
Drives LTO-2	6	€ 450,00	€ 450,00
			€ 1100,00

Tabla 2.19 - Costo mantenimiento de licencias de hardware Bs. As.<sup>4</sup>

Los problemas relevados en esta infraestructura puntual, se mencionan a continuación dividiéndolos en:

1- Costos:

- a- Mantenimiento del servidor de backup.
- b- Necesidad de adquirir más cintas debido al incremento de los datos.

2- Hardware:

- a- Importante cantidad de incidentes mensuales producidos por el hardware de la librería.
- b- La falta de mantenimiento y limpieza de los drives provocan que las cintas queden presas y se requiera intervención manual para liberarlos.

3- Seguridad:

- a- Las cintas son muy viejas y no confiables. Propensas a dañarse, provocando la perdida de los datos.
- b- Alto porcentajes de los Jobs de backup cancelados.
- c- Imposibilidad de cumplir con los 21 días de retención de la información según las políticas de la compañía.

<sup>4</sup> Costo de mantenimiento para el año 2013

#### 4- Mantenimiento:

- a- Librería L40 sin mantenimiento.
- b- Servidor de backup llego a su fin de vida y esta fuera de soporte y mantenimiento.

Como fue mencionado con anterioridad, las librerías de backup han llegado a su fin de vida y no se dispone mas de soporte por parte del proveedor. Por este motivo, no se especifica con mayor detalle las características técnicas que ofrecen, ya que sin dudas deben ser reemplazadas en la nueva solución de backup.

En los apartados siguientes se hacen estudios de las nuevas tecnologías ofrecidas en el mercado, donde se aborda con mayor detalle las especificaciones técnicas de las mismas.

## *Capítulo III – Marco Teórico*

---

## **CAPITULO III – MARCO TEÓRICO**

En este capítulo, se aborda con mayor detalle algunos conceptos teóricos sobre las **Librerías de Cintas Tradicionales**, como también de las **Librerías Virtuales (VTL)**. Cuales son sus características, que ventajas y desventajas tienen, como funcionan, que tecnología utilizan, entre otras preguntas, serán respondidas a lo largo de este apartado.

Un punto muy fuerte que también se desarrollará y es sin dudas uno de los aspectos fundamentales a la hora de hablar de soluciones de backup en la actualidad, es el concepto de Deduplicación. Se estudiarán las técnicas, los tipos y los algoritmos utilizados.

Por último, se definen métricas que permitirán comparar los diferentes factores específicos para la elección de la solución de backup propuesta más conveniente. Para esto se diseña una matriz de ponderación de los factores a tener en cuenta según la importancia de los mismos.

### **3.1 LIBRERÍAS DE CINTAS TRADICIONALES**

Se comenzará estudiando en primera medida las librerías tradicionales de cintas que son ampliamente utilizadas desde hace muchos años por grandes cantidades de empresas.

Las librerías de cintas, son dispositivos robotizados que permiten la utilización de varias unidades de cintas magnéticas de una forma automatizada.

Existe una gran variedad de modelos de estos dispositivos ofrecidos por diversos fabricantes, como Hewlett Packard, IBM y SUN (adquirida hace un tiempo por la empresa Oracle) que son los más conocidos y/o utilizados por las empresas. En el **Anexo A** se pueden observar los modelos comercializados en la actualidad por estos fabricantes.

Entre las características ofrecidas por los fabricantes sobre las librerías de cintas, se verifican algunas en común como:

- Utilizan drives LTO Ultrium en las versiones LTO-5, LTO-4 y LTO-3.
- Puerto Ethernet para la administración remota, facilitando la administración y control de la librería.

- Encriptación de datos para la protección contra el acceso no autorizado.
- Permiten la rápida expansión de la librería aumentando la capacidad y rendimiento satisfaciendo la demanda de datos.
- Compatibles con todas las interfaces de conexión (SCSI, SAS, Fiber Channel).
- Compatible con amplia gama de Sistemas Operativos y software de backup.

### **3.1.1 VENTAJAS Y DESVENTAJAS GENERALES**

#### **3.1.1.1 Ventajas**

- Menor costo de adquisición que las librerías basadas en discos duros.
- Operaciones con diversos tipos de cintas.
- Permite la gestión remota vía Web.
- Aumento de la capacidad de almacenamiento bajo demanda.
- La cinta magnética es todavía necesaria para fines de archivo a largo plazo, y para el cumplimiento regulatorio.
- Las unidades de cintas (drives) tienen una utilización mínima de energía eléctrica cuando no están en uso.

#### **3.1.1.2 Desventajas**

- Escritura secuencial en las cintas.
- Tiempos globales de las operaciones afectados por las operaciones de robótica y posicionamiento de la banda.
- Importante número de errores físicos (debido a las operaciones mecánicas).
- Las cintas no son confiables y sensibles a las condiciones ambientales (calor, humedad, etc.).
- Requiere mayor ventana de backup.
- Requiere mayor espacio de almacenamiento.

### **3.2 LIBRERÍAS DE CINTAS VIRTUALES (VTL)**

Si bien las librerías tradicionales de cintas son muy utilizadas, desde hace algunos años se comenzó a escuchar sobre una nueva tecnología dentro de las soluciones de resguardo de los datos, las **Librerías de cintas Virtuales**.

Las VTL son sistemas de almacenamiento a disco, con un software incorporado capaz de emular el funcionamiento de las librerías tradicionales, organizando los datos en cintas virtuales (con varios formatos compatibles, incluido LTO), con sus slots y sus drives también virtuales. Para el usuario final, a través del software gestor de backups, la impresión es la de estar usando una librería normal de cinta, pero con las ventajas de los discos duros.

Los diferentes fabricantes compiten fuertemente en el desarrollo de herramientas, componentes y técnicas para diferenciarse uno de otro. En el **Anexo B**, se muestran los productos ofrecidos por los fabricantes Hewlett Packard, SUN/Oracle, EMC e IBM.

Las características comunes entre los fabricantes que se destacan a nivel general para el resguardo de la información en disco son:

- Fácil integración con infraestructuras existentes.
- Compatibilidad con diferentes plataformas de sistemas operativos y software de backup.
- Ofrecen software de Deduplicación para la optimización de recursos.
- Posibilidad de replicación.
- Administración centralizada de diferentes sitios.
- Compatibles con todas las interfaces de conexión (SCSI, SAS, FC).
- Encriptación de datos.

## **3.2.1 VENTAJAS Y DESVENTAJAS GENERALES**

### **3.2.1.1 Ventajas**

- Mayor velocidad en la realización del backup y el restore.
- Ahorro de espacio en disco (compresión y deduplicación).
- Mayor facilidad de gestión, monitorización y mantenimiento.
- Mayor escalabilidad en la infraestructura de backup.
- Mayor fiabilidad y reducción de tiempos de fallo.
- Ahorro de costos como resultado final.
- Posibilidad de integración con los sistemas de backup actuales. (Puede conectarse junto a una librería física tradicional).
- Evita manipulación de cintas y reduce el costo de las mismas.
- Permite mayor número de grabaciones.
- Permite la replicación.
- Pueden emular librerías y drives en gran cantidad.
- Mayor ejecución de Jobs de backup en simultáneo.
- Reducción de las ventanas de backup.

### **3.2.1.2 Desventajas**

- Altos costos de adquisición. Requiere una importante inversión.
- Alto consumo de procesador en los métodos de deduplicación.
- No permite la posibilidad de una retención de datos de largo tiempo. No se puede cumplir con regulaciones legales que requieren de este punto.
- Amplio consumo de energía. Los discos consumen energía todo el tiempo por más que no se estén haciendo backup.



### **3.3 LIBRERÍAS FÍSICAS VS. LIBRERÍAS VIRTUALES**

En la tabla 3.1 se resumen algunas características a nivel general que tienen las librerías de cintas tradicionales como las librerías Virtuales.

Ítem	Librería tradicional	Librería Virtual
Costo Inversión	Bajo	Alto
Performance Bkp/Restore	Media	Alto
Deduplicación	No	Si
Escalabilidad	Si	Si
Replicación	No	Si
Externalizarían de cintas	Si	No
Compresión	No	Si
Fiabilidad	Media	Alta
Interfaz de conexión	SCSI/SAS/FC	SAS/FC
Compatibilidad Software	Si	Si
Administración Remota	Web	Web
Retención de datos de largo plazo	SI	SI
Consumo de energía	Bajo	Alto
Tipo de escritura	Secuencial	Aleatorio
Operaciones Mecánicas	Alta	Bajo
Encriptación de datos	No	Si
Ventana de backup	Amplia	Reducida

Tabla 3.1 - Librerías Tradicionales vs. Librerías Virtuales de cintas.

### **3.4 DEDUPLICACIÓN**

Si bien la Deduplicación de datos es un pequeño aspecto dentro de un sistema de backup en toda su dimensión, vale la pena dedicar un desarrollo específico dentro de éste capítulo, debido a la importancia que tiene dentro de cualquier solución de backup y es un punto que los fabricantes están mejorando día a día.

#### **3.4.1 ¿QUE ES LA DEDUPLICACIÓN DE DATOS?**

El organismo internacional *SNIA (Storage Network Industry Association)*, define la **Deduplicación** como el proceso de identificación de datos redundantes en un volumen de información y su posterior sustitución por referencias únicas, de forma que ocupen mucho menos espacio.

En la figura 3.1 se muestra gráficamente como son reemplazados los bloques repetidos dentro de un segmento de datos por referencias a la ubicación de los mismos.

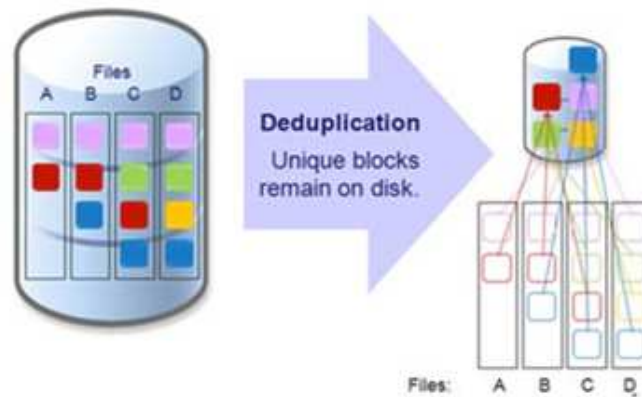


Fig. 3.1 - Deduplicación

Todos los sistemas de deduplicación constan de los siguientes componentes:

- a- **Repositorio:** es el conjunto de segmentos únicos de datos. Se aloja en disco.
- b- **Índice:** es un mapa binario de la localización en el repositorio de aquellos segmentos únicos.

El aspecto más importante que determina el rendimiento de la técnica de deduplicación es si el índice puede alojarse por completo en la memoria RAM del sistema de deduplicación.

En la figura 3.2 se muestra un ejemplo simple de cómo funciona la deduplicación. En este caso se comparan los datos de dos días diferentes (día 1 y día 2) y se almacenan en destino solamente los datos del día 1 y los datos del día 2 que no se encuentran en el día 1.

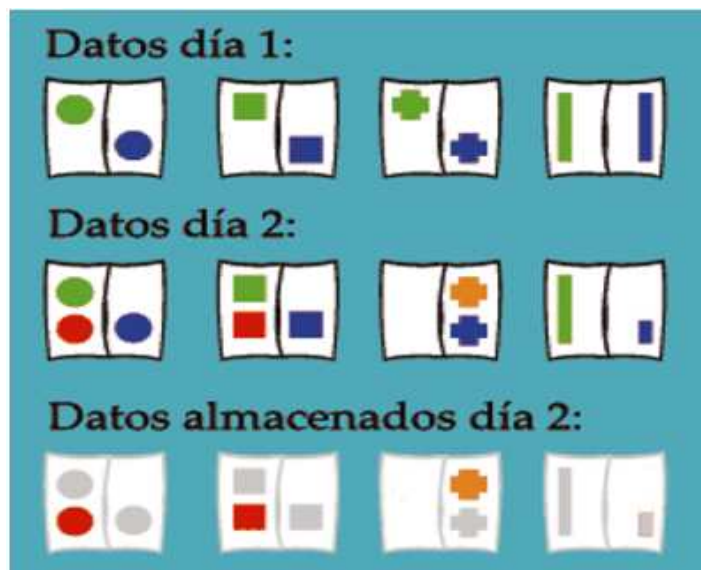


Fig. 3.2 - Ejemplo de Deduplicación

Para restaurar los datos del día 2 basta con disponer las referencias de los datos del día 1 y los datos únicos del día 2.

Esta es una simplificación del proceso para su fácil comprensión, puesto que una tecnología de deduplicación es un proceso más complejo que el mostrado en este ejemplo.

El **objetivo de la Deduplicación**, consiste en almacenar elementos de datos únicos solo una vez, y al mismo tiempo poder reconstruir todo el contenido en su formato original según demanda, con un 100% de confiabilidad a velocidades de disco.

### **3.4.2 RATIO DE DEDUPLICACIÓN**

Uno de los aspectos que caracterizan las técnicas de deduplicación es el denominado **Ratio de Deduplicación o Ratio de Reducción de Espacio**.

En general, se define el Ratio de Deduplicación como la relación entre las cantidades de bytes de entrada frente a la cantidad de bytes de salida de dicho proceso, como se muestra en la figura 3.3.

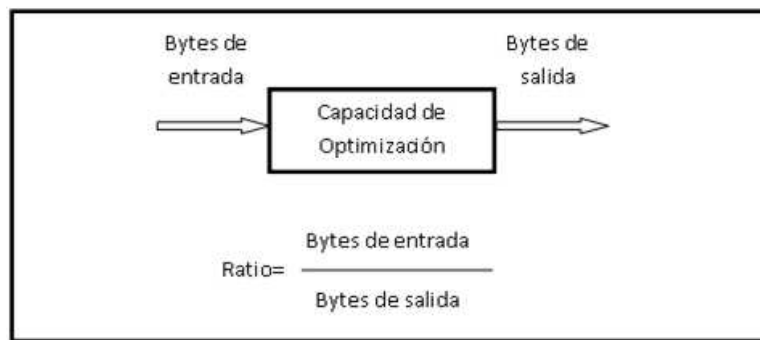


Fig. 3.3 - El Ratio de Reducción de espacio

Este ratio se expresa generalmente en términos **n:1** e indica que el espacio necesario para almacenar el conjunto de datos deduplicados es “**n**” veces menor que el necesario para almacenar ese mismo conjunto de datos sin deduplicar. Algunos ejemplos son 10:1 y 20:1).

También puede utilizarse el porcentaje de reducción de espacio, definido mediante la expresión **Porcentaje = 1 - (1/Ratio)**, donde se indica que el porcentaje de reducción es 1 menos la inversa del ratio de reducción.

Si tabulamos diferentes valores del ratio y su porcentaje asociado, podremos comprobar como relativamente pequeños ratios de deduplicación conllevan grandes reducciones de espacio alcanzándose un punto a partir del cual, grandes aumentos del ratio producen pequeños incrementos del porcentaje de reducción. Este comportamiento se observa en la tabla 3.2, donde se puede verificar que el incremento en el porcentaje de reducción de espacio entre los ratios 100:1 y 500:1 es de 0.8% (99% a 99,8%).

Ratio de reducción	Porcentaje de Reducción = $1 - (1/\text{ratio de reducción})$
2:1	$1/2 = 50\%$
5:1	$4/5 = 80\%$
10:1	$9/10 = 90\%$
20:1	$19/20 = 95\%$
100:1	$99/100 = 99\%$
500:1	$499/500 = 99,8\%$

Tabla 3.2 - Relación ratio vs porcentaje de reducción.

En la figura 3.4 puede comprobarse el crecimiento lento de la curva a partir de valores de ratios de reducción superiores a 5:1

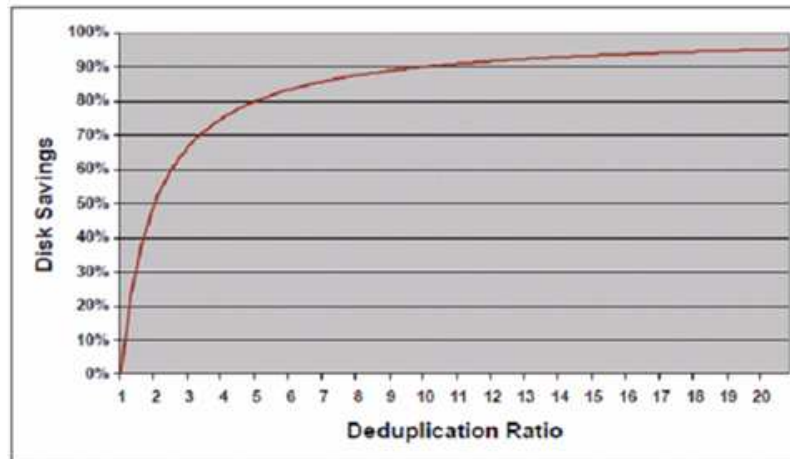


Fig. 3.4 - La curva de ahorro de espacio.

Existen factores que influyen en el ratio de Deduplicación. Ellos son:

- **Las políticas de backup de datos:**

A mayor frecuencia de backup Full, mayor porcentaje de deduplicación, ya que habrá datos redundantes de un día para otro.

- **Ajustes de conservación de los datos:**

Cuanto más tiempo se conserven los datos en disco, mayor oportunidad tendrá el motor de deduplicación de encontrar redundancias.

- **Tipo de datos:**

Algunos datos son intrínsecamente más proclives a la duplicación que otros. Archivos como imágenes o videos tienen bajo porcentaje de deduplicación.

- **Tasa de variación:**

A menor ratio de variación, mayor probabilidad de encontrar datos duplicados.

- **Dominio de deduplicación:**

Cuanto más amplio es el ámbito del proceso de inspección y comparación, mayor es la probabilidad de detección de datos duplicados.

### **3.4.3 BENEFICIOS**

Los beneficios que las empresas pueden encontrar con la utilización de la deduplicación son:

- Optimiza el consumo de energía que se tiene en las organizaciones, pues se simplifica el proceso de respaldo y recuperación.
- Optimiza los procesos de ventanas de respaldo y mejora el cumplimiento de servicio a sus usuarios o clientes finales.
- Optimización de recursos y transferencia de conocimiento.
- Permite cumplir y estar preparados para el crecimiento exponencial de la información sin incurrir en un exagerado costo de energía o espacio.
- Permite periodos más largos de retención de los datos.
- Mejora la utilización de recursos de red (ancho de banda) para el resguardo de la información.
- Mayor seguridad y administración optimizada. Se eliminan los riesgos, los retrasos y los costos asociados con el manejo y almacenamiento de cintas físicas.

#### **3.4.4 VENTAJAS**

Algunos puntos que justifican la utilización de Deduplicación son mencionados a continuación.

##### **a- *Oficinas remotas***

Puede ayudar a responder ante la situación donde existen oficinas remotas que carecen de personal local formado para administrar los backup. La utilización de una librería virtual por ejemplo con deduplicación para almacenar datos de backup, eliminará la necesidad de garantizar que haya siempre una cinta disponible y suprimirá el requisito de disponer de una persona que monte una cinta para las restauraciones.

Si a eso se le suma la capacidad de replicar datos deduplicados en la red de área amplia (WAN), tenemos una solución de resguardo que conlleva pocas tareas de administración general. Además, la replicación de datos deduplicados en la WAN reduce las

necesidades de ancho de banda de la red, lo cual la convierte en una alternativa más barata que los discos en espejo.

#### ***b- Deduplicación de datos y archivos duplicados***

Los entornos con grandes cantidades de archivos duplicados o similares tienen mucho que ganar con la reducción de costos de almacenamiento. Los mejores resultados de reducción de datos de la deduplicación se obtienen cuando el proceso encuentra grandes volúmenes de segmentos de datos idénticos. En los casos en que se realizan backup completos con frecuencia y los datos cambian a un ritmo entre moderado y lento, la reducción de datos puede ser muy impresionante, y desembocar en ahorros significativos en materia de almacenamiento. Los ratios de reducción entre 5:1 y 10:1 no son infrecuentes, pero en algunos entornos se han observado ratios de 20:1 y más.

#### ***c- Reducción de la manipulación de cintas***

En los entornos que siguen necesitando operadores de cintas y racks para almacenarlas porque la librería de cintas roza su capacidad máxima, la deduplicación ofrece una gran oportunidad de reducir la manipulación de éstas, y permite destinar los recursos a otras áreas que los requieran.

#### ***d- Recuperación de espacio***

En vista del costo del espacio de los centros de datos, es muy sensato recuperar parte del espacio ocupado por una librería de cintas muy grande y sustituirla por algún backup a discos con función de deduplicación que ocupan menos superficie.

#### ***e- Mejora de las cintas***

Cualquier organización que se plantee una mejora de su tecnología de cinta debería considerar seriamente la deduplicación en disco. Aunque retirar y sustituir un subsistema de cinta que sigue satisfaciendo los requisitos no es forzosamente ventajoso desde el punto de vista financiero, la necesidad de actualizar la tecnología siempre brinda una oportunidad de evaluar otras opciones.

### **3.4.5 DESVENTAJAS**

Algunos puntos en contra que tiene la Deduplicación son los siguientes:

#### ***a- Tipo de datos***

No todos los datos se prestan a la deduplicación: los archivos de imágenes, vídeo y audio, u otros tipos de datos comprimidos disminuyen poco con la deduplicación.

#### ***b- Cifrado***

Para las entidades preocupadas por la seguridad que implantan el cifrado de datos en origen, la deduplicación en el marco de la salvaguarda no es la mejor opción, pues la primera función del cifrado es que los datos resulten irreconocibles sin las claves. Esto anula la mayor parte de las ventajas de la deduplicación, a menos que el cifrado se realice después de la deduplicación.

#### ***c- Datos efímeros***

Los datos con parámetros de retención muy bajos normalmente arrojan ratios de deduplicación muy reducidos. Esto se debe a que para ser eficaz, la deduplicación tiene que desarrollar una base de segmentos de datos idénticos. Los datos de paso o retenidos a muy corto plazo normalmente no residen en los discos de almacenamiento el tiempo suficiente para permitir que los algoritmos de deduplicación desarrollen una hipótesis. La deduplicación es sin lugar a dudas más adecuada para la retención a largo plazo.

### **3.4.6 TÉCNICAS DE DEDUPLICACIÓN**

Existen diversas técnicas de deduplicación que se clasifican de acuerdo a diferentes criterios. En los siguientes puntos se detalla cada una de éstas.

#### **3.4.6.1 Por Nivel de Análisis (File o Sub-File)**

Según el nivel al que se realice el análisis de redundancia, se puede hablar de técnicas de deduplicación con menor o mayor granularidad. Existen dos niveles de análisis posibles, por *Nivel de Archivos* y por *Nivel de Bloques*.



### 3.4.6.1.1 Nivel de archivos (File)

Las técnicas que utilizan esta metodología, analizan el conjunto de datos (por ejemplo filesystems) calculando la clave hash de cada archivo y detectando duplicación cuando detectan una clave hash existente. El resultado de esta técnica es que solo se guarda un solo archivo y un indicador al mismo para cada copia encontrada. Esta técnica se conoce también como SIS (Single Instance Store) o almacenamiento de instancia única.

### 3.4.6.1.2 - Nivel de bloques (Sub-File)

Otras técnicas de deduplicación trabajan a nivel inferior al archivo (sub-file) y dividen el conjunto de datos en segmentos de longitud fija o variable. Los segmentos suelen ser del orden de 4K o inferiores. Cuanto menor sea el tamaño del segmento, mayores son las posibles ganancias de la deduplicación puesto que la probabilidad de encontrar redundancia de segmentos pequeños es mayor que con segmentos grandes.

Entre la deduplicación por archivos y por bloques, se pueden encontrar las siguientes diferencias:

**1)-** La deduplicación a nivel de archivo puede ser menos eficiente que a nivel de bloques, ya que un cambio en el archivo provoca que se vuelva a salvar todo el archivo. A diferencia, la deduplicación por Bloques sólo guardaría los bloques modificados entre una versión del archivo y la siguiente.

Por ejemplo, en un archivo PowerPoint, puede cambiar algo tan sencillo como el encabezamiento de la página para reflejar un cambio de fecha o de presentador y eso hará que se vuelva a guardar todo el archivo por segunda vez y no solo el bloque modificado.

Las ratios de reducción pueden ser del orden del 5:1 o menos solamente a nivel de archivos, mientras que la deduplicación por bloques ha conseguido reducir el volumen de datos guardados del orden del 20:1 al 50:1.

**2)-** La deduplicación a nivel de archivo puede resultar más eficiente que a nivel de bloques, ya que los índices de deduplicación por archivos son significativamente menores. Esto permite disminuir el tiempo de procesamiento para localizar los archivos

duplicados, mejorando considerablemente el rendimiento del backup y la recuperación de los datos.

En cambio la deduplicación a nivel de bloques, exige volver a ensamblar los fragmentos basándose en el índice maestro que contiene los segmentos únicos y los indicadores que remiten a estos segmentos.

Como los enfoques basados a nivel de archivos almacenan archivos únicos e indicadores que apuntan a dichos archivos, hay menos necesidad de re-ensamblaje.

**3)-** La deduplicación a nivel de bloques tiene ciertas desventajas a considerar.

**a-** La utilización de un algoritmo de Hash para calcular el ID único del segmento introduce el riesgo de generar colisiones Hash provocando un falso positivo cuando dos segmentos diferentes tienen el mismo ID. Aunque las colisiones de Hash y la corrupción de datos resultantes es una posibilidad real, su incidencia es muy reducida.

**b-** Almacenar IDs únicos en un índice puede demorar el proceso de inspección a medida que aumenta su tamaño y requiere mayor I/O de disco (a menos que se controle el tamaño del índice y la comparación de datos se realice en la memoria).

### **3.4.6.2 Por Algoritmo**

En cuanto al algoritmo matemático subyacente sobre el que se basan, las técnicas de deduplicación pueden dividirse en *técnicas basadas en Hashing* y *técnicas basadas en diferencias binarias*.

#### **3.4.6.2.1 Técnicas basadas en Hashing**

Las técnicas basadas en algoritmos de Hashing (principalmente basadas en los conocidos algoritmos MD5 y SHA-1 ó SHA-2) consisten en la división del conjunto de datos en segmentos de longitud fija o variable y en el cálculo de un valor hash de cada uno de ellos, detectando redundancia cuando dos segmentos presentan el mismo valor del hash calculado.

Estas técnicas presentan dos problemas a resolver:

- Rendimiento vs Capacidad
- Integridad

### A- Rendimiento vs Capacidad

Como se ha indicado, las técnicas basadas en hashing consisten en la detección de aquellos segmentos para los que el hash calculado es igual al de algún segmento ya presente en el repositorio.

Según este planteamiento, es necesario que el índice tenga para cada segmento único un registro que contenga el valor del hash así como un puntero direccionando al segmento dentro del repositorio.

En la tabla 3.2 se muestra para diferentes volúmenes de repositorios, el tamaño del índice necesario, utilizando valores de segmentos de 8Kb.

El tamaño de índice se calcula según la expresión:

$$\text{Tamaño índice} = (\text{tamaño hash} + \text{tamaño puntero}) * \text{número de segmentos}$$

El valor de hash utilizado es el que utiliza SHA-1 (20 bytes). El tamaño de puntero tiene en cuenta que el número máximo de direcciones con 32 bits es de 4.294.967.296 ( $2^{32}$ ) por lo que para mapear un repositorio de 50TB requeriremos punteros de 5 bytes en lugar de 4. Es decir que a medida que se incrementa la cantidad de segmentos, el tamaño del puntero deberá ser mayor.

Como se puede observar en la tabla 3.3, el aumento en el volumen del repositorio requiere el incremento del tamaño del índice de una forma directamente proporcional.

Repositorio (Tb)	Segmentos (8kb)	Hash (bytes)	Puntero (bytes)	Índice (Gb)
1	134.217.728	20	4	3
25	3.355.443.200	20	4	75
50	6.710.886.400	20	5	156
100	13.421.772.800	20	5	313

Tabla 3.3 - Tamaño de índice vs tamaño de repositorio (segmento 8kb).

Con objeto de conseguir el **rendimiento** requerido por un sistema de deduplicación de gama alta, es necesario que el índice se aloje completamente en memoria RAM del dispositivo para acelerar lo máximo posible las operaciones de búsqueda y actualización sobre él.

Cuando a partir de un determinado punto, el tamaño del índice supera el tamaño de la memoria RAM disponible del dispositivo, y es necesario alojarlo tanto en RAM como en disco, el rendimiento presenta el comportamiento que muestra la figura 3.5.

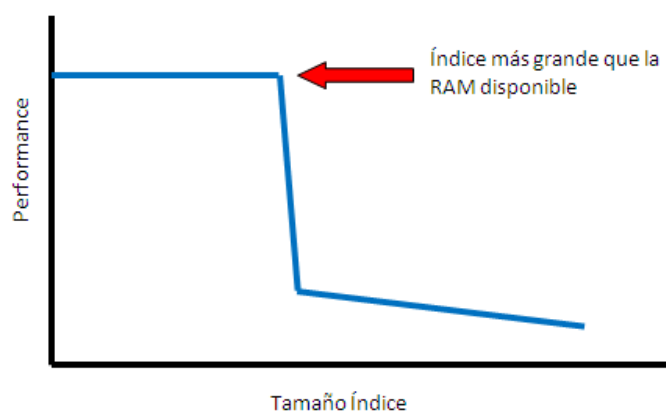


Fig. 3.5 - Impacto en la performance por incremento del tamaño de índice

Las técnicas de deduplicación basadas en hash, intentan resolver este problema de la brusca pérdida de rendimiento cuando el tamaño de índice supera una cierta cantidad razonable de memoria RAM y debe ser alojado parcialmente en disco. Este comportamiento suele marcar un tope en la capacidad del repositorio.

Para solucionar este efecto, es común aumentar el tamaño del segmento con objeto de mapear un número menor de segmentos en el repositorio. No obstante, esto provocará que el ratio de deduplicación disminuya. Es importante recordar que, cuanto mayor sea la granularidad en el análisis, mejores serán los ratios de deduplicación obtenidos.

En la tabla 3.4 se puede observar que para valores superiores de tamaño de segmento (16Kb), el efecto se sigue presentando aunque ligeramente desplazado. A partir de un determinado tamaño de índice, la memoria RAM del dispositivo no va a ser capaz de alojar por completo dicho índice y habrá que alojarlo parcialmente en RAM y disco.

Repositorio (Tb)	Segmentos (16kb)	Hash (bytes)	Puntero (bytes)	Índice (Gb)
1	67.108.864	20	4	2
25	1.677.721.600	20	4	38
50	3.355.443.200	20	4	75
100	6.710.886.400	20	5	156

Tabla 3.4 - Tamaño de índice vs tamaño de repositorio (segmento 16kb).

Como se pudo observar en el análisis realizado, el incremento del tamaño de repositorio plantea un posible problema de rendimiento por la disminución brusca de rendimiento a partir de un determinado punto. Por tanto, las técnicas basadas en hashing pueden no ser las adecuadas en escenarios donde sea requerido un compromiso agresivo de capacidad y rendimiento.

## B- Integridad

Estas técnicas también deben solucionar la problemática conocida de dichos algoritmos de hashing en los que es posible matemáticamente (con una probabilidad bastante baja) que dos segmentos tengan el mismo valor de hashing aunque sean diferentes a nivel binario. Este hecho puede provocar que se descarte un segmento como duplicado referenciándose a un segmento diferente. Esta situación provocará un error de consistencia en fase de recuperación puesto que el segmento que se recuperará cuando se acceda a él será diferente al debido.

Este problema es mucho más grave (de producirse) que una posible situación de error físico en el disco puesto que los sistemas actuales de almacenamiento implementan mecanismos de protección basadas en RAID que logran que los errores físicos de discos aislados no afecten a la recuperación del dato.

### 3.4.6.2.2 Técnicas basadas en diferencias binarias

Las técnicas basadas en algoritmos de diferencias o basadas en delta differencing, consisten en la división del conjunto de datos en segmentos (de longitud fija o variable) y en la búsqueda de posibles segmentos similares, efectuándose a continuación una comparación binaria.

El resultado del proceso de comparación binaria, es un conjunto de referencias a segmentos ya existentes en el repositorio, como la copia de los bloques nuevos o modificados que serán almacenados en el repositorio, siendo anotados en el índice para posteriores referencias.

A diferencia de las técnicas basadas en hashing, las técnicas basadas en delta no presentan el problema de integridad que describíamos anteriormente y suelen especificar 100% de integridad.

### **3.4.6.3 Con o Sin conocimiento de metadatos o estructura de datos**

Según el conocimiento o no, que se tenga de las estructuras de datos que se van a resguardar, las técnicas de deduplicación se dividen en *Técnicas Content-Aware* y *Técnicas Agnósticas*.

#### **3.4.6.3.1 Técnicas Content-Aware**

Cuando la técnica de deduplicación conoce la estructura de datos o tiene acceso a los metadatos que describen a los datos, ya resguardados en el repositorio de backup, se la denomina “Content-Aware”.

El objetivo principal de estas técnicas es, que al conocer la estructura de la información alojada será posible la identificación de objetos y sus relaciones dentro del espacio de almacenamiento a deduplicar y conseguir ratios mayores de deduplicación que las técnicas que trabajan a nivel de segmento (nivel de análisis inferior) y sin conocimiento de objetos.

Este planteamiento conlleva a vincular estrechamente la técnica con el sistema de backup en particular que utiliza el almacenamiento. Generalmente en los sistemas de backup más comerciales y extendidos, se utilizan técnicas Content-Aware.

Las técnicas Content-Aware se apoyan en el conocimiento de la estructura de datos que los sistemas de backup utilizan para almacenar las imágenes de backup. Son capaces de analizar e identificar dentro de la estructura de datos las imágenes de backup y dentro de

ellas los objetos que las componen (archivos y objetos de bases de datos principalmente). Este análisis a nivel de objetos suele ser el primer paso en su algoritmo.

Una vez identificados los objetos de las imágenes de backup, utilizan los metadatos alojados junto a ellos, para discernir por ejemplo que hay dos archivos con el mismo nombre y atributos contenidos en dos imágenes de backup y por lo tanto susceptibles de deduplicación.

Los metadatos son utilizados por éstas técnicas para relacionar objetos con instancias diferentes de dichos objetos ya almacenados en el repositorio y descartar referencias idénticas a los mismos. De este modo, construyen una lista de objetos candidatos potenciales a deduplicar, apoyándose en procedimientos de diferencias binarias para confirmar la duplicación. Este análisis a nivel de diferencias binarias suele ser el segundo paso en su algoritmo.

Los sistemas de deduplicación que utilizan técnicas Content-Aware deben conocer con un nivel de profundidad suficiente, qué tipos de datos están procesando para deduplicar especificando matrices de compatibilidad a nivel de:

- Versión y Sistema de backup de origen.
- Tipo y versión de los filesystems copiados.
- Tipo y versión de las bases de datos copiadas.

Como no siempre los sistemas de deduplicación Content-Aware se construyen con acceso a la propiedad intelectual de los sistemas de backup comerciales, muchos de estos sistemas efectúan procesos de “ingeniería inversa” para conocer la estructura y representación de metadatos y datos dentro de las imágenes de backup. Este hecho provoca que sean muy susceptibles a cambios de estructura de los sistemas de backup origen y que los niveles de deduplicación se vean afectados por posibles errores en la ingeniería inversa realizada para la interpretación de dicha estructura.

Por ese mismo motivo, estos sistemas suelen incorporar mecanismos robustos para garantizar la integridad de los datos, no efectuando la deduplicación de algún objeto sino pasa el segundo análisis a nivel de diferencias binarias.

### 3.4.6.3.2 Técnicas Agnósticas

Otras técnicas de deduplicación son capaces de trabajar sin conocimiento de metadatos o de la estructura de los datos almacenados. Estas técnicas se denominan Agnósticas (agnostic-stream) y permiten ser aplicadas independientemente del contenido, a una gran variedad de posibles datos de entrada.

Las técnicas Agnósticas, a diferencia de las técnicas Content-Aware, no especifican matrices de compatibilidad, siendo independientes de los sistemas de backup, de los filesystems y bases de datos copiadas, ya que no necesitan conocer la estructura de la información almacenada sino que analizan a nivel de segmento independientemente de su contenido.

### 3.4.6.4 Por Agente de deduplicación

Según el lugar donde se implemente el proceso de deduplicación, las diferentes técnicas pueden diferenciarse en deduplicación *En Origen* o *En Destino*.

#### 3.4.6.4.1 En Origen

Las técnicas de deduplicación en origen consisten, en que el proceso se ejecute antes del envío de la información hacia la red en el sistema propietario del almacenamiento primario.

En este tipo de técnicas de deduplicación en origen, el agente de backup realiza el proceso de deduplicación antes de enviar la información al servidor de backup. La información que se envía hacia la red con destino al servidor de backup es información deduplicada con el consiguiente ahorro en el volumen enviado.

La ventaja de estas técnicas es el ahorro de infraestructura de redes puesto que el ancho de banda consumido es reducido. A su vez los posibles inconvenientes que pueden surgir son:

- Impacto de procesamiento en el sistema origen.
- Ratios de deduplicación más pequeños, debido a que no se tienen en cuenta segmentos recibidos desde otros sistemas.



Estas técnicas suelen ser beneficiosas para el backup centralizado de la información alojada en oficinas remotas puesto que permiten optimizar el canal WAN de comunicaciones con el Centro de Datos principal corporativo.

#### **3.4.6.4.2 En Destino**

Las técnicas de deduplicación en destino, consisten en efectuar el proceso de deduplicación una vez recibido el conjunto de datos desde la red, siendo realizado por un sistema diferente al propietario (Generalmente el sistema de backup o incluso el dispositivo utilizado por éste).

A diferencia de las técnicas de deduplicación en origen, la información se recibe íntegra desde la red procediéndose a la deduplicación sobre dichos datos recibidos.

#### **3.4.6.5 Instante de Deduplicación**

En cuanto al instante en que se ejecuta el proceso de deduplicación, diferenciamos en *Deduplicación Inline* y *Deduplicación Diferida*.

##### **3.4.6.5.1 Deduplicación Inline**

Las técnicas de deduplicación inline se ejecutan mientras se reciben los datos desde la red y antes de su almacenamiento en el servidor correspondiente. En definitiva, se realiza durante la operación de backup.

La deduplicación inline ha de conseguir alcanzar unos niveles mínimos de rendimiento con objeto de no ver afectada la ventana de backup. En cuanto al rendimiento, suele ser aceptada generalmente una velocidad de 450-500 Mbps como velocidad mínima de entrada para ser calificable como Enterprise.

##### **3.4.6.5.2 Deduplicación Diferida**

Las técnicas de deduplicación diferida, consisten en aplicar los procesos de reducción fuera de la ventana de backup. En un primer momento, los datos se reciben de la red y son almacenados sin deduplicar en un buffer temporal de disco.

Una vez finalizada la ventana de backup, el sistema de deduplicación ejecuta el proceso de reducción del buffer temporal almacenando los segmentos únicos en el repositorio de deduplicación y actualizando su índice asociado.

Al no ejecutarse ningún procesamiento inline, esta técnica no impacta en la operación de backup. No obstante, requiere almacenamiento en disco adicional para poder ubicar el buffer temporal de recepción.

#### **3.4.6.6 Por Tipo de Solución: Appliance o Software**

Una última clasificación se refiere a si el proceso es ejecutado por un *Appliance* (cualquier dispositivo tanto hardware como software no considerable como sistema de propósito general y destinado a una tarea concreta) o por el contrario, está ejecutado por algún *Software* instalable sobre algún sistema de propósito general (con las restricciones correspondientes publicadas por el fabricante).

#### **3.4.7 IDEAS ERRÓNEAS SOBRE LA DEDUPLICACIÓN**

No se debe considerar que las bibliotecas de cintas virtuales con capacidad de deduplicación (VTL) sean una fuente infinita de dispositivos de cinta. Aunque los fabricantes le pueden permitir configurar 128 unidades de cinta lógicas o más, eso no se traduce automáticamente en un aumento masivo del rendimiento. Por ejemplo, encaminar las corrientes de datos a más de 100 unidades de cinta virtuales a través de un enlace de 1 Gbit/seg, el rendimiento no se incrementará debido al limitante del ancho de banda del enlace, el cual continúa siendo 1 Gbit/seg. Puede encontrarse con los mismos cuellos de botella de rentabilidad después de haber invertido decenas de miles de dólares.

Muchos proveedores explotarán el hecho de que los arrays de discos con función de deduplicación pueden ser más rápidas que la cinta, pero más allá de eso, siguen teniendo sus limitaciones. La deduplicación de datos a disco no es tecnología de espejo o de instantánea; los datos se tendrán que reensamblar y, si se administran a través de un producto de backup, también se tendrán que volver a escribir en un sistema de archivo, en un formato legible por las aplicaciones que acceden a los mismos. Dependiendo de la tecnología de deduplicación que se utilice, el rendimiento de las grandes operaciones de restauración también puede ser decepcionante.

## **3.5 MATRIZ DE PONDERACIÓN**

Para seleccionar la mejor opción que reemplace la infraestructura de backup que es cuestión de estudio de este trabajo, se utiliza como herramienta la **Matriz de Ponderación** permitiendo mayor claridad al momento de la toma de decisiones a nivel estratégico.

En los puntos siguientes se hace una pequeña referencia teórica sobre el concepto y características de las matrices de ponderación. Luego, se definen los criterios que son importantes para la empresa a la hora de la selección de una nueva infraestructura, posibilitando la creación de las matrices y la posterior toma de decisiones.

### **3.5.1 ¿QUÉ ES UNA MATRIZ DE PONDERACIÓN?**

La **Matriz de Ponderación o de Selección**, es una herramienta o instrumento que permite la selección de opciones sobre la base de la ponderación y aplicación de criterios. Se basa en la priorización de los criterios en función a la elaboración de matrices y al cálculo de algunos valores que en forma consecuente indican que criterios son las más importantes según su influencia hacia el resto y según su dependencia del resto.

Hace posible, determinar alternativas y los criterios a considerar para adoptar una decisión, priorizar y clarificar problemas, oportunidades de mejora y proyectos y, en general establecer prioridades entre un conjunto de elementos para facilitar la toma de decisiones.

### **3.5.2 ¿CÓMO SE ELABORA UNA MATRIZ DE PONDERACIÓN?**

Los pasos para la elaboración de una Matriz de Ponderación son:

#### **1- Definir el Objetivo:**

El planteamiento del objetivo debe ser claro y explícito.

#### **2- Definir los Criterios de Decisión:**

Si los criterios no están determinados, el equipo deberá elaborar una lista consensuada. Los criterios deben definirse nítidamente para que su significado no ofrezca duda a los miembros del equipo. Suele utilizarse la técnica de tormenta de ideas.

### **3- Ponderar los Criterios:**

Mediante una matriz de pares se ponderan los distintos criterios, confrontándolos con los demás. Para ello, y partiendo del eje vertical, se compara el primer criterio con los restantes, asignando el valor mas apropiado según la tabla de valores existente al efecto.

### **4- Comparar las Opciones:**

Se comparan todas las opciones entre si en función de cada uno de los criterios. Se crea para ello tantas matrices de pares como criterios se han definido, estableciendo las comparaciones de las opciones a analizar en cada uno de los criterios.

### **5- Seleccionar la mejor opción (Matriz Final):**

Se utiliza una matriz de pares en la que se compara cada opción sobre la base de la combinación de criterios. En ésta matriz resumen se sitúan los criterios en el eje vertical y las opciones en horizontal. Para cada celda de la matriz de priorización se multiplica el valor obtenido de “ponderación de criterio” (para cada criterio) por el valor de “calificación de la opción” (para cada opción). Luego sumar cada fila para obtener el puntaje final de cada una de las opciones. Finalmente seleccionar la opción de mayor puntaje.

## **3.6 ESTUDIO DE SELECCIÓN DE ALTERNATIVAS**

En este apartado se hace un estudio de diferentes alternativas permitiendo a través de la definición y ponderación de criterios y opciones llegar a la toma de decisiones para la selección del mejor hardware a utilizar en cada sitio como también la selección del mejor proveedor de dicho hardware.

Siguiendo los pasos indicados en el capítulo 3.5, se procede a la definición de criterios y elaboración de las diferentes matrices de selección según la necesidad del estudio que se esta desarrollando en este proyecto.

### **1- Objetivo**

El objetivo principal, es definir la mejor infraestructura de backup que se necesita para reemplazar la existente.

Para esto, se definen algunos objetivos secundarios, que ayudan a la selección de la mejor alternativa. Ellos son:

- a- Seleccionar el mejor hardware que existe en el mercado según los criterios definidos por la empresa.
- b- Seleccionar al mejor proveedor de acuerdo a los criterios definidos por la empresa.
- c- Seleccionar el mejor hardware según se adapte al tamaño del sitio.

## **2- Definición de Criterios**

Los criterios definidos por la empresa para la toma de decisiones son agrupados en diferentes categorías para una mejor comprensión. Estos son:

- Técnicos
  - Performance Bkp/Restore
  - Capacidad Nativa
  - Deduplicación
  - Vigencia tecnológica
  - Compatibilidad Software
  - Escalabilidad
- Ambientales
  - Tamaño HW
  - Consumo energético
- Operativos
  - Tipo de operación
  - Facilidad de operación
- Seguridad
  - Externalización
  - Retención de datos
  - Encriptación de datos
  - Fiabilidad

- Comerciales
  - Costo inversión
  - Garantía
  - Costo de mantenimiento
  - Plazo Entrega
  - Servicio técnico

### 3- Ponderación de los Criterios

Se crea una escala valorativa que permite hacer una comparación entre dos criterios, identificando cuál es la importancia de uno respecto al otro de acuerdo a las necesidades de la empresa.

En la tabla 3.5 se observan los valores definidos para comparar la **importancia** entre criterios.

Valor	Grado de Importancia
1	Cuando dos aspectos tienen igual importancia
2	Para el aspecto que tenga mayor importancia
3	Para el aspecto que tenga mucho mayor importancia
1/2 = 0,5	Para el aspecto que tenga menor importancia
1/3 = 0,33	Para el aspecto que tenga mucho menor importancia

Tabla 3.5 - Escala Valorativa de Criterios

Para calcular la importancia de cada criterio, se debe crear una matriz de pares, es decir una tabla con filas y columnas donde se nombren todos los criterios definidos y comparar la importancia de cada uno de ellos contra todos los otros según la escala valorativa de la tabla 3.5.

Una vez completada la matriz, sumar la fila de cada criterio. Luego sumar todos los valores de cada fila para obtener un total. Para calcular la ponderación de los criterios se divide la suma de cada fila sobre el total obtenido de la sumatoria de todos los valores de los criterios.

En la tabla 3.6 se muestra la matriz de ponderación de los criterios definidos con la ponderación de cada uno de ellos ya calculada. En ella podemos observar de acuerdo al análisis que el criterio de mayor importancia es el **Costo de Inversión** con un valor de ponderación de **0.093**, seguido por la **Performance de Backup/Restore** con **0.089**.

Criterios	Criterios																	Total	Ponderación del Criterio		
	Performance Bkpl/Rest	Capacidad Nativa	Deduplicación	Vigencia tecnológica	Compatibilidad SW	Escalabilidad	Tamaño HW	Consumo energético	Tipo de operación	Facilidad de operación	Externalización	Retención de datos	Encriptación de datos	Fiabilidad	Costo inversión	Garantía	Costo de mantenimiento			Plazo Entrega	Servicio técnico
Performance Bkpl/Rest	2	2	3	1	2	3	3	3	3	3	2	2	2	2	2	2	3	3	3	41,00	0,089
Capacidad Nativa	0,5	1	2	0,5	1	3	3	3	3	1	2	2	2	0,5	2	1	3	3	33,50	0,073	
Deduplicación	0,5	1	3	1	2	3	3	3	3	1	2	2	2	0,5	3	2	3	3	38,00	0,083	
Vigencia tecnológica	0,33	0,5	0,33	0,33	0,5	1	0,5	0,5	0,5	0,33	0,5	0,5	0,33	0,33	0,33	0,33	2	1	10,14	0,022	
Compatibilidad Software	1	2	1	3	0,5	0,5	0,33	1	1	0,33	2	2	1	1	2	1	3	3	25,66	0,056	
Escalabilidad	0,5	1	0,5	2	2	2	2	2	2	0,5	1	0,5	0,33	0,33	2	1	3	3	25,66	0,056	
Tamaño HW	0,33	0,33	0,33	1	2	0,5	1	0,5	0,5	0,33	0,33	0,5	0,33	0,33	0,5	0,33	2	0,5	11,64	0,025	
Consumo energético	0,33	0,33	0,33	2	3	0,5	1	0,5	0,5	0,33	0,33	0,5	0,33	0,33	0,5	0,33	2	0,5	13,64	0,030	
Tipo de operación	0,33	0,33	0,33	2	1	0,5	2	2	2	1	0,33	0,5	0,5	0,33	1	0,33	2	1	15,98	0,035	
Facilidad de operación	0,33	0,33	0,33	2	1	0,5	2	2	1	0,5	0,5	0,5	0,5	0,33	2	0,33	2	2	18,15	0,039	
Externalización	0,5	1	1	3	3	2	3	3	3	2	2	2	1	0,5	2	0,5	2	2	33,50	0,073	
Retención de datos	0,5	0,5	0,5	2	0,5	1	3	3	2	2	0,5	2	1	0,5	2	0,5	2	1	24,50	0,053	
Encriptación de datos	0,5	0,5	0,5	2	0,5	2	2	2	2	0,5	0,5	2	1	0,5	2	0,5	2	2	23,00	0,050	
Fiabilidad	0,5	0,5	0,5	3	1	3	3	3	2	2	1	1	1	0,5	2	0,5	2	2	28,50	0,062	
Costo inversión	1	2	2	3	1	3	3	3	3	2	2	2	2	3	2	3	3	3	43,00	0,093	
Garantía	0,5	0,5	0,33	3	0,5	0,5	2	2	1	0,5	0,5	0,5	0,5	0,33	1	3	2	19,16	0,042		
Costo de mantenimiento	0,5	1	0,5	3	1	1	3	3	3	2	2	2	2	0,5	1	3	2	33,50	0,073		
Plazo Entrega	0,33	0,33	0,33	0,5	0,33	0,33	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,33	0,33	0,33	0,5	7,64	0,017		
Servicio técnico	0,33	0,33	0,33	1	0,33	0,33	2	2	1	0,5	0,5	1	0,5	0,33	0,5	0,5	2	13,98	0,030		
<b>TOTAL:</b>																			<b>460,15</b>	<b>1</b>	

Tabla 3.6 - Matriz de Ponderación de Criterios.

#### 4- Comparación de Opciones

En este punto se creó una escala valorativa para la comparación de las opciones. La tabla 3.7 muestra esta nueva escala en base al **impacto** que una opción tiene respecto a las otras.

Valor	Grado de Impacto
1	Cuando dos aspectos tienen igual impacto
2	Para el aspecto que tenga mayor impacto
3	Para el aspecto que tenga mucho mayor impacto
1/2 = 0,5	Para el aspecto que tenga menor impacto
1/3 = 0,33	Para el aspecto que tenga mucho menor impacto

Tabla 3.7 - Escala Valorativa de Opciones.

Si bien los valores son iguales a la escala de la tabla 3.5, la diferencia está en el grado de impacto que una opción tiene sobre las otras y no en el grado de importancia.

Como se definió en los objetivos, para llegar al objetivo general existen 3 objetivos secundarios.

Para el primer objetivo propuesto (**a- Seleccionar el mejor hardware que existe en el mercado según los criterios definidos por la empresa**), las opciones a comparar son:

- Librerías Tradicionales de Cintas (Tape)
- Librerías Virtuales de Cintas (VTL)
- VTL + Tape
- Disk to Disk to Tape (VTL + VTL + Tape)

En la figura 3.6 se muestran todas las tablas de comparación para las opciones de hardware seleccionadas según el impacto de cada una de ellas en cada criterio. La ponderación de cada opción se calcula de la misma forma que la matriz descrita en el punto 3 “Ponderación de Criterios”.





En la tabla 3.8 se indica que la mejor opción a nivel de hardware es la utilización de librería virtual junto con una librería tradicional de cintas (VTL + Tape) con un promedio de **0.268**.

Se llegó a este resultado sumando los resultados de la multiplicación entre la PC (Ponderación de criterio) por la CO (Calificación de la Opción) de cada criterio.

Matriz resumen - Mejor Hw para backup	Performance Blip/Flex		Capacidad Nativa		Deduplicación		Vigencia tecnológica		Compatibilidad SW		Eresalabilidad		Tamaño Hw		Consumo energético		Tipo de operación		Facilidad de operación		Externalización		Retención de datos		Encriptación de datos		Fiabilidad		Costo inversión		Garantía		Costo de mantenimiento		Plazo Entrega		Servicio técnico		Puntaje Final
	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O			
Librería Física	0.089	0.087	0.073	0.103	0.083	0.062	0.022	0.080	0.056	0.250	0.056	0.457	0.025	0.371	0.030	0.371	0.035	0.345	0.039	0.296	0.073	0.313	0.053	0.286	0.050	0.087	0.062	0.118	0.093	0.420	0.042	0.250	0.073	0.420	0.017	0.250	0.030	0.250	0.247
VTL	0.089	0.328	0.073	0.207	0.083	0.313	0.022	0.170	0.056	0.250	0.056	0.228	0.025	0.371	0.030	0.371	0.035	0.276	0.039	0.296	0.073	0.062	0.053	0.107	0.050	0.228	0.062	0.130	0.093	0.330	0.042	0.250	0.073	0.330	0.017	0.250	0.030	0.250	0.249
VTL + Tape	0.089	0.328	0.073	0.345	0.083	0.313	0.022	0.270	0.056	0.250	0.056	0.228	0.025	0.186	0.030	0.186	0.035	0.276	0.039	0.296	0.073	0.313	0.053	0.357	0.050	0.228	0.062	0.294	0.093	0.180	0.042	0.250	0.073	0.180	0.017	0.250	0.030	0.250	0.268
D2D2T	0.089	0.261	0.073	0.345	0.083	0.313	0.022	0.480	0.056	0.250	0.056	0.087	0.025	0.072	0.030	0.072	0.035	0.103	0.039	0.111	0.073	0.313	0.053	0.250	0.050	0.457	0.062	0.457	0.093	0.070	0.042	0.250	0.073	0.070	0.017	0.250	0.030	0.250	0.236

Tabla 3.8 - Matriz Final para la selección del mejor Hardware.

Para desarrollar las matrices de ponderación de los otros dos objetivos secundarios, solo es necesario repetir los pasos 4 y 5 de la metodología. Esto permitirá continuar con el análisis de selección para el cumplimiento del objetivo principal.

A continuación se aborda el segundo objetivo (**b- Seleccionar al mejor proveedor de acuerdo a los criterios definidos por la empresa**).

Los valores de los criterios de ponderación (CP) a utilizar en este punto ya han sido calculados en el punto anterior. Estos van a ser:

- Costo de Inversión (**0.093**)
- Garantía (**0.042**)
- Costo de Mantenimiento (**0.073**)
- Plazo de entrega (**0.017**)
- Servicio Técnico (**0.030**)

Las opciones de proveedores a comparar contra los criterios definidos son:

- EMC
- HP
- IBM
- Oracle

Se utiliza el mismo cuadro valorativo definido con anterioridad para la comparación de las opciones.

En la figura 3.7 se muestran las tablas con las comparaciones (CO) de los proveedores en estudio según los criterios de ponderación que fueron determinados.

Costo Inversión	EMC	HP	IBM	Oracle	Suma	Calificación de la opción
EMC		0,5	3	2	5,5	0,330
HP	2		3	2	7	0,420
IBM	0,33	0,33		0,5	1,16	0,070
Oracle	0,5	0,5	2		3	0,180
TOTAL:					16,66	1

Garantía	EMC	HP	IBM	Oracle	Suma	Calificación de la opción
EMC		1	2	2	5	0,357
HP	1		2	2	5	0,357
IBM	0,5	0,5		1	2	0,143
Oracle	0,5	0,5	1		2	0,143
TOTAL:					14	1

Costo de Mantenimiento	EMC	HP	IBM	Oracle	Suma	Calificación de la opción
EMC		2	3	2	7	0,442
HP	0,5		2	2	4,5	0,284
IBM	0,33	0,5		2	2,83	0,179
Oracle	0,5	0,5	0,5		1,5	0,095
TOTAL:					15,83	1

Plazo Entrega	EMC	HP	IBM	Oracle	Suma	Calificación de la opción
EMC		1	2	1	4	0,296
HP	1		2	1	4	0,296
IBM	0,5	0,5		0,5	1,5	0,111
Oracle	1	1	2		4	0,296
TOTAL:					13,5	1

Servicio Técnico	EMC	HP	IBM	Oracle	Suma	Calificación de la opción
EMC		2	2	1	5	0,357
HP	0,5		1	0,5	2	0,143
IBM	0,5	1		0,5	2	0,143
Oracle	1	2	2		5	0,357
TOTAL:					14	1

Fig. 3.7 - Tablas de comparación entre opciones de proveedores según su impacto para cada criterio.

La matriz final para la selección de proveedores, indica que el mejor proveedor es la empresa **EMC** con un valor final de **0.094**, seguida por la empresa **Hewlett Packard** con **0.084**.

Los valores para cada una de las empresas están indicados en la tabla 3.9.

Matriz resumen - Mejor Proveedor	Costo de Inversión		Garantía		Costo de Mantenimiento		Plazo de Entrega		Servicio Técnico		Puntaje Final
	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	
EMC	0,093	0,330	0,042	0,357	0,073	0,442	0,017	0,296	0,030	0,357	0,094
	0,031		0,015		0,032		0,005		0,011		
HP	0,093	0,420	0,042	0,357	0,073	0,284	0,017	0,296	0,030	0,143	0,084
	0,039		0,015		0,021		0,005		0,004		
IBM	0,093	0,070	0,042	0,143	0,073	0,179	0,017	0,111	0,030	0,143	0,032
	0,006		0,006		0,013		0,002		0,004		
Oracle	0,093	0,180	0,042	0,143	0,073	0,095	0,017	0,296	0,030	0,357	0,045
	0,017		0,006		0,007		0,005		0,011		

Tabla 3.9 - Matriz Final para la selección del mejor Proveedor

Por último, se realizan algunas matrices de selección que ayudan a definir que hardware se adapta mejor al tamaño de cada uno de los sitios cuestión del estudio. Con estas tablas se cumple con el último objetivo secundario (**c- Seleccionar el mejor hardware según se adapte al tamaño del sitio**) planteado anteriormente.

Los valores de los criterios de ponderación (CP) a utilizar en este punto ya han sido calculados. Estos van a ser:

- Librerías Tradicionales de Cintas (Tape) **(0.247)**
- Librerías Virtuales de Cintas (VTL) **(0.249)**
- VTL + Tape **(0.268)**
- Disk to Disk to Tape (VTL + VTL + Tape) **(0.236)**

En las siguientes matrices se realiza una comparación entre las siguientes opciones, permitiendo descubrir cual se adapta mejor a los diferentes sitios. Ellas son:

- Pequeño
- Mediano
- Grande

Se utiliza el mismo cuadro valorativo definido con anterioridad para la comparación de las opciones.

En la figura 3.8 se muestran las tablas con las comparaciones (CO) de los diferentes tamaños de sitios en estudio según los criterios de ponderación que fueron determinados.

Librería	Pequeño	Mediano	Grande	Total	Calificación de la opción
Pequeño		1	2	3	0,429
Mediano	1		2	3	0,429
Grande	0,5	0,5		1	0,143
TOTAL:				7	1

VTL	Pequeño	Mediano	Grande	Total	Calificación de la opción
Pequeño		1	0,5	1,5	0,231
Mediano	1		1	2	0,308
Grande	2	1		3	0,462
TOTAL:				6,5	1

VTL + Tape	Pequeño	Mediano	Grande	Total	Calificación de la opción
Pequeño		0,5	0,5	1	0,143
Mediano	2		1	3	0,429
Grande	2	1		3	0,429
TOTAL:				7	1

D2D2T	Pequeño	Mediano	Grande	Total	Calificación de la opción
Pequeño		0,33	0,33	0,66	0,072
Mediano	3		0,5	3,5	0,382
Grande	3	2		5	0,546
TOTAL:				9,16	1

Fig. 3.8 - Tablas de comparación entre opciones de tamaños de sitios según su impacto para cada criterio.

En la matriz final quedan definidas las mejores opciones de hardware para cada uno de los tamaños de sitio. Por ejemplo en la tabla 3.10, se identifica que para un **sitio pequeño** la mejor opción es la utilización de **Librería de cintas tradicionales** con un valor de **0.106**.

Matriz resumen - Mejor Hw por sitio	Librería Física		VTL		VTL + Tape		D2D2T	
	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O
Pequeño	0,247	0,429	0,249	0,231	0,268	0,143	0,236	0,072
	0,106		0,057		0,038		0,017	
Mediano	0,247	0,429	0,249	0,308	0,268	0,429	0,236	0,382
	0,106		0,077		0,115		0,090	
Grande	0,247	0,143	0,249	0,462	0,268	0,429	0,236	0,546
	0,035		0,115		0,115		0,129	

Tabla 3.10 - Matriz Final para la selección del mejor Hardware para cada sitio

### **3.7 PROPUESTA DE SOLUCIONES**

En este apartado del trabajo se diseñan las nuevas infraestructuras propuestas como soluciones a las actuales, de acuerdo a los análisis realizados en el capítulo anterior a través de las matrices de selección.

#### **3.7.1 PROPUESTA SITIO GRANDE - CURITIBA**

Los resultados arrojados por las matrices de selección, para este tipo de sitios lo recomendado es la utilización de D2D2T. Para el caso de estudio de esta investigación, se utiliza como hardware seleccionado una Librería Virtual de Cintas (VTL) junto con una librería de cintas tradicional.

Esta elección se definió, debido a que en Curitiba se dispone de un solo centro de cómputos y no se justifica hacer una inversión tan grande para la utilización de D2D2T dentro del mismo edificio, ya que la mayor ventaja de esta infraestructura es la replicación de datos entre diferentes edificios.

El proveedor mejor posicionado según indica la matriz, es la empresa EMC que se especializa en el desarrollo de storages de discos de la cual se puede adquirir la VTL. Como EMC no comercializa librerías de cintas tradicionales, se tomará como proveedor de las mismas a la empresa Hewlett Packard.

La figura 3.9 define el diseño de la solución propuesta para el reemplazo de la infraestructura del sitio grande de Curitiba. Se observa que la solución utiliza solo un servidor de backup, se reemplaza la librería StorageTek por una HP MSL4048 y se agrega una Datadomain DD670 (VTL) de EMC.

De acuerdo a la volumetría de datos resguardados, el incremento vegetativo previsto para los siguientes años, la cantidad de servidores clientes actuales, entre otras características se dimensionó el hardware a comprar no limitando la solución, dejando la posibilidad de ampliación en caso de necesitarlo. Este análisis se puede observar en el anexo C apartado C.1.

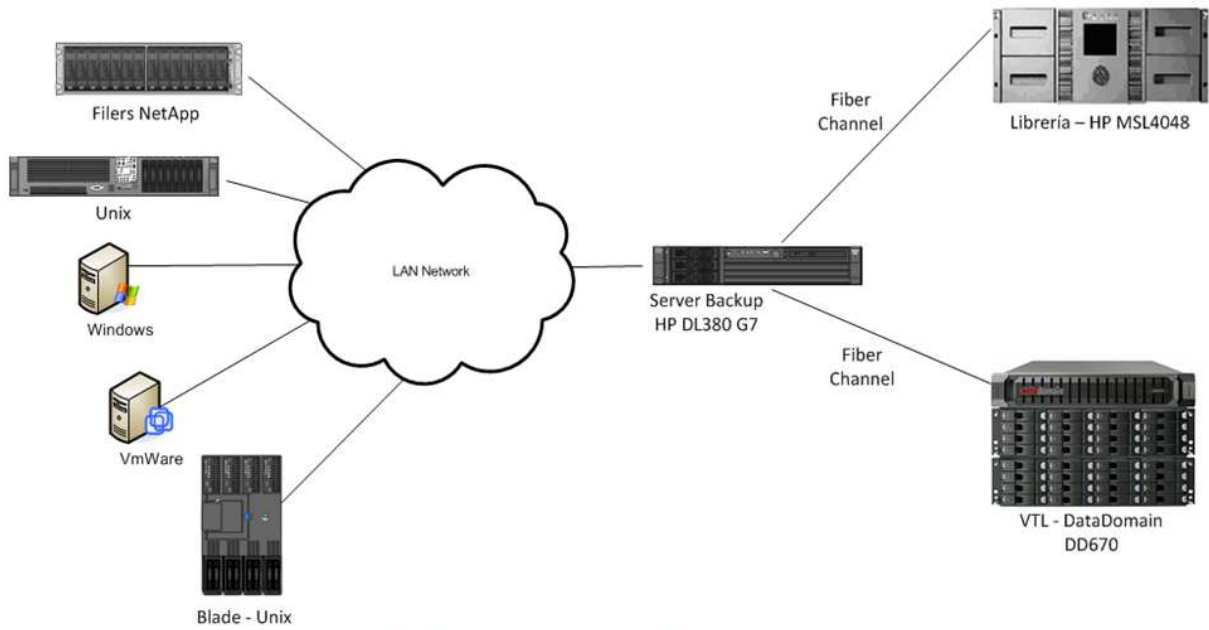


Fig. 3.9 - Diseño infraestructura propuesta para sitio de Curitiba - Sitio Grande

En las tablas 3.11, 3.12 y 3.13 se detallan las características del servidor de backup, de la VTL y de la librería de cintas propuestas para la nueva solución respectivamente.

Servidor Backup	
Marca	HP
Modelo	HP DL380 G7
Sist. Operativo	Linux RedHat Enterprise 5.6
Memoria	16 GB
Disco	2 x 300 GB
Conexión	4 Gigabit Eth + 4 FC
Unidad Rack	2U
Consumo eléctrico	460 watts/hr
Adm. Remota	ILO-3
Mantenimiento	Garantía de 3 años
Costo	8000 €

Tabla 3.11 - Descripción del nuevo servidor de Curitiba

Librería Virtual - VTL	
Marca	EMC
Modelo	DataDomain DD670
Sist. Operativo	DD OS versión 5.4
Disco	42 TB
Tipo RAID	RAID 6
Conexión	8 FC
Deduplicación	Si
Unidad Rack	6U
Consumo eléctrico	688 watts/hr
Adm. Remota	ILO-3
Mantenimiento	Garantía de 3 años
Costo	116600 €

Tabla 3.12 - Descripción de la VTL de Curitiba

Librería de Backup	
Marca	HP
Modelo	MSL4048
# Drives	4 (LTO-5 FC)
# Slots	48
# Mailslots	3
Capacidad Nativa / Comprimida	72 TB / 144 TB
Interfaz conexión	FC
Panel Operacional	SI
Adm. Remota	ILO-3
Rackeable	SI
Unidad Rack	4U
Consumo eléctrico	312 Watts/hr
Mantenimiento	Garantía 1 año
Costo	25000 €

Tabla 3.13 - Descripción de la nueva librería de Curitiba

### **3.7.2 PROPUESTA SITIO MEDIANO - CÓRDOBA**

Para el tamaño mediano de sitio como es el caso de Córdoba que cuenta con un solo centro de cómputos, se decidió invertir directamente en la compra de una nueva librería tradicional de cintas, ya que la externalización de cintas outside es una política interna muy importante que tiene la empresa.



Si bien en la matriz de selección, la mejor opción es la de utilizar una VTL junto a una librería de cintas, la empresa informó que está proyectando retirar del resguardo los dos filers en el corto plazo, lo que provoca que el volumen de datos disminuya en un 50% aproximadamente.

El proveedor seleccionado para la adquisición de la librería de cintas es la empresa Hewlett Packard.

En la figura 3.10 se muestra el diseño de la solución propuesta para el reemplazo de la infraestructura del sitio mediano de Córdoba.

De acuerdo a la volumetría de datos resguardados, el incremento vegetativo previsto para los siguientes años, la cantidad de servidores clientes actuales, entre otras características se dimensionó el hardware a comprar no limitando la solución, dejando la posibilidad de ampliación en caso de necesitarlo. Este análisis se puede observar en el anexo C apartado C.2.

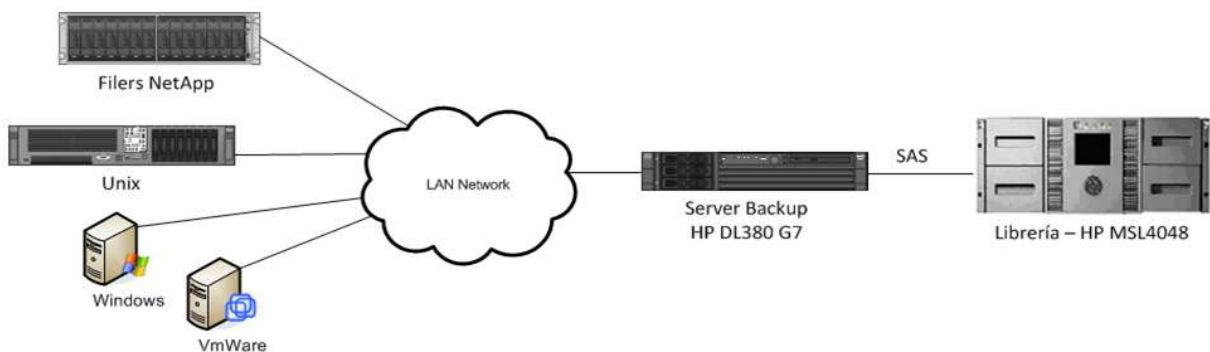


Fig. 3.10 - Diseño infraestructura propuesta para sitio de Córdoba - Sitio Mediano

En las tablas 3.14 y 3.15 se detallan las características del servidor de backup y de la librería de cintas propuestas para la nueva solución respectivamente.

Servidor Backup	
Marca	HP
Modelo	HP DL380 G7
Sist. Operativo	Linux RedHat Enterprise 5.6
Memoria	16 GB
Disco	2 x 300 GB
Conexión	4 Gigabit Eth + 2 SAS
Unidad Rack	2U
Consumo eléctrico	460 watts/hr
Adm. Remota	ILO-3
Mantenimiento	Garantía de 3 años
Costo	10000 €

Tabla 3.14 - Descripción del nuevo servidor de Córdoba

Librería de Backup	
Marca	HP
Modelo	MSL4048
# Drives	2 (LTO-4 SAS)
# Slots	48
# Mailslots	3
Capacidad Nativa / Comprimida	38.4 TB / 76.8 TB
Interfaz conexión	SAS
Panel Operacional	SI
Adm. Remota	ILO-3
Rackeable	SI
Unidad Rack	4U
Consumo eléctrico	312 Watts/hr
Mantenimiento	Garantía 1 año
Costo	19000 €

Tabla 3.15 - Descripción de la nueva librería de Córdoba

### 3.7.2 PROPUESTA SITIO PEQUEÑO - BUENOS AIRES

Para los sitios pequeños como Buenos Aires, la matriz de selección de hardware da como mejor opción la utilización de una librería tradicional de cintas. Nuevamente para éste caso la empresa proveedora seleccionada es Hewlett Packard.

En la figura 3.11 se especifica el nuevo diseño de la infraestructura de backup.

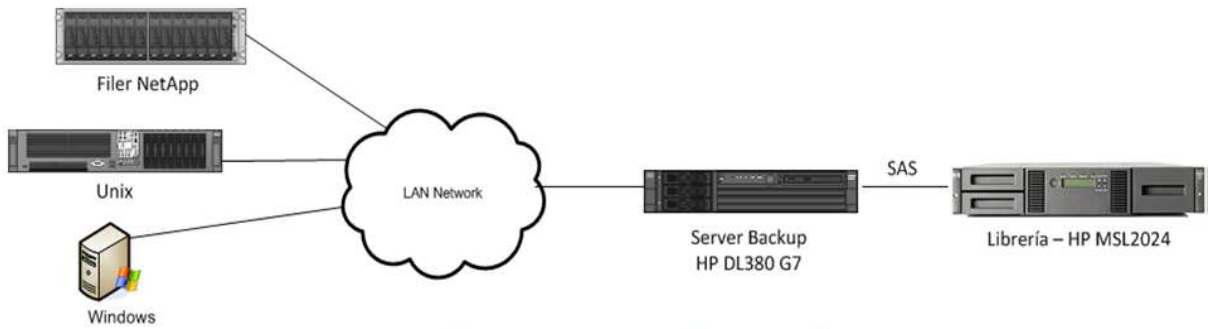


Fig. 3.11 - Diseño infraestructura propuesta para sitio de Buenos Aires - Sitio Pequeño

La descripción del hardware a reemplazar en la solución actual se detalla en las tablas 3.16 y 3.17. En el análisis de dimensionamiento de hardware se detalla en el anexo C apartado C.3.

Servidor Backup	
Marca	HP
Modelo	HP DL380 G7
Sist. Operativo	Linux RedHat Enterprise 5.6
Memoria	16 GB
Disco	2 x 300 GB
Conexión	4 Gigabit Eth + 2 SAS
Unidad Rack	2U
Consumo eléctrico	460 watts/hr
Adm. Remota	ILO-3
Mantenimiento	Garantía de 3 años
Costo	10000 €

Tabla 3.16 - Descripción del nuevo servidor de Buenos Aires

Librería de Backup	
Marca	HP
Modelo	MSL2024
# Drives	1 (LTO-4 SAS)
# Slots	24
# Mailslots	1
Capacidad Nativa / Comprimida	19.2TB / 38.4 TB
Interfaz conexión	SAS
Panel Operacional	SI
Adm. Remota	ILO-3
Rackeable	SI
Unidad Rack	2U
Consumo eléctrico	168 Watts/hr
Mantenimiento	Garantía 1 año
Costo	11000 €

Tabla 3.17 - Descripción de la nueva librería de Buenos Aires

### **3.8 COMPARACIÓN SOLUCIÓN ACTUAL VS. PROPUESTA**

Para determinar si la solución propuesta mejora la solución en uso actualmente, se hace una comparación tomando los 10 criterios más importantes según se estableció en la matriz de la tabla 3.5. Ellos son:

Performance Bkp/Restore	(0,089)
Deduplicación	(0,083)
Capacidad Nativa	(0,073)
Externalización	(0,073)
Costo de mantenimiento	(0,073)
Fiabilidad	(0,062)
Escalabilidad	(0,056)
Retención de datos	(0,053)
Consumo energético	(0,030)
Tamaño HW	(0,025)

Se puede observar que fueron descartados dos puntos importantes como lo es el costo de inversión (no hay costo de inversión en la infraestructura actual), y la compatibilidad con el software (el nuevo hardware debe ser compatible si o si por políticas de casa central).

El análisis comparativo utilizado es el desarrollado en el punto 3.5, donde se estudia la importancia de una solución respecto a la otra y según el resultado que da, se multiplica por el valor del criterio de estudio ya calculado. Una vez realizado este procedimiento para cada criterio se suman los valores y la alternativa de infraestructura con mayor puntaje es la mejor solución a utilizar.

En los puntos siguientes se realiza para cada sitio el estudio de las infraestructuras.

### 3.8.1 COMPARACIÓN SITIO GRANDE - CURITIBA

En la figura 3.12 se muestran las tablas con las comparaciones (CO) entre la infraestructura actual versus la solución propuesta según los criterios de ponderación que fueron determinados.

Performance Bkp/Restore	Actual	Propuesta	Total	Calificación de la opción
Actual		0.5	0.50	0.200
Propuesta	2		2.00	0.800
TOTAL:			2.5	1

Capacidad Nativa	Actual	Propuesta	Total	Calificación de la opción
Actual		0.5	0.50	0.200
Propuesta	2		2.00	0.800
TOTAL:			2.5	1

Deduplicación	Actual	Propuesta	Total	Calificación de la opción
Actual		0.33	0.33	0.099
Propuesta	3		3.00	0.901
TOTAL:			3.33	1

Externalización	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1.00	0.500
Propuesta	1		1.00	0.500
TOTAL:			2	1

Escalabilidad	Actual	Propuesta	Total	Calificación de la opción
Actual		0.33	0.33	0.099
Propuesta	3		3.00	0.901
TOTAL:			3.33	1

Tamaño HW	Actual	Propuesta	Total	Calificación de la opción
Actual		0.5	0.50	0.200
Propuesta	2		2.00	0.800
TOTAL:			2.5	1

Costo de Mantenimiento	Actual	Propuesta	Total	Calificación de la opción
Actual		0.33	0.33	0.099
Propuesta	3		3.00	0.901
TOTAL:			3.33	1

Fiabilidad	Actual	Propuesta	Total	Calificación de la opción
Actual		0.33	0.33	0.099
Propuesta	3		3.00	0.901
TOTAL:			3.33	1

Consumo Energético	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1.00	0.500
Propuesta	1		1.00	0.500
TOTAL:			2	1

Retención de Datos	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1.00	0.500
Propuesta	1		1.00	0.500
TOTAL:			2	1

Fig. 3.12 - Tablas de comparación entre infraestructura Actual vs. Propuesta - Sitio Grande.

A simple vista se observa en la tabla 3.18 una amplia superioridad de la nueva solución. Si bien es un resultado esperado debido a que el hardware actual es muy viejo y en el transcurso de los años la tecnología ha avanzado mucho, es válido reflejar los resultados ya que hacen al estudio en cuestión.

Matriz resumen - Mejor Infraestructura Sitio Grande	Performance Bkp/Res		Deduplicación		Capacidad Nativa		Externalización		Costo Mantenimiento		Fiabilidad		Escalabilidad		Retención de Datos		Consumo Energético		Tamaño de HW		Puntaje Final
	F.C	C.O	F.C	C.O	F.C	C.O	F.C	C.O	F.C	C.O	F.C	C.O	F.C	C.O	F.C	C.O	F.C	C.O	F.C	C.O	
	Actual	0,089	0,200	0,083	0,099	0,073	0,200	0,073	0,500	0,073	0,099	0,062	0,099	0,056	0,099	0,053	0,500	0,030	0,500	0,025	
Propuesta	0,089	0,800	0,083	0,901	0,073	0,800	0,073	0,500	0,073	0,901	0,062	0,901	0,056	0,901	0,053	0,500	0,030	0,500	0,025	0,800	0,474
	0,018		0,008		0,015		0,037		0,007		0,006		0,006		0,027		0,015		0,005		

Tabla 3.18 - Matriz Final para la selección de la mejor infraestructura - Sitio Grande.

En la tabla 3.19 se resumen los problemas relevados en el análisis del trabajo con las soluciones y mejoras que tienen con la nueva infraestructura.

Problemas		Solución - Mejora
<b>Complejidad</b>	<p>a- Utiliza dos servidores de backup, conectados a una librería (vía SCSI y FC) para atender la cantidad actual de servidores clientes.</p> <p>b- Utilización de sistemas operativos diferentes en los servidores de backup.</p> <p>c- La cantidad de horas hombre utilizadas para la administración de la infraestructura es muy alta.</p>	<p>a. Se pasa a utilizar un solo servidor nuevo, conectado por fibra tanto a la VTL como a la librería.</p> <p>b. Al ser un solo servidor se utilizaría como único sistema operativo Linux RedHat Enterprise 5.6.</p> <p>c. Se disminuye la cantidad de horas hombre debido a la simplificación de la infraestructura.</p>
<b>Costos</b>	<p>a- Licencia para los dos servidores de backup adquiridas a Atempo.</p> <p>b- Licencia para la utilización de 12 drives.</p> <p>c- Mantenimiento de los dos servidores de backup.</p> <p>d- Necesidad de adquirir más cintas, ya que se utiliza gran cantidad por el volumen actual de datos.</p>	<p>a. Se necesita solo una licencia para el servidor de backup.</p> <p>b. Se reduce el costo de licencia a 4 drives.</p> <p>c. Ahorro en mantenimiento de servidores de backup, ya que se tiene 4 años de garantía.</p> <p>d. La cantidad se reduce ampliamente ya que la capacidad de almacenamiento de las cintas son superiores.</p>
<b>Performance</b>	<p>a- La información más crítica de la empresa tiene un tiempo de backup de 26 horas y el tiempo de restore es de 5 días. Tiempos inaceptables para el negocio.</p> <p>b- El backup es realizado por la misma LAN de producción a 1000 MB/s. Esto afecta visiblemente el ancho de banda en los horarios productivos.</p>	<p>a. Disminución de la ventana de backup y restore de la información a menos de la mitad del tiempo actual.</p> <p>b. Se propone la configuración de una VLAN exclusiva para la realización de los backups.</p>
<b>Hardware</b>	<p>a- Importante cantidad de incidentes mensuales producidos por el hardware de la librería.</p> <p>b- Actualmente existen dos drives offline por problemas.</p> <p>c- La falta de mantenimiento y limpieza de los drives provocan que las cintas queden presas y se requiera intervención manual para liberarlos.</p>	<p>a. Reducción de los incidentes por falla del hardware.</p> <p>b y c. La librería es nueva.</p>
<b>Seguridad</b>	<p>a- Las cintas son muy viejas y no confiables. Propensas a dañarse, provocando la pérdida de los datos.</p> <p>b- Alto porcentajes de los Jobs de backup cancelados.</p> <p>c- Imposibilidad de cumplir con los 21 días de retención de la información según las políticas de la compañía.</p>	<p>a. Parque de cintas LTO-5 todas nuevas.</p> <p>b. Porcentaje casi nulo de errores de backups.</p> <p>c. Capacidad de retención de los datos por 21 días en VTL como en la librería de cintas.</p>
<b>Mantenimiento</b>	<p>a- Librería L700 sin mantenimiento.</p> <p>b- Servidor de backup número 1 llegó a fin de vida y esta fuera de soporte y mantenimiento.</p>	<p>a. La librería y la VTL tienen 4 años de garantía</p> <p>b. Servidor se da de baja.</p>
<b>Escalabilidad</b>	<p>a- Solución trabajando al 100 %.</p>	<p>a. Posibilidad de ampliación en VTL y librería.</p>

Tabla 3.19 - Resumen de los problemas con sus soluciones y mejoras en sitio Grande.

### 3.8.2 COMPARACIÓN SITIO MEDIANO - CÓRDOBA

Al igual que en Brasil, Córdoba tiene una infraestructura muy vieja, la cuál es ampliamente superada por la propuesta como mejora.

En la figura 3.13 se observa el impacto de cada infraestructura sobre la otra según los criterios de ponderación de estudio.

Performance BkpfRestore	Actual	Propuesta	Total	Calificación de la opción
Actual		0,5	0,50	0,200
Propuesta	2		2,00	0,800
TOTAL: 2,5 1				

Capacidad Nativa	Actual	Propuesta	Total	Calificación de la opción
Actual		0,5	0,50	0,200
Propuesta	2		2,00	0,800
TOTAL: 2,5 1				

Deduplicación	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1,00	0,500
Propuesta	1		1,00	0,500
TOTAL: 2 1				

Externalización	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1,00	0,500
Propuesta	1		1,00	0,500
TOTAL: 2 1				

Escalabilidad	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1,00	0,500
Propuesta	1		1,00	0,500
TOTAL: 2 1				

Tamaño HW	Actual	Propuesta	Total	Calificación de la opción
Actual		0,5	0,50	0,200
Propuesta	2		2,00	0,800
TOTAL: 2,5 1				

Costo de Mantenimiento	Actual	Propuesta	Total	Calificación de la opción
Actual		0,33	0,33	0,099
Propuesta	3		3,00	0,901
TOTAL: 3,33 1				

Fiabilidad	Actual	Propuesta	Total	Calificación de la opción
Actual		0,33	0,33	0,099
Propuesta	3		3,00	0,901
TOTAL: 3,33 1				

Consumo Energético	Actual	Propuesta	Total	Calificación de la opción
Actual		2	2,00	0,800
Propuesta	0,5		0,50	0,200
TOTAL: 2,5 1				

Retención de Datos	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1,00	0,500
Propuesta	1		1,00	0,500
TOTAL: 2 1				

Fig. 3.13 - Tablas de comparación entre infraestructura Actual vs. Propuesta - Sitio Mediano.

La gran diferencia se ve reflejada en la tabla 3.20.

Matriz resumen - Mejor Infraestructura Site Mediano	Performance Bkpf/Res		Deduplicacion		Capacidad Nativa		Externalizacion		Costo Mantenimiento		Fiabilidad		Escalabilidad		Retencion de Datos		Consumo Energetico		Tamaño de HW		Puntaje Final
	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	
	Actual	0,089	0,200	0,083	0,500	0,073	0,200	0,073	0,500	0,073	0,099	0,062	0,099	0,056	0,500	0,053	0,500	0,030	0,200	0,025	
Propuesta	0,089	0,800	0,083	0,500	0,073	0,800	0,073	0,500	0,073	0,901	0,062	0,901	0,056	0,500	0,053	0,500	0,030	0,800	0,025	0,800	0,428
	0,018		0,042		0,015		0,037		0,007		0,006		0,028		0,027		0,006		0,005		
	0,071		0,042		0,058		0,037		0,066		0,056		0,028		0,027		0,024		0,020		

Tabla 3.20 - Matriz Final para la selección de la mejor infraestructura - Sitio Mediano.



En la tabla 3.21 se resumen los problemas relevados en el análisis del trabajo con las soluciones y mejoras que tienen con la nueva infraestructura.

Problemas		Solución - Mejora
<b>Costos</b>	<ul style="list-style-type: none"> <li>a- Mantenimiento del servidor de backup.</li> <li>b- Licencia de Sistema Operativo (HP-UX) es muy elevada.</li> <li>c- Necesidad de adquirir más cintas, ya que se utiliza gran cantidad por el volumen actual de datos.</li> </ul>	<ul style="list-style-type: none"> <li>a. Nuevo servidor de backup con garantía de 4 años.</li> <li>b. Se utiliza como sistema operativo Linux RedHat Enterprise 5.6.</li> <li>c. La cantidad se reduce ampliamente ya que la capacidad de almacenamiento de las cintas son superiores.</li> </ul>
<b>Performance</b>	<ul style="list-style-type: none"> <li>a- El backup es realizado por la misma LAN de producción a 100 MB/s. Esto afecta visiblemente el ancho de banda en los horarios productivos.</li> </ul>	<ul style="list-style-type: none"> <li>a. Se propone la configuración de una VLAN exclusiva para la realización de los backups.</li> </ul>
<b>Hardware</b>	<ul style="list-style-type: none"> <li>a- Importante cantidad de incidentes mensuales producidos por el hardware de la librería.</li> <li>b- Actualmente existe un drive offline por problemas.</li> <li>c- La falta de mantenimiento y limpieza de los drives provocan que las cintas queden presas y se requiera intervención manual para liberarlos.</li> </ul>	<ul style="list-style-type: none"> <li>a. Reducción de los incidentes por falla del hardware.</li> <li>b y c. La librería es nueva.</li> </ul>
<b>Seguridad</b>	<ul style="list-style-type: none"> <li>a- Las cintas son muy viejas y no confiables. Propensas a dañarse, provocando la pérdida de los datos.</li> <li>b- Alto porcentajes de los Jobs de backup cancelados.</li> <li>c- Imposibilidad de cumplir con los 21 días de retención de la información según las políticas de la compañía.</li> </ul>	<ul style="list-style-type: none"> <li>a. Parque de cintas LTO-4 todas nuevas.</li> <li>b. Disminución de errores de backup.</li> <li>c. Capacidad de retención de los datos por 21 días en la librería de cintas.</li> </ul>
<b>Mantenimiento</b>	<ul style="list-style-type: none"> <li>a- Librería L180 sin mantenimiento.</li> <li>b- Servidor de backup llegó a su fin de vida y está fuera de soporte y mantenimiento.</li> </ul>	<ul style="list-style-type: none"> <li>a. La librería tienen 4 años de garantía.</li> <li>b. Servidor se reemplaza.</li> </ul>
<b>Escalabilidad</b>	<ul style="list-style-type: none"> <li>a- Solución trabajando al 100 %.</li> </ul>	<ul style="list-style-type: none"> <li>a. Posibilidad de ampliación en la nueva librería.</li> </ul>

Tabla 3.21 - Resumen de los problemas con sus soluciones y mejoras en Sitio Mediano.

### 3.8.3 COMPARACIÓN SITIO PEQUEÑO – BUENOS AIRES

El caso de Buenos Aires no difiere de los anteriores, por lo que los resultados obtenidos en la comparación del sitio pequeño nos dan como mejor solución la propuesta en este proyecto.

En la figura 3.14 se detalla la comparación de los criterios de ponderación, y en la tabla 3.22 los resultados finales.

Performance Bkp/Restore				
	Actual	Propuesta	Total	Calificación de la opción
Actual		0,5	0,50	0,200
Propuesta	2		2,00	0,800
TOTAL:			2,5	1

Capacidad Nativa				
	Actual	Propuesta	Total	Calificación de la opción
Actual		0,5	0,50	0,200
Propuesta	2		2,00	0,800
TOTAL:			2,5	1

Deduplicación				
	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1,00	0,500
Propuesta	1		1,00	0,500
TOTAL:			2	1

Externalización				
	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1,00	0,500
Propuesta	1		1,00	0,500
TOTAL:			2	1

Escalabilidad				
	Actual	Propuesta	Total	Calificación de la opción
Actual		0,33	0,33	0,099
Propuesta	3		3,00	0,901
TOTAL:			3,33	1

Tamaño HW				
	Actual	Propuesta	Total	Calificación de la opción
Actual		0,5	0,50	0,200
Propuesta	2		2,00	0,800
TOTAL:			2,5	1

Costo de Mantenimiento				
	Actual	Propuesta	Total	Calificación de la opción
Actual		0,33	0,33	0,099
Propuesta	3		3,00	0,901
TOTAL:			3,33	1

Fiabilidad				
	Actual	Propuesta	Total	Calificación de la opción
Actual		0,33	0,33	0,099
Propuesta	3		3,00	0,901
TOTAL:			3,33	1

Consumo Energético				
	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1,00	0,500
Propuesta	1		1,00	0,500
TOTAL:			2	1

Retención de Datos				
	Actual	Propuesta	Total	Calificación de la opción
Actual		1	1,00	0,500
Propuesta	1		1,00	0,500
TOTAL:			2	1

Fig. 3.14 - Tablas de comparación entre infraestructura Actual vs. Propuesta - Sitio Pequeño.

Matriz Resumen - Mejor Infraestructura Site Pequeño	Performance Bkp/Res		Deduplicación		Capacidad Nativa		Externalización		Costo Mantenimiento		Fiabilidad		Escalabilidad		Retención de Datos		Consumo Energético		Tamaño de HW		Puntaje Final
	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	P.C	C.O	
	Actual	0,089	0,200	0,083	0,500	0,073	0,200	0,073	0,500	0,073	0,099	0,062	0,099	0,056	0,099	0,053	0,500	0,030	0,500	0,025	
Propuesta	0,089	0,800	0,083	0,500	0,073	0,800	0,073	0,500	0,073	0,901	0,062	0,901	0,056	0,901	0,053	0,500	0,030	0,500	0,025	0,800	0,441

Tabla 3.22 - Matriz Final para la selección de la mejor infraestructura - Sitio Pequeño.

En la tabla 3.23 se resumen los problemas relevados en el análisis del trabajo con las soluciones y mejoras que tienen con la nueva infraestructura.

Problemas		Solución - Mejora
<b>Costos</b>	<p>a- Mantenimiento del servidor de backup.</p> <p>b- Necesidad de adquirir más cintas debido al incremento de los datos.</p>	<p>a. Nuevo servidor de backup con garantía de 4 años.</p> <p>b. La cantidad se reduce ampliamente ya que la capacidad de almacenamiento de las cintas son superiores.</p>
<b>Hardware</b>	<p>a- Importante cantidad de incidentes mensuales producidos por el hardware de la librería.</p> <p>b- La falta de mantenimiento y limpieza de los drives provocan que las cintas queden presas y se requiera intervención manual para liberarlos.</p>	<p>a. Reducción de los incidentes por falla del hardware.</p> <p>b. La librería es nueva.</p>
<b>Seguridad</b>	<p>a- Las cintas son muy viejas y no confiables. Propensas a dañarse, provocando la pérdida de los datos.</p> <p>b- Alto porcentajes de los Jobs de backup cancelados.</p> <p>c- Imposibilidad de cumplir con los 21 días de retención de la información según las políticas de la compañía.</p>	<p>a. Parque de cintas LTO-4 todas nuevas.</p> <p>b. Disminución de errores de backup.</p> <p>c. Capacidad de retención de los datos por 21 días en la librería de cintas.</p>
<b>Mantenimiento</b>	<p>a- Librería L140 sin mantenimiento.</p> <p>b- Servidor de backup llego a su fin de vida y esta fuera de soporte y mantenimiento.</p>	<p>a. La librería tienen 4 años de garantía.</p> <p>b. Servidor se reemplaza.</p>

Tabla. 3.23 - Resumen de los problemas con sus soluciones y mejoras en Sitio Pequeño.

## *Capítulo IV – Conclusión*

---

## **CAPITULO IV - CONCLUSIÓN**

En este proyecto final de grado se tomó como referencia una empresa que tiene una serie de problemas a resolver en la infraestructura de backup utilizada. Todos estos problemas fueron relevados para conocer el estado de la situación actual en que se encuentra la infraestructura. Se investigó diferentes fabricantes para conocer los productos ofrecidos y a través de métodos de selección de alternativas se propuso una nueva solución para cada tamaño de sitio. Finalmente se hizo una comparación entre la solución actual y la propuesta determinando cual es la mejor.

Este trabajo ha permitido adquirir conocimientos sobre las ofertas y avances del hardware de distintos proveedores que se utilizan en la actualidad para el resguardo seguro de la información.

Se conoció y desarrolló el concepto de Deduplicación de datos, que es sin dudas un avance extraordinario para paliar el incremento constante de información que afectan a las organizaciones, permitiendo grandes ratios de reducción de datos que son almacenados.

Se diseñó una metodología de selección basada en una matriz de ponderación, que permitió descubrir la mejor opción según las necesidades de la empresa, a nivel hardware como a nivel de proveedores. Es importante mencionar que esta herramienta es utilizable para empresas y organizaciones de cualquier índole.

Se propusieron nuevas soluciones de infraestructura de backup para cada tamaño de sitio según los resultados obtenidos en el procedimiento de selección.

Por último, se realizó un análisis comparativo entre las diferentes infraestructuras de backup (la actual versus la propuesta) de cada tamaño de sitio, y se comprobó que la solución propuesta para cada sitio es mejor que la actual, ya que permite eliminar todas las situaciones problemáticas que fueron detectadas.

Un aspecto muy importante a señalar, es que el procedimiento de selección desarrollado puede aplicarse en un futuro para analizar diferentes alternativas de software de backup.

## *Bibliografía*

---

## **BIBLIOGRAFÍA**

[1] Definición de backup

[http://es.wikipedia.org/wiki/Copia\\_de\\_seguridad](http://es.wikipedia.org/wiki/Copia_de_seguridad)

[2] Definición y características de RAID:

[http://es.wikipedia.org/wiki/RAID#RAID\\_0\\_.28Data\\_Striping.29](http://es.wikipedia.org/wiki/RAID#RAID_0_.28Data_Striping.29)

[3] Interfaz de conexión Discos Rígidos

[http://es.wikipedia.org/wiki/Disco\\_duro#Tipos\\_de\\_conexi.C3.B3n](http://es.wikipedia.org/wiki/Disco_duro#Tipos_de_conexi.C3.B3n)

[4] Capacidad de los Discos Rígidos

<http://eldestornillador.wordpress.com/2009/09/23/la-capacidad-de-los-discos-duros/>

[5] Deduplicación:

<http://www.searchstorage.es/respaldo-de-datos/tecnologia-de-deduplicacion-de-datos-casos-que-justifican-la-deduplicacion/>

[6] Técnicas de Deduplicación de Datos y Aplicación en librerías virtuales de cintas

PFC\_FRANCISCO\_JAVIER\_JIMENEZ\_PATRICIO.pdf

[7] Diferencia entre deduplicación y otras tecnologías

<http://www.lortu.es/Technology.aspx>

[8] Calculadora de Deduplicación

<http://www.openstore.com/products/dedpup-calaculator.php>

[9] Site HP

[www.hp.com](http://www.hp.com)

[10] Site IBM

[www.ibm.com](http://www.ibm.com)

[11] Site EMC

[www.emc.com](http://www.emc.com)

[12] Site Oracle

[www.oracle.com](http://www.oracle.com)

[13] Site Wikipedia

<http://es.wikipedia.org/wiki/Wikipedia:Portada>

[14] Matriz de selección

<http://es.scribd.com/doc/72251692/6taherramientadecalidadmatrizdeseleccion>



# *Anexos*

---

## ANEXOS

### ANEXO A – LIBRERÍAS DE CINTAS

En este apartado, se muestran las características técnicas principales de las librerías de cintas según la información provista por los fabricantes en sus sitios Web.

El tipo de cintas soportadas y la cantidad de drives de lectura/escritura, la cantidad de slots para almacenar las cintas, la interfaz para la conectividad con el servidor, la capacidad máxima de datos que puede albergar, el tamaño físico, entre otras características, son detalladas a continuación.

Uno de los grandes fabricantes es la empresa **Hewlett Packard**. En la figura A.1 y A.2 se observan todos los modelos de librerías de cintas que está comercializando.



Technical specifications	HP StorageWorks MSL2024 Tape Library High-capacity data storage	HP StorageWorks MSL4048 Tape Library Lots of storage in a small space
<b>Number and type of tape drives supported</b>	1 LTO-5 Ultrium 3280 1 LTO-4 Ultrium 1840 1/2 LTO-4 Ultrium 1760 1 LTO-3 Ultrium 960 1/2 LTO-3 Ultrium 920 1/2 LTO-2 Ultrium 448	1/2 LTO-5 Ultrium 3280 1/2 LTO-4 Ultrium 1840 1/4 LTO-4 Ultrium 1760 1/2 LTO-3 Ultrium 960 1-4 LTO-3 Ultrium 920
<b>Number of cartridge slots</b>	24	48
<b>Capacity</b>	38.4 TB Compressed 2:1 19.2 TB Native Maximum, depending on number of drives and drives selected	76.8 TB Compressed 2:1 38.4 TB Native Maximum, depending on number of drives and drives selected
<b>Transfer rate</b>	1.15 TB/hr Maximum, compressed 2:1 Depending on number of drives and drives selected	2.3 TB/hr Maximum Compressed 2:1 Depending on number of drives and drives selected
<b>Host Interface</b>	8 Gb Fibre Channel 4 Gb Fibre Channel Ultra320 (VD SCSI) 3 Gb/sec SAS Ultra160 (VD SCSI) Depending on drive type	8 Gb Fibre Channel 4 Gb Fibre Channel Ultra320 (VD SCSI) 3 Gb/sec SAS Depending on drive type
<b>Encryption capability</b>	AES 256-bit for LTO-5 and LTO-4 Ultrium drives only	AES 256-bit for LTO-5 and LTO-4 Ultrium drives only
<b>Form factor</b>	2U	4U
<b>Warranty</b>	1-year parts exchange, next business day	1-year parts exchange, next business day
<b>QuickSpec URLs</b>	<a href="http://h18000.www1.hp.com/products/quickspecs/13258_div/13258_div.html">http://h18000.www1.hp.com/products/quickspecs/13258_div/13258_div.html</a>	
<b>Recommended Services</b>	<p><b>3-Year HP Support Plus 24 Service:</b> For a higher return on your storage technology, this 3-year combined reactive support option delivers integrated onsite hardware/software support services available 24x7.</p> <p><b>HP StorageWorks ESL, MSL, EML Tape Libraries Installation and Startup:</b> To get your Fibre Channel-based HP Tape Library up-and-running fast, this service provides initial hardware installation of your product into your technology environment (system configuration is not included).</p>	

Fig. A.1 - Librerías de cintas de Hewlett Packard.




			
Technical specifications	HP StorageWorks MSL8048 Tape Library Flexible, reliable, and available	HP StorageWorks MSL8096 Tape Library Full featured with exceptional storage density	HP StorageWorks MSL6020/ MSL6060 Tape Libraries Scalable, flexible, and manageable
<b>Number and type of tape drives supported</b>	1-4 LTO-5 Ultrium 3280 1-4 LTO-4 Ultrium 1840 1-4 LTO-4 Ultrium 1760 1-4 LTO-3 Ultrium 960 1-4 LTO-3 Ultrium 920 1-4 LTO-2 Ultrium 448 Drives supported, none are included.	2-4 LTO-5 Ultrium 3280 2-4 LTO-4 Ultrium 1840 2-4 LTO-4 Ultrium 1760 2-4 LTO-3 Ultrium 960	1-16 LTO-4 Ultrium 1840 1-16 LTO-3 Ultrium 960 Supported, depending on model
<b>Number of cartridge slots</b>	48	96	30 to 240 (Ultrium-based library, multi-unit) Maximum, depending on model
<b>Capacity</b>	76.8 TB Compressed 38.4 TB Native	288 TB Compressed 2:1 144 TB Native Maximum, depending on number of drives and drives selected	384 TB Maximum Compressed 2:1, 8 modules with multi-unit scalability
<b>Transfer rate</b>	11 TB/hr Compressed 2:1	4.0 TB/hr Compressed 2:1 Maximum, depending on number of drives and drives selected	13.8 TB/hr Compressed 2:1 Maximum, 8 modules, 16 drives with multi-unit configuration
<b>Host interface</b>	8 Gb Native Fibre Channel 4 Gb Native Fibre Channel Ultra320 (VD SCSI) 3 Gb/sec SAS Depending on drive type	8 Gb Native Fibre Channel 4 Gb Native Fibre Channel Ultra320 (VD SCSI) 3 Gb/sec SAS Depending on drive type	4 Gb Native Fibre Channel 2 Gb Native Fibre Channel Ultra320 (VD SCSI) Ultra160 (VD SCSI) Depending on model
<b>Encryption capability</b>	AES 256-bit for LTO-5 and LTO-4 Ultrium drives only	AES 256-bit for LTO-5 and LTO-4 Ultrium drives only	AES 256-bit for LTO-4 Ultrium drives only
<b>Form factor</b>	8U	8U	5U or 10U, depending on model
<b>Warranty</b>	1-year onsite, next business day	1-year onsite, next business day	1-year onsite, next business day
<b>QuickSpec URLs</b>	<a href="http://h18000.www1.hp.com/products/quickspecs/13258_div/13258_div.html">http://h18000.www1.hp.com/products/quickspecs/13258_div/13258_div.html</a>		<a href="http://h18006.www1.hp.com/products/quickspecs/11863_na/11863_na.html">http://h18006.www1.hp.com/products/quickspecs/11863_na/11863_na.html</a>
<b>Recommended Services</b>	<b>3-Year HP Support Plus 24 Service:</b> For a higher return on your storage technology, this 3-year combined reactive support option delivers integrated onsite hardware/software support services available 24x7. <b>HP StorageWorks ESL, MSL, EML Tape Libraries Installation and Startup:</b> To get your Fibre Channel-based HP Tape Library up and running fast, this service provides initial hardware installation of your product into your technology environment (system configuration is not included).		<b>3-Year HP Support Plus 24 Service:</b> For a higher return on your storage technology, this 3-year combined reactive support option delivers integrated onsite hardware/software support services available 24x7.

Fig. A.2 - Librerías de cintas de Hewlett Packard.

Otro de los fabricantes que está muy introducido en el mercado es la empresa SUN/Oracle. En la figura A.3 se muestran los modelos de las librerías.






STORAGETEK MODULAR LIBRARY SYSTEMS			
			
<b>Name</b>	<b>StorageTek SL8500 Modular Library System</b>	<b>StorageTek SL3000 Modular Library System</b>	<b>StorageTek SL500 Modular Library System</b>
<b>Features and benefits</b>	<ul style="list-style-type: none"> <li>• <b>Consolidate resources.</b> With extreme scalability and support for mixed media and diverse operating systems, the SL8500 is ideal for consolidating many smaller libraries into one high-performance system.</li> <li>• <b>Increase availability.</b> Minimize downtime with redundant and hot-swappable components such as, drives, power supplies, robotics and electronics.</li> <li>• <b>Increase throughput.</b> High-performance robotics help you keep pace with unpredictable workload spikes and future throughput needs.</li> <li>• <b>Grow with ease.</b> The SL8500 can scale from 1,000 slots to more than 100,000 slots, easily handling years of explosive growth. Upgrade the SL8500 with additional drives and cartridge slots, without system downtime.</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Grow without compromise.</b> The SL3000 midrange library scales from 200 to more than 3,000 slots and from 1 to 56 tape drives. With capacity on demand, install physical capacity in advance and tap into it incrementally.</li> <li>• <b>Minimize risk.</b> Sustain 24/7 operations with many components borrowed from the SL8500 and on-the-fly replacement of robotics, power supplies, and drives. Redundant robotics boost performance and availability.</li> <li>• <b>Stretch storage dollars.</b> Consolidate storage within mainframe and open systems environments, or both. Manage storage efficiently with mixed-media support, no-restriction sharing, and flexible native partitioning.</li> <li>• <b>Simplify administration.</b> Monitor and manage the library locally or remotely. Share the SL3000 across heterogeneous environments.</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Achieve enterprise scalability.</b> Offers industry-leading modular scalability in both capacity and performance.</li> <li>• <b>Provide industry-leading reliability.</b> Leverages SL8500 technology. Designed to last 10+ years. All major parts are FRUs with mean time to replace of less than 30 minutes.</li> <li>• <b>Operate with industry-leading density and low TCO.</b> Buy only the capacity you need when you need it. Consolidate multiple small libraries by creating up to eight logical partitions to save management time, software license fees, and maintenance costs. Reduce environmental costs with high density.</li> <li>• <b>Simplify administration.</b> Oracle's StorageTek Library Console software allows for local or remote library management and monitoring.</li> </ul>
<b>Capacity</b>	<ul style="list-style-type: none"> <li>• 1,000 to 100,000 slots maximum</li> <li>• 1 to 640 drives of any combination of supported drives, including               <ul style="list-style-type: none"> <li>• T10000 (FICON, FC)</li> <li>• T9840 (ESCON, FICON, FC)</li> <li>• T9940 (ESCON, FICON, FC)*</li> <li>• LTO (FC)</li> <li>• DLTS4, SDLT (FC)†</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• 200 to 5,925 slots</li> <li>• 1 to 56 drives of any combination of supported drives, including               <ul style="list-style-type: none"> <li>• T10000 (FICON, FC)</li> <li>• T9840 (ESCON, FICON, FC)</li> <li>• LTO (FC)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• 30 to 575 slots, LTO only (FC, SCSI, SAS)</li> <li>• 1 to 18 drives</li> <li>• Expand with drive expansion modules or cartridge expansion modules.</li> <li>• Capacity on demand allows for scalability in small increments as needed.</li> </ul>
<b>Operating systems</b>	A wide array including Oracle Solaris, Windows, z/OS, AS/400, AIX, HP-UX, and Linux	A wide array including Oracle Solaris, Windows, z/OS, AS/400, AIX, HP-UX, and Linux	Oracle Solaris, Windows, UNIX, and Linux
STORAGETEK TAPE SOLUTIONS			
			
<b>Name</b>	<b>StorageTek SL48 Tape Library</b>	<b>StorageTek SL24 Tape Autoloader</b>	
<b>Features and benefits</b>	<ul style="list-style-type: none"> <li>• <b>Integrate easily.</b> Extensive testing enables the SL48 to work with a variety of servers, operating systems, and ISV packages. And its support for multiple interfaces makes it ideal for backup and archiving. Supports SCSI, SAS, and FC interfaces.</li> <li>• <b>Keep it simple.</b> SL48 tape libraries arrive rack-ready and include Web-based management and an integrated bar code reader.</li> <li>• <b>Protect your investments.</b> Native capacity of up to 72 TB simplifies backup/recovery jobs, and the 4U form factor saves rack space. Tape drive and interface choices provide investment protection. Supports up to four logical partitions, making consolidation easy.</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Increase reliability and flexibility.</b> Ideal for backups for midsized businesses and first-time automation users, the SL24 supports SCSI, SAS, and FC interfaces.</li> <li>• <b>Simplify installation and use.</b> SL24 tape autoloaders arrive rack-ready, and their modular design and removable magazines make it easy to load/unload data cartridges and upgrade drives.</li> <li>• <b>Boost ROI.</b> Direct attach capability and support for up to two drives allow for a native capacity of up to 36 TB. The broad choice of tape drives and interfaces provides investment protection. Supports up to two logical partitions.</li> </ul>	
<b>Capacity</b>	<ul style="list-style-type: none"> <li>• 48 slots, LTO only</li> <li>• Maximum of four half-height, two full-height, or two half-height with one full-height</li> </ul>	<ul style="list-style-type: none"> <li>• 24 slots, LTO only</li> <li>• Maximum of one full-height or two half-height drives</li> </ul>	
<b>Operating systems</b>	Oracle Solaris, Windows, UNIX, and Linux	Oracle Solaris, Windows, UNIX, and Linux	

Fig. A.3 - Librerías de cintas de SUN/Oracle.

La empresa **IBM** tiene en su cartera de productos los siguientes modelos de librerías de cintas que se muestra en la figura A.4

Productos de cinta empresariales/de gama media			
	librerías de cintas de gama media		
	TSS310	TSS500	
<b>Nombre</b>	3576	3584	<b>Controladora de cintas empresarial para System z</b>
<b>Tipo</b>	3576	3584	3592
<b>modelo de máquina, I/VE/C#</b>	L5B E9U	L53 D53 S54	Controladora de cintas para System z
<b>Puntos fuertes del producto</b>	Varios SO Alto rendimiento Gran capacidad Diseño modular	Varios SO Alto rendimiento Gran capacidad Tecnología de ranuras de alta densidad (HD) Compatibilidad con LTFS Compatibilidad con TSLM Admite unidades LTO6, LTO5 y LTO4	3592 C07
<b>Número de unidades</b>	1-18	1-192 Hasta 2700 en un complejo	3584
<b>Número máx. de cartuchos<sup>1</sup></b>	409	20.000, 300.000 en un complejo	TS3500
<b>Cifrado de una escritura, muchas lecturas (WORM)</b>	si/si	si/si	
<b>Capacidad nativa</b>	Gen 6: 1022,5 TB Gen 5: 613,5 TB Gen 4: 327,2 TB Gen 3: 163,6 TB	50 PB por librería. Hasta 750 PB en un complejo	
<b>Capacidad típica<sup>2</sup></b>	Gen 6: 2,56 petabytes (PB) Gen 5: hasta 1,22 PB	25 PB por librería. Hasta 1,675 exabytes (EB) en un complejo	
<b>Rendimiento nativo (sin comprimir)</b>	Gen 6: hasta 2,88 gigabytes por segundo (GBps) Gen 5: hasta 2,52 GBps	26,9 GBps	
<b>Interfaz</b>	FC a 4 Gbps FC a 8 Gbps (LTO6 y 5 únicamente) SAS a 3 Gbps (LTO4 únicamente) SCSI LVD (LTO3 únicamente)	FC a 8 Gbps	
<b>Librerías de cintas admitidas</b>	N/A	N/A	
<b>Compatibilidad con plataformas<sup>3</sup></b>	System p, System i, System x, Microsoft Windows, HP-LUX, Sun Solaris, Linux	System p, System i, System x, Microsoft Windows, HP-LUX, Sun Solaris, Linux	
<b>Compatibilidad con aplicaciones<sup>4</sup></b>	A, B, C, D, E, F, G, H, J, L <sup>5</sup>	A, B, C, D, E, F, G, H, J, L <sup>5</sup>	
<b>Soportes</b>	Consulte la sección Soportes de cinta, página 8	Consulte la sección Soportes de cinta, página 8	
<b>Periodo de garantía</b>	Un año	Un año	
<b>Tipo de garantía</b>	Siguiente día laborable (de 09:00 a 17:00)	Reparación in situ (24 horas al día)	
			Conexión de System z a unidades TS1140, TS1130 y TS1120 Alto rendimiento
			Compatibilidad con diversas plataformas Gestión avanzada Escalaible Tecnología de ranuras HD
			1-16
			1-192 en una cadena de librerías Hasta 2700 en un complejo de librerías
			si/si
			si/si
			N/A
			15.000 por librería <sup>1</sup> 225.000 en un complejo de librerías
			N/A
			Hasta 60 PB por librería 900 PB en un complejo
			N/A
			Hasta 180 PB Hasta 2,7 EB en un complejo
			Puede variar <sup>6</sup>
			Hasta 48 GBps
			FC a 8 Gbps TS3500
			FC a 8 Gbps
			N/A
			System p, System i, System x, System z; Microsoft Windows; HP-LUX; Sun Solaris; Linux A, B, C, D, E, F, G, H, J, L <sup>5</sup>
			System p, System i, System x, System z; Microsoft Windows; HP-LUX; Sun Solaris; Linux Consulte la unidad
			N/A
			Consulte la sección Soportes de cinta, página 8
			Un año
			Un año
			Reparación in situ (24 horas al día)
			Reparación in situ (24 horas al día)

Fig. A.4 - Librerías de cintas de IBM

## ANEXO B – LIBRERÍAS DE CINTAS VIRTUALES

En este apartado, se muestran las características técnicas principales de las librerías de cintas virtuales según la información provista por los fabricantes en sus sitios Web.

Características como el tipo de RAID que utiliza, cantidad de librerías y drives que se pueden crear, tasa de transferencia, capacidad usable de discos, tipo de deduplicación, replicación, entre otras, son detalladas en la documentación incluida a continuación.

El primero de los fabricantes es la empresa **Hewlett Packard**. En la figura B.1 se observan todos los modelos de librerías basadas en discos que esta comercializando.

Technical specifications	HP StorageWorks D2D2500 Backup System Provides consolidated, disk-based data protection in an intelligent self-managing 1U rack-mountable solution	HP StorageWorks D2D4000 Backup System Provides consolidated, disk-based data protection for small and medium size data centers in an intelligent self-managing 2U rack-mountable solution	HP StorageWorks D2D4112 Backup System Provides scalable consolidated, disk-based data protection for small and medium size data centers in an intelligent self-managing 2U rack-mountable solution
<b>Models</b>	D2D2502, D2D2504	D2D4004, D2D4004L, D2D400R, D2D4009L	D2D4112
<b>Form factor</b>	1U	2U	2U
<b>Total raw capacity</b>	2 TB (D2D2502) 4 TB (D2D2504)	4 TB (D2D4004) expandable to 9 TB 9 TB (D2D4009)	12 TB scalable to 24 TB
<b>Total usable capacity</b>	1.5 TB (D2D2502) 3 TB (D2D2504)	3 TB (D2D4004) expandable to 7.5 TB 7.5 TB (D2D4009)	9 TB scalable to 18 TB
<b>Transfer rate (Maximum)</b>	180 GB/hr (D2D2502) 288 GB/hr (D2D2504)	325 GB/hr (D2D4004L, D2D4009, D2D4009L) 250 GB/hr (D2D4004)	700 GB/hr
<b>Interface</b>	1 Gb iSCSI (2) Ports	Two 1 Gb iSCSI Ports (D2D4004) Two 4 Gb Fibre Channel Ports and Two 1 Gb iSCSI Ports (D2D4004L) Two 4 Gb Fibre Channel Ports and Two 1 Gb iSCSI Ports (D2D4009L) Two 1 Gb iSCSI Ports (D2D4009)	Two 4 Gb Fibre Channel Ports and Two 1 Gb iSCSI Ports
<b>Drive type</b>	500 GB 3G 7.2K iFF SATA, Included (D2D2502) 1 TB 3G 7.2K iFF MDL SATA, Included (D2D2504)	750 GB 3G 7.2K iFF SATA	1 TB 3G 7.2K iFF MDL SATA
<b>Number of drives (in base unit)</b>	4	6 (D2D4004) 12 (D2D4009)	12
<b>RAID support</b>	Hardware RAID 6	Hardware RAID 5	Hardware RAID 5
<b>Data deduplication and replication</b>	Dynamic deduplication and low bandwidth replication	Dynamic deduplication and low bandwidth replication	Dynamic deduplication and low bandwidth replication
<b>Storage expansion options</b>	None	D2D4004 capacity upgrade kit (D2D4004 only)	D2D4112 capacity upgrade kit
<b>Tape devices emulated</b>	HP LTO-2/LTO-3/LTO-4/LTO-5 tape drives in 1/8 G2 Autoloader, MS12024, MS14048, HP D2D Backup System Generic Library	HP LTO-2/LTO-3/LTO-4/LTO-5 tape drives in 1/8 G2 Autoloader, MS12024, MS14048, HP D2D Backup System Generic Library	HP LTO-2/LTO-3/LTO-4/LTO-5 tape drives in 1/8 G2 Autoloader, MS12024, MS14048, HP D2D Backup System Generic Library
<b>Number of Virtual Tape Libraries and NAS Targets</b>	4 (D2D2502) 8 (D2D2504)	16	24
<b>Number of virtual tape cartridges emulated</b>	192 (D2D2502) 384 (D2D2504)	1536	3456
<b>Targets for backup applications</b>	Virtual tape library and NAS (CIFS) CIFS = Common Internet File System	Virtual tape library and NAS (CIFS)	Virtual tape library and NAS (CIFS)
<b>Direct attach tape drives Note: VTL targets only</b>	HP LTO-2/LTO-3/LTO-4/LTO-5 tape drives in 1/8 G2 Autoloader, MS12024, MS14048	HP LTO-2/LTO-3/LTO-4/LTO-5 tape drives in 1/8 G2 Autoloader, MS12024, MS14048	HP LTO-2/LTO-3/LTO-4/LTO-5 tape drives in 1/8 G2 Autoloader, MS12024, MS14048
<b>Warranty</b>	1 year, next-day, parts exchange, limited warranty, 9x5 phone	1 year parts, labor, onsite next business day response	1 year parts, labor, onsite next business day response
<b>QuickSpec URL</b>	<a href="http://h18000.www1.hp.com/products/quickspecs/12318_na/12318_na.html">http://h18000.www1.hp.com/products/quickspecs/12318_na/12318_na.html</a>		
<b>Recommended Services</b>	<b>3-Year HP Support Plus 24:</b> Available 24x7, the 3-year combined reactive support option delivers onsite hardware support and over-the-phone software support around-the-clock. Leverage the full strength of HP Technology Services—customers can trust the service professionals at HP to work collaboratively with them, putting our strategic and technical know-how to work across their entire infrastructure.		

Fig. B.1 - Librerías de cintas Virtuales de Hewlett Packard.

A continuación se muestran los productos ofrecidos por **SUN/Oracle**. El fabricante dispone 2 líneas de productos para las librerías virtuales. Una de las líneas es la **versión Prime**, la cual ofrece diversos modelos para las **series R y C**. En la figura B.2 se pueden ver los modelos de VTL Prime que comercializa la empresa con alguna de sus características.

#### StorageTek VTL Prime Specifications

Storage	VTL Prime R-Series			VTL Prime C-Series		
	VTL4302R-Z ST2540	VTL4305R-Z ST2540	VTL4310R-Z ST2540	VTL4320C-Z ST6140	VTL4329C-Z ST6140	VTL4365C-Z ST6140
VTL cache (TB)	1	2	4	8	13.5	26
Dedupe repository (TB)	2.5	5	10	20	29	65
Server	X4200	X4200	X4200	X4600	X4600	X4600
Host interface	ISCSI/FC*	ISCSI/FC*	FC	FC	FC	FC
# of dual port 4Gig FC HBA	1	1	2	2	2	2
Storage interface	FC	FC	FC	FC	FC	FC
<b>VTL software</b>						
Tape libraries				128		
Tape drives				1,024		
Tape cartridges				64,000		

\* Optional item

Fig. B.2 - Librerías de cintas Virtuales de SUN/Oracle. - Línea PRIME -

SUN/Oracle por otra parte, comercializa la línea **Plus** con los modelos mostrados en la figura B.3.

#### Sun StorageTek® Virtual Tape Library Plus Specifications

	Capacity, raw (uncompressed)	Host/total connectivity*	Performance	Virtual tape libraries/drives/carts	Appliance package
VTL1140 appliance racks	8-56 TB raw	2/6 2 Gb/sec FC	Up to 350 MB/sec	16/128/2048	One 19-in. rack
VTL2540 appliance racks	8-224 TB raw	8/28 2 Gb/sec FC	Up to 700 MB/sec	64/512/4096	Up to four 19-in.
VTL3540 appliance racks	8-224 TB raw	16/56 2 Gb/sec FC	Up to 1400 MB/sec	128/1024/8192	Up to four 19-in.

\*This is the maximum host connectivity not the base connectivity specifications.

Fig. B.3 - Librerías de cintas Virtuales de SUN/Oracle. - Línea PLUS -

La diferencia más importante entre ambas líneas, es que la línea Plus no incluye deduplicación de datos.

La empresa **EMC**, la cual se ubica en una excelente posición dentro del mercado con sus productos para las Librerías Virtuales con sus modelos Datadomain ofrece los modelos de la figura B.4.



## Especificaciones del sistema Data Domain



	DD140	DD610	DD630	DD670 <sup>3</sup>	DD860 <sup>3</sup>	DD890 <sup>3</sup>	GDA <sup>3</sup>	DD Archiver <sup>3</sup>
<b>Capacidad lógica</b> <sup>1, 2</sup>	Entre 9 y 43 TB	Entre 40 y 195 TB	Entre 84 y 420 TB	Entre 0.6 y 2.7 PB	Entre 1.4 y 7.1 PB	Entre 2.9 y 14.2 PB	Entre 5.7 y 28.5 PB	Entre 5.7 y 28.5 PB
<b>Capacidad cruda</b> <sup>2</sup>	1.5 TB	Hasta 6 TB	Hasta 12 TB	Hasta 76 TB	Hasta 192 TB	Hasta 384 TB	Hasta 768 TB	Hasta 768 TB
<b>Máx. rendimiento (otros)</b>	450 GB/h	675 GB/h <sup>4</sup>	1.1 TB/h <sup>4</sup>	3.6 TB/h <sup>5</sup>	5.1 TB/h <sup>5</sup>	8.1 TB/h <sup>6</sup>	10.7 TB/h <sup>6</sup>	4.3 TB/h <sup>7</sup>
<b>Máx. rendimiento (DD Boost)</b>	490 GB/h	1.3 TB/h	2.1 TB/h	5.4 TB/h	9.8 TB/h	14.7 TB/h	26.3 TB/h	9.8 TB/h

1. Combinación de datos de respaldo de empresa típicos (sistemas de archivos, bases de datos, correos electrónicos y archivos de desarrolladores). El valor inferior del rango de capacidad representa un respaldo completo semanal o mensual y un respaldo incremental diario o semanal, utilizando toda la capacidad del sistema. El valor superior del rango representa un respaldo completo diario, utilizando toda la capacidad del sistema.
2. Todos los valores de capacidad se calculan en base 10 (es decir, 1 TB = 1,000,000,000,000 bytes) y utilizando la configuración con mayor capacidad cruda.
3. Incluye soporte para estantes adicionales (disponibles por separado).
4. El rendimiento máximo se alcanzó mediante una interfaz VTL y Fibre Channel de 4 Gbps.
5. El rendimiento máximo se alcanzó mediante Symantec OpenStorage y Ethernet de 10 Gb.
6. El rendimiento máximo se alcanzó mediante una interfaz VTL y Fibre Channel de 8 Gbps.
7. El rendimiento máximo se alcanzó mediante NFS y Ethernet de 10 Gb.

Fig. B.4 - Librerías de cintas Virtuales de EMC - DataDomain -

El último fabricante a mencionar es la empresa **IBM** que dentro de su gama de backup a disco ofrece a sus clientes los modelos que se muestran en la figura B.5.

## Enterprise Tape Products/Other Backup Tape Products



### Virtualización de cinta en sistemas abiertos de gama media

#### (Dispositivos IBM ProtecTIER y gateway IBM ProtecTIER)

### Virtualización de cintas empresarial para System z

#### IBM Virtualisation Engine

Producto	TS7620	TS7650A	TS7650G	TS7720	TS7740
Tipo modelo de máqui	3959 SM2	3958 AP1	3596DD4	3957 VEB (requiere modelos y tipos de máquina adicionales)	3957 V07 (requiere modelos y tipos de máquina adicionales)
Puntos fuertes del producto	Elimina datos redundantes en una relación de hasta 25:1	Elimina datos redundantes en una relación de hasta 25:1	Elimina datos redundantes en una relación de hasta 25:1	Aumenta el rendimiento Escalable Un gran tamaño de caché para una recuperación rápida	Aumenta el rendimiento Escalable Ayuda a reducir los costos
Número de unidades	Hasta 64	Hasta 256, hasta 512 en cluster	Hasta 256*, hasta 512* en cluster	Hasta 256* Hasta 1536 (GRID para 6 sitios***)	Hasta 256* Hasta 1536 (GRID para 6 sitios***)
Cifrado de una escritura, muchas lecturas (WORM)	no/no	no/no	no/no	si/si	si/si
Número de unidades	8192	Hasta 128.000	Hasta 512.000*	Hasta 4.000.000*	Hasta 4.000.000*
Capacidad nativa	Hasta 11 TB	Hasta 36 TB	Hasta 1 PB	Hasta 624 TB Hasta 3,7 PB (GRID para 6 sitios***)	Hasta 28 TB Hasta 1728 TB (GRID para 6 sitios***)
Capacidad típica <sup>†</sup>	Hasta 275 TB (capacidad nominal basada en una relación de deduplicación de 25:1)	Hasta 900 TB (capacidad nominal basada en una relación de deduplicación de 25:1)	Hasta 25 TB (capacidad nominal basada en una relación de deduplicación de 25:1)	Hasta 1872 PB	Hasta 84 TB
Rendimiento nativo (sin comprimir)	Hasta 100 MBps	Hasta 500 MBps	Hasta 1600 MBps, hasta 2500 MBps en cluster	Hasta 900 MBps	Hasta 900 MBps
Interfaz	FC a 8 Gbps	FC a 8 Gbps	FC a 8 Gbps	FC a 8 Gbps	FC a 8 Gbps
Librerías de cintas admitidas	N/A	N/A	N/A	TS3500, en una red con TS7740	TS3500, 3494**
Compatibilidad con plataformas <sup>‡</sup>	System p, System i, System x, Microsoft Windows, HP-LUX, Sun Solaris, Linux	System p, System i, System x, Microsoft Windows, HP-LUX, Sun Solaris, Linux	System p, System i, System x, Microsoft Windows, HP-LUX, Sun Solaris, Linux	System z	System z
Compatibilidad con aplicaciones <sup>§</sup>	A, B, C, E, F, G, H <sup>†</sup>	A, B, C, E, F, G, H <sup>†</sup>	A, B, C, D, E, F, G, H <sup>†</sup>	A <sup>†</sup>	A <sup>†</sup>
Soportes	N/A	N/A	N/A	N/A	N/A
Periodo de garantía	Un año	Un año	Un año	Un año	Un año
Tipo de garantía	Reparación in situ (24 horas al día)	Reparación in situ (24 horas al día)	Reparación in situ (24 horas al día)	Reparación in situ (24 horas al día)	Reparación in situ (24 horas al día)

\* Especificación de recurso virtual

\*\* Este producto ha dejado de comercializarse

\*\*\* Capacidad típica para seis modelos similares, admite la combinación de modelos

Fig. B.5 - Librerías de cintas Virtuales de IBM.

## **ANEXO C – ESTUDIO SELECCIÓN HARDWARE**

### **C.1 DIMENSIONAMIENTO HARDWARE SITIO CURITIBA**

Para la selección de la librería, es necesario dimensionar la cantidad de cintas que se van a necesitar de acuerdo al tamaño de datos a resguardar y las políticas de retención. Como se sabe el proveedor elegido es la empresa HP según el resultado de la matriz de selección.

En la tabla C.1, se indica la configuración para el modelo HP MSL4048 con 4 drives LTO-5 de fibra óptica. Según la cantidad de datos y la política de retención de cintas se puede observar que se van a necesitar aproximadamente 108 cintas LTO-5 en total para el resguardo de toda la información, utilizando 18 para los backups incrementales y 22 para backup full, teniendo 66 cintas de backup externalizadas. El mínimo de slots necesarios es de 41 que serán ocupados dentro de la librería. También es mencionado el costo aproximado de inversión.

<b>Configuración Librería - Curitiba</b>	
Modelo Librería	MSL4048
Slots	48
Drives LTO-5	4
Costo	25.000 €
<hr/>	
Volumen Full (GB)	64000
Volumen Incremental (GB)	3600
Volumen por cinta (GB)	3000
Backup Full/semana	1
Backup Inc./semana	5
Retención backup Full (semanas)	3
Retención backup Inc. (semanas)	3
<hr/>	
N° de cintas para backup Full	22
N° de cintas Full externalizadas	66
N° de cintas para backup Inc.	18
<hr/>	
<b>Número de cintas total</b>	<b>108</b>
<b>Número de slots ocupados</b>	<b>41</b>

Tabla C.1 - Configuración librería Curitiba

Para definir el tamaño de la VTL se utilizó el DataDomain Deduplication Calculator provisto por el proveedor EMC.

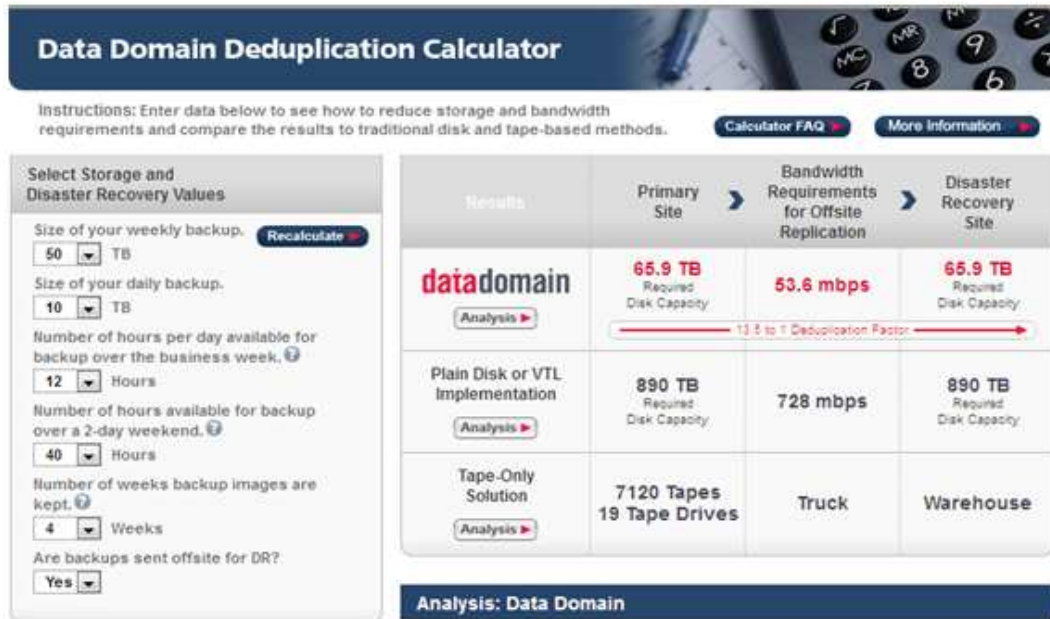


Fig. C.1 - Dimensionamiento DataDomain

Esta herramienta es una estimación de lo necesario para el volumen actual de datos, sin embargo el tamaño del Datadomain se debe calcular de manera mas fina, ya que los datos dependiendo de su formato, tienen una deduplicación de 50:1 mas la compresión, lo que disminuye en gran medida la ocupación en disco.

## C.2 DIMENSIONAMIENTO HARDWARE SITIO CÓRDOBA

En la tabla C.2, se indica la configuración para el modelo HP MSL4048 con 4 drives LTO-4 con interfaz de conexión SAS. Según la cantidad de datos y la política de retención de cintas se van a necesitar aproximadamente 33 cintas LTO-4 en total. Para los backups incrementales se utilizan 12 cintas y 5 para backup full, teniendo 15 cintas de backup externalizadas. Los slots necesarios en un principio es de 18 que serán ocupados dentro de la librería. También es mencionado el costo aproximado de inversión.

Configuración Librería - Córdoba	
Modelo Librería	MSL4048
Slots	48
Drives LTO-4	4
Costo	19.000 €
Volumen Full (GB)	7900
Volumen Incremental (GB)	1100
Volumen por cinta (GB)	1600
Backup Full/semana	1
Backup Inc./semana	5
Retención backup Full (semanas)	3
Retención backup Inc. (semanas)	3
N° de cintas para backup Full	5
N° de cintas Full externalizadas	15
N° de cintas para backup Inc.	12
<b>Número de cintas total</b>	<b>33</b>
<b>Número de slots ocupados</b>	<b>18</b>

Tabla C.2 - Configuración librería Córdoba

### C.3 DIMENSIONAMIENTO HARDWARE SITIO BUENOS AIRES

Para el sitio de Buenos Aires, la propuesta es una librería HP MSL2024 con un solo drive LTO-4 con interfaz de conexión SAS. Para la cantidad de datos y la política de retención, se necesitan 15 cintas LTO-4 aproximadamente en total para el resguardo de toda la información. Se utilizan 6 para backup incrementales y 2 para backup full, teniendo 6 cintas de backup externalizadas. El mínimo de slots necesarios es de 9 que serán ocupados dentro de la librería. También es mencionado el costo aproximado de inversión.

Configuración Librería - Bs. As.	
Modelo Librería	MSL2024
Slots	24
Drives LTO-4	1
Costo	11.000 €
Volumen Full (GB)	1800
Volumen Incremental (GB)	400
Volumen por cinta (GB)	1600
Backup Full/semana	1
Backup Inc./semana	5
Retención backup Full (semanas)	3
Retención backup Inc. (semanas)	3
N° de cintas para backup Full	2
N° de cintas Full externalizadas	6
N° de cintas para backup Inc.	6
<b>Número de cintas total</b>	<b>15</b>
<b>Número de slots ocupados</b>	<b>9</b>

Tabla C.3 - Configuración librería Buenos Aires