



Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

**ESTUDIO DE FACTIBILIDAD TÉCNICA PARA EL USO DE UN BANCO DE
DATOS VIRTUAL POR PARTE DE LAS EMPRESAS GUATEMALTECAS
COMO SERVICIO DE RESPALDO DE INFORMACIÓN**

Juan Vidal Luis

Asesorado por el Ing. Luis Fernando Quiñonez López

Guatemala, agosto de 2012

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**ESTUDIO DE FACTIBILIDAD TÉCNICA PARA EL USO DE UN BANCO DE
DATOS VIRTUAL POR PARTE DE LAS EMPRESAS GUATEMALTECAS
COMO SERVICIO DE RESPALDO DE INFORMACIÓN**

TRABAJO DE GRADUACIÓN

PRESENTADO A JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

JUAN VIDAL LUIS

ASESORADO POR EL ING. LUIS FERNANDO QUIÑONEZ LÓPEZ

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, AGOSTO DE 2012

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERÍA



NÓMINA DE JUNTA DIRECTIVA

DECANO	Ing. Murphy Olympto Paiz Recinos
VOCAL I	Ing. Alfredo Enrique Beber Aceituno
VOCAL II	Ing. Pedro Antonio Aguilar Polanco
VOCAL III	Ing. Miguel Ángel Dávila Calderón
VOCAL IV	Br. Juan Carlos Molina Jiménez
VOCAL V	Br. Mario Maldonado Muralles
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Murphy Olympto Paiz Recinos
EXAMINADOR	Ing. Marlon Antonio Pérez Turk
EXAMINADORA	Inga. Floriza Felipa Ávila Pesquera
EXAMINADORA	Inga. Sonia Yolanda Castañeda Ramírez
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

ESTUDIO DE FACTIBILIDAD TÉCNICA PARA EL USO DE UN BANCO DE DATOS VIRTUAL POR PARTE DE LAS EMPRESAS GUATEMALTECAS COMO SERVICIO DE RESPALDO DE INFORMACIÓN

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería en Ciencias y Sistemas, con fecha agosto de 2012.



Juan Vidal Luis

Guatemala, 03 de mayo de 2012

Señores

Comisión de Revisión de Trabajos de Graduación

Carrera de Ingeniería en Ciencias y Sistemas

Facultad de Ingeniería

Universidad de San Carlos de Guatemala

Guatemala, Ciudad

Respetables Señores:

El motivo de la presente es para informarles que como asesor del estudiante **JUAN VIDAL LUIS**, he procedido a revisar el trabajo de graduación titulado **“ESTUDIO DE FACTIBILIDAD TECNICA PARA EL USO DE UN BANCO DE DATOS VIRTUAL POR PARTE DE LAS EMPRESAS GUATEMALTECAS COMO SERVICIO DE RESPALDO DE INFORMACION”** y de acuerdo a mi criterio, el mismo se encuentra concluido y cumple con los objetivos definidos al inicio.

Sin otro particular, atentamente me suscribo de ustedes.



Ing. Luis Fernando Quiñonez López

Asesor de trabajo de graduación

Colegiado: 7514

LUIS QUIÑÓNEZ
INGENIERO EN C.C. Y SISTEMAS
COLEGIADO No. 7514



Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 06 de Junio de 2012

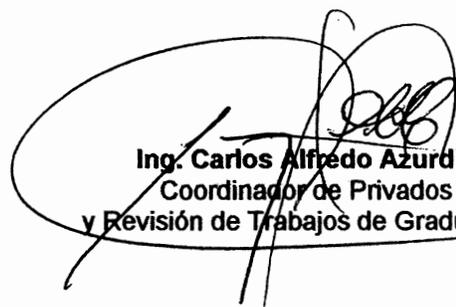
Ingeniero
Marlon Antonio Pérez Turk
Director de la Escuela de Ingeniería
En Ciencias y Sistemas

Respetable Ingeniero Pérez:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante **JUAN VIDAL LUIS** carné **2004-13517**, titulado: **"ESTUDIO DE FACTIBILIDAD TECNICA PARA EL USO DE UN BANCO DE DATOS VIRTUAL POR PARTE DE LAS EMPRESAS GUATEMALTECAS COMO SERVICIO DE RESPALDO DE INFORMACION"**, y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,


Ing. Carlos Alfredo Azurdia
Coordinador de Privados
y Revisión de Trabajos de Graduación



UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA



FACULTAD DE INGENIERÍA
ESCUELA DE CIENCIAS Y SISTEMAS
TEL. 24767644

E
S
C
U
E
L
A

D
E

C
I
E
N
C
I
A
S

Y

S
I
S
T
E
M
A
S

*El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, del trabajo de graduación titulado **“ESTUDIO DE FACTIBILIDAD TÉCNICA PARA EL USO DE UN BANCO DE DATOS VIRTUAL POR PARTE DE LAS EMPRESAS GUATEMALTECAS COMO SERVICIO DE RESPALDO DE INFORMACIÓN”**, presentado por el estudiante JUAN VIDAL LUIS, aprueba el presente trabajo y solicita la autorización del mismo.*

“ID Y ENSEÑAD A TODOS”



Ing. Marvin Antonio Pérez Turk
Director, Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 09 de agosto 2012



El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al trabajo de graduación titulado: **ESTUDIO DE FACTIBILIDAD TÉCNICA PARA EL USO DE UN BANCO DE DATOS VIRTUAL POR PARTE DE LAS EMPRESAS GUATEMALTECAS COMO SERVICIO DE RESPALDO DE INFORMACIÓN**, presentada por el estudiante universitario **Juan Vidal Luis**, autoriza la impresión del mismo.

IMPRÍMASE:

Ing. Murphy Olympo Paiz Recinos
Decano

Guatemala, 24 de agosto de 2012.



/gdech

AGRADECIMIENTOS A:

- Dios** Por guiarme siempre por el camino correcto y bendecirme en cada momento de la vida.
- Mis amigos** Por apoyarme en el largo y complicado viaje en la Universidad. Por compartir las alegrías y ayudar a cargar las penas. Por recordarme que siempre contaré con ellos.
- Los catedráticos y auxiliares conscientes** Por compartir su conocimiento y guardar el prestigio de la carrera. Por ayudarme a ver que me falta mucho por aprender y demostrar que el esfuerzo extra para evitar la mediocridad vale la pena.
- Ing. Francisco Lobos e Ing. Luis Fernando Quiñonez López** Por el tiempo dedicado y sus valiosas recomendaciones para la elaboración de este trabajo.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES.....	VII
GLOSARIO.....	IX
RESUMEN.....	XIII
OBJETIVOS.....	XV
INTRODUCCIÓN.....	XVII
1. MARCO TEÓRICO.....	1
1.1. Cloud-Computing.....	1
1.2. Infraestructura	3
1.2.1. Componentes de Cloud-Computing.....	3
1.2.1.1. Cliente	4
1.2.1.2. Centro de datos	5
1.2.1.3. Servidores distribuidos	5
1.3. Virtualización	6
1.3.1. Concepto.....	6
1.3.2. Paravirtualización	7
1.3.3. Cloud-Computing y virtualización	8
1.4. Pay-as-use	8
1.4.1. Software como Servicio (Software as a Service - SaaS)	11
1.4.2. Plataforma como servicio (Plataform as a Service - PaaS)	12
1.4.3. <i>Hardware</i> como Servicio (<i>Hardware</i> as a Service - HaaS).....	13

2.	ANTECEDENTES	15
2.1.	Implementación de Cloud-Computing en las empresas	15
2.1.1.	Amazon EC2.....	15
2.1.2.	Google Apps	22
2.1.3.	Implementación de Cloud-Computing.....	25
2.1.4.	Implicaciones	27
2.1.5.	Ventajas y desventajas del modelo de Cloud-Computing	27
2.2.	Servicios orientados a Cloud-Computing, cambios y oportunidades.....	29
2.2.1.	Video.....	29
2.2.1.1.	Video <i>streaming</i>	29
2.2.2.	Hospedaje de archivos.....	30
2.2.3.	Música.....	30
2.2.4.	Imágenes	30
2.2.5.	Pagos y transferencias.....	30
2.2.6.	Anuncios publicitarios	31
2.2.7.	Sistemas operativos.....	31
2.2.8.	Virtualización.....	31
2.2.9.	Ofimática	32
2.3.	Cambios y oportunidades.....	32
3.	RESPALDO DE INFORMACIÓN DIGITAL	35
3.1.	Introducción a los conceptos de respaldo de datos.....	35
3.1.1.	Razones para hacer un respaldo de datos	37
3.1.1.1.	Respaldo de información.....	37
3.1.1.2.	Fallas en el medio	38
3.1.1.3.	Operación incorrecta	38
3.1.1.4.	Ataques externos.....	38

3.1.1.5.	Incompatibilidad en el <i>hardware</i>	39
3.1.1.6.	<i>Bugs</i> o problemas en <i>Drivers</i> o en aplicaciones	39
3.1.1.7.	Desaparición del medio	39
3.1.2.	Costo de pérdida de información	40
3.1.2.1.	Pérdida de clientes	40
3.1.2.2.	Pérdida de órdenes	41
3.1.2.3.	Moral.....	41
3.1.2.4.	Imagen.....	41
3.1.2.5.	Presupuesto	42
3.1.2.6.	Tiempo.....	42
3.1.3.	Tipos de <i>Backup</i>	44
3.1.3.1.	Respaldo total.....	44
3.1.3.2.	Respaldos incrementales	45
3.1.3.3.	Respaldos diferenciales.....	45
3.1.4.	Estrategias de respaldo.....	46
3.1.4.1.	Esquema de tres cintas	48
3.1.4.2.	Esquema de seis cintas.....	48
3.1.4.3.	Esquema de 19 cintas	49
3.1.4.4.	Esquema de 10 cintas	49
3.1.1.	<i>Hardware</i> para <i>Backup</i>	49
3.1.1.1.	Confiabilidad.....	50
3.1.1.2.	Ciclo de trabajo.....	51
3.1.1.3.	Velocidad de transferencia	52
3.1.1.4.	Flexibilidad.....	52
3.1.1.5.	Tiempo-de-datos.....	53
3.1.1.6.	Capacidad	53
3.1.1.7.	Amovilidad	54
3.1.1.8.	Costo	54

3.1.1.9.	Cintas magnéticas	55
3.1.1.10.	Unidades ópticas	56
3.1.1.11.	Unidades de disco	57
3.1.1.12.	Otros dispositivos	58
3.2.	Importancia del respaldo de datos de la información digital para las grandes y medianas empresas	58
4.	BANCO DE DATOS VIRTUAL	63
4.1.	Adaptación más que un nuevo paradigma	63
4.2.	Reducción de costos	65
4.3.	Disponibilidad	67
4.4.	Seguridad	68
4.5.	Banco de Datos Virtual.....	69
5.	IMPLEMENTACIÓN DE UN BANCO DE DATOS VIRTUAL	73
5.1.	<i>DataCenter</i>	73
5.1.1.	SAN / NAS	73
5.1.1.1.	Características principales de una SAN	74
5.1.1.1.1.	Rendimiento	74
5.1.1.1.2.	Disponibilidad	74
5.1.1.1.3.	Escalabilidad	74
5.1.1.1.4.	Seguridad.....	75
5.1.2.	iSCSI.....	76
5.2.	Infraestructura	77
5.3.	Componentes	79
5.3.1.	Sistema operativo de Cloud-Computing.....	79
5.3.2.	Plataforma de virtualización	80
5.3.3.	<i>DataCenter</i>	80
5.4.	Implementación	81

5.4.1.	<i>Hardware</i>	81
5.4.1.1.	Servidor Cloud-Computing	82
5.4.1.2.	Servidor SAN/NAS.....	82
5.4.1.3.	Servidor <i>DataCenter</i>	82
5.4.1.4.	Switch Lan	82
5.4.2.	Software.....	83
5.4.2.1.	VMWare ESXi.....	83
5.4.2.2.	VMWare vSphere	86
5.4.2.3.	OpenFiler.....	88
5.4.3.	Factibilidad técnica de implementar un Banco de Datos Virtual.....	90
CONCLUSIONES		93
RECOMENDACIONES.....		95
BIBLIOGRAFÍA		97

ÍNDICE DE ILUSTRACIONES

FIGURAS

1.	La nube en el diagrama de red representa al <i>Internet</i>	1
2.	Componentes que intervienen en Cloud-Computing.....	4
3.	Esquema de SaaS.....	12
4.	Esquema de PaaS.....	13
5.	Esquema de HaaS.....	14
6.	Esquema del <i>hardware</i> mínimo de la implementación del Banco de datos virtual.....	81
7.	Servidor ESXi administrando una granja de servidores SAN/NAS en un <i>DataCenter</i>	84
8.	Consola de administración de VMWare ESXi	86
9.	VMWare vSphere convierte una infraestructura de IT en una Cloud-Computing privada.....	87
10.	Servidor SAN/NAS con varios discos de diferentes capacidades de almacenamiento.....	89

TABLAS

I.	Servicios de procesamiento computacional ofrecidos por Amazon EC2	19
II.	Servicios de desarrollo y ambiente de desarrollo ofrecidos por Amazon EC2	19

III.	Precios de <i>backup</i> de datos a la fecha 04 febrero 2012	66
IV.	Países preferentes para instalar <i>DataCenter</i> . Consultado el 05 marzo de 2012	79

GLOSARIO

AWS	<i>Amazon Web Services</i> , Servicios Web de Amazon.
Backup	Respaldo de información, copia de seguridad.
BIOS	Siglas para <i>Basic Input/Output System</i> , es un sistema que se ejecuta al iniciar la computadora y controla todos los dispositivos de entrada y salida.
Cluster	Conjunto de computadoras con características similares, que juntos funcionan como si fuesen un solo computador.
DataCenter	Centro de procesamiento de datos. Infraestructura tecnológica con la capacidad de procesar y almacenar grandes volúmenes de información.
Failover	Técnica de tolerancia a fallos, que consiste en tener listo un componente en caso de que el principal falle, y éste pase a ser el principal inmediatamente.
IPV4	Versión 4 del Protocolo de <i>Internet</i> (IP). Las direcciones IP son números de 32 bits que deben ser únicas dentro de una misma red y que permitan identificar a cada dispositivo conectado a ella.
IPV6	Sexta versión del <i>Internet Protocol</i> . Está desarrollada para ser la sucesora de la versión 4 y resolver la limitante de número de direcciones disponibles, al ser un número de 128 bits.

ISP	Siglas para <i>Internet Service Provider</i> (Proveedor de servicios de <i>Internet</i>).
IT	<i>Information Technologies</i> . Tecnologías de la información.
Mirroring	Técnica que consiste en hacer una copia espejo. Es decir, replicar el estado de los datos tal y como están en un dispositivo de almacenamiento al momento de cambiar algún dato.
Multicast	Servicio de red, el cual permite enviar datos desde una única fuente, hacia varios destinatarios al mismo tiempo.
NAS	<i>Network Attached Storage</i> . Tecnología de almacenamiento que provee la capacidad de almacenamiento a través de una red dedicada.
On-demand	Servicio que puede ser distribuido como se va requiriendo; es decir, ha pedido.
P2P	<i>Peer to Peer</i> . Red de computadoras donde no hay servidores ni clientes, sino se conectan todos los nodos entre sí.
PDA	Siglas para <i>Personal Digital Assistant</i> (Asistente Digital Personal), computadora de mano.
Rack	Bastidor o gabinete utilizado para alojar equipo informático, sobre todo de redes.
SAN	<i>Storage Area Network</i> . Red de área de almacenamiento, es una red que conecta dispositivos de almacenamiento.

- SCSI** Siglas para *Small Computer System Interface*. Interfaz para la transferencia de datos.
- Streaming** Servicio de distribución de datos en *Internet*, significa 'flujo' y se refiere a la transmisión continua de datos como audio, video u otro.
- Tuning** Optimización para la eficiente operación de un sistema.
- Web** Palabra inglesa para telaraña, regularmente se utiliza cuando se quiere referir al *Internet* como ambiente de ejecución.

RESUMEN

La necesidad constante de respaldar los datos para el uso futuro de ella, es una actividad que realizan tanto empresas pequeñas, medianas y grandes como usuarios finales en sus hogares. El *backup* de datos consiste en respaldar información digital a medios de almacenamiento digitales. Existen diferentes tipos de medios de almacenamiento digitales y se diferencian entre ellos según la tecnología, capacidad, confiabilidad, tiempo de vida y precio.

Respaldar la información tiene un costo asociado el cual no es perceptible para un usuario final, pero si para grandes empresas que manejan grandes cantidades de información digital. Los *backups* de datos actuales que se generan en empresas medianas y grandes tienen la particularidad de utilizar espacios físicos para el almacenamiento de los medios de almacenamiento digitales como cintas magnéticas, CD/DVD-ROM, discos duros, etc. Los *backups* son realizados por operadores de IT periódicamente.

El modelo *pay-as-use* permite a las empresas utilizar recursos de IT y pagar por ellos únicamente el consumo por cierto período de tiempo. Cloud-Computing, es el componente clave que hace posible la implementación del modelo *pay-as-use*. Realizar el *backup* de los datos de una empresa mediana o grande por medio del *Internet* a un servidor de un tercero y pagar una tarifa mensual por GB almacenado es un ejemplo de la implementación del modelo *pay-as-use*. Cloud-Computing adicional es la infraestructura a la cual se accede para consumir recursos por medio del *Internet*. Un centro de datos es un sistema basado en Cloud-Computing, que permite ofrecer el servicio de almacenamiento de datos a los clientes que deseen pagar por el servicio.

Un Banco de Datos Virtual, es un sistema de Cloud-Computing que haciendo uso de su principal recurso, el centro de datos, permite a los clientes depositar información en sus servidores a través del *Internet*. El banco de datos garantiza al cliente por medio del cobro de la tarifa mensual por la GB almacenado en sus servidores, la integridad, seguridad, disponibilidad, y accesibilidad de su información en todo momento.

OBJETIVOS

General

Determinar la factibilidad técnica para las medianas y grandes empresas guatemaltecas de utilizar un Banco de Datos Virtual como servicio de respaldo de su información digital.

Específicos

1. Identificar el tipo de almacenamiento de datos más utilizado actualmente en el medio para respaldos de información digital.
2. Determinar el nivel de aceptación actual hacia el uso de un Banco de Datos Virtual.
3. Demostrar la factibilidad de almacenar información digital en un Banco de Datos Virtual y la reducción de costos que esto representa, fusionando los temas de respaldo y recuperación de datos y prestación de servicio en la nube.
4. Realizar una pequeña demostración de la infraestructura para un Banco de Datos Virtual, y tener una mínima idea del esfuerzo y de la infraestructura tecnológica necesaria para implementar un servicio de *backup online* a nivel empresarial.

INTRODUCCIÓN

El propósito de este trabajo, es el de mostrar los beneficios del Cloud-Computing y aprovecharlos para el respaldo de información digital, y así fusionar ambos conceptos para una posible aplicación empresarial. Específicamente el modelo *pay-as-use*, dicho modelo, es muy utilizado actualmente por medianas y grandes empresas.

La implementación de este modelo de preferencia en el departamento de IT, al cual atañe este trabajo, ayuda a reducir sus costos de operación. Tal y como se utilizan los servicios de electricidad o telefonía, el modelo *pay-as-use*, permite a los departamentos de IT el uso de un modelo que les da la opción de pagar por los servicios que utilizan por el tiempo que lo hacen, es decir, por consumo. Aun cuando los departamentos de IT optan por utilizar medios de respaldo como cintas magnéticas, discos CD-ROM/DVD-ROM, discos duros, etc.

Estos medios únicamente optan por almacenarse en sitios seguros dentro de las empresas o en el mejor de los casos, fuera de sus instalaciones en cajas de seguridad de bancos o empresas que se dedican al almacenamiento de información digital en cajas de seguridad.

Actualmente existen empresas que prestan el servicio de información, algunas de ellas ocupándose únicamente de brindar el servicio de almacenamiento, otros también ofrecen el servicio de seguridad de la información. Sin embargo, ninguna compañía ofrece un servicio equiparativo al modelo de servicio de la banca financiera. La información digital de las empresas es actualmente un recurso de valiosa importancia.

La protección y el respaldo de la misma pueden convertirse en un nuevo modelo de servicio. El presente trabajo mostrará técnicamente cómo hacer uso del modelo de servicio de Cloud-Computing, los beneficios de la virtualización para la creación de un sistema de almacenamiento masivo de datos y el uso del modelo de servicio de la banca financiera, para la creación o adaptación de un nuevo modelo de trabajo para el respaldo de información digital a la cual le llamaremos Banco de Datos Virtual.

También se configurará una pequeña demostración de un Banco de Datos Virtual, utilizando y describiendo herramientas y tecnologías asociadas, para darse una idea del trabajo y la infraestructura que conlleva implementar un Banco de Datos Virtual.

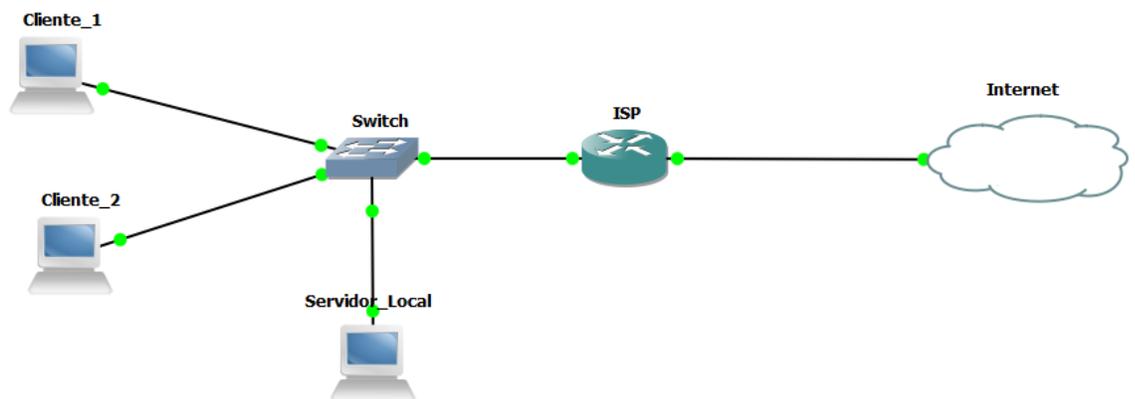
El trabajo va orientado no solo a la fusión de dos conceptos, uno como necesidad y otro como tendencia tecnológica relativamente joven; sino también a la posible utilidad que pueda dársele a nivel empresarial, entiéndase pequeña, mediana y grande empresa tomando en cuenta las necesidades de cada una y sus características y sobre todo al reconocer si existe o no algún antecedente en el ámbito guatemalteco.

1. MARCO TEÓRICO

1.1. Cloud-Computing

Cloud-Computing, es un tema reciente que toma su nombre como metáfora del *Internet*. Esto debido a que típicamente el *Internet* es representado en los diagramas de red como una nube (Velte, 2009).

Figura 1. La nube en el diagrama de red representa al *Internet*



Fuente: elaboración propia.

Prácticamente, todo lo que se encuentra en el *Internet* es lo que puede considerarse parte de la nube, empresas como Amazon, Oracle, Google, Microsoft, IBM, etc., ofrecen plataformas de recursos, aplicaciones o servicios de almacenamiento, los cuales son ofrecidos a las empresas como servicios. Las empresas pueden optar por contratar estos servicios tal como si contratasen algún servicio de electricidad o telefonía.

En esencia, Cloud-Computing es una construcción que le permite a cualquiera acceder a aplicaciones que realmente residen en una locación fuera de su computador o en algún dispositivo conectado a *Internet*, la mayor parte del tiempo, podría estar en un centro de datos muy distante (Velte, 2009).

La evolución de las redes de computadoras, ha permitido en gran medida el auge de Cloud-Computing. El cambio del protocolo de IPv4 a IPv6, permite un mejor uso de los recursos de la red para el envío de paquetes, debido a que soporta *multicast* y hace posible el envío de un paquete a múltiples destinos en una simple operación de envío; esto es ideal para aplicaciones como replicación de datos y *mirroring*, lo cual puede traducirse en servicios en Cloud-Computing como el almacenamiento de datos en centros de datos remotos.

El anuncio de *Google* de probar sus “súper-redes” de alta velocidad (*Google*, 2012), podría terminar de convencer a más empresas de cualquier tamaño el adquirir y confiar más en los servicios disponibles en el *Internet*. Dado que el mayor pro y contra de Cloud-Computing radica en la conexión continua, de alta disponibilidad y alta velocidad entre el cliente y el *Internet*.

Generalmente los departamentos de IT requieren de elevados costos de operación, la mayor parte de los recursos humanos y financieros asignados a los departamentos se utilizan en mantenimiento de infraestructura. Los empleados del departamento de IT invierten mucho de su tiempo productivo en ocuparse de tareas de mantenimiento, soporte y *tuning* a la infraestructura tecnológica de la empresa, principalmente el centro de datos, y las aplicaciones corporativas; no permitiéndoles dedicarle más tiempo a la ejecución de estrategias de proyectos.

El uso de un centro de datos remoto permite ahorrar costos e inclusive reducir personal operativo. Esto se traduce en ahorro de costos de operación en los departamentos de IT de las empresas y en general, ahorro de costos de funcionamiento de las empresas.

1.2. Infraestructura

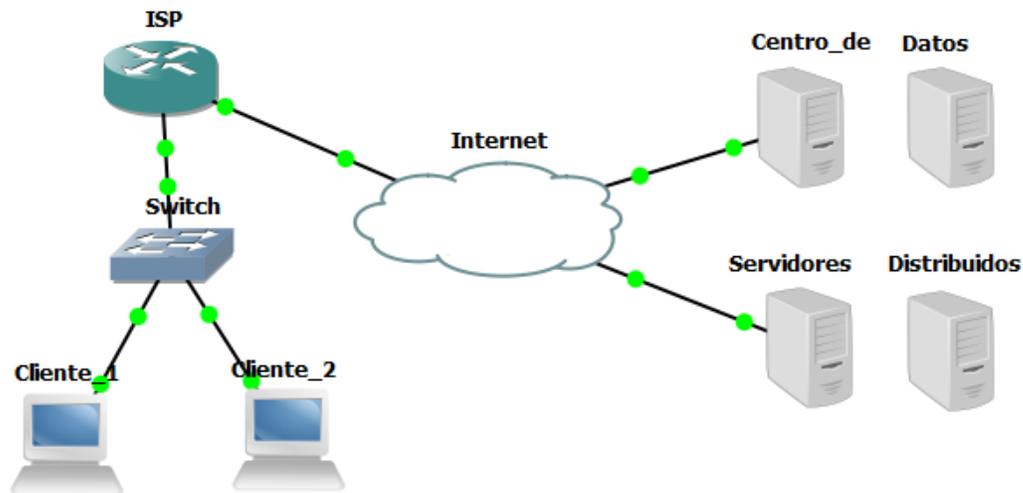
Existen diferentes infraestructuras de Cloud-Computing. El diseño, implementación e integración de la infraestructura, dependerá exclusivamente del tipo de servicio que el proveedor de servicios ofrecerá a sus clientes. Independientemente del tipo de servicio que prestará un proveedor de servicios de Cloud-Computing, siempre contendrá a los componentes de cliente, centro de datos y servidores distribuidos.

1.2.1. Componentes de Cloud-Computing

Los servicios que se ofrecen por medio de Cloud-Computing se ejecutan en la infraestructura del proveedor de servicios. Sin embargo, los elementos que componen la Cloud-Computing son por lo general: los clientes, el centro de datos y los servidores distribuidos. Esto ejemplifica el hecho de que Cloud-Computing no es únicamente en sí el *Internet*, los servicios que adquiere el cliente están en la red, más el cliente forma parte de los componentes de la Cloud-Computing en todo momento.

A continuación se describe a cada componente, pero cabe aclarar que el ISP no es en sí un componente, ya que estos son parte de las redes y permiten a la interconexión entre redes, en el caso de la figura 1.2, ilustra la conexión de dos diferentes clientes a *Internet*.

Figura 2. Componentes que intervienen en Cloud-Computing



Fuente: elaboración propia.

1.2.1.1. Cliente

Clientes son, en una arquitectura de Cloud-Computing, típicamente computadoras que están justamente en su escritorio. Sin embargo, no se limita únicamente a computadoras de escritorio, también pueden ser teléfonos inteligentes móviles, laptops, PDAs, computadoras de tableta.

Clientes son todos los dispositivos que utilizan los usuarios finales para interactuar con las aplicaciones y administrar su información en la “nube”. Se dividen en tres categorías:

- Móviles: dispositivos móviles como *PDA*s o teléfonos inteligentes.
- *Thin*: computadoras que no poseen discos duros internos, dejan que los servidores hagan el trabajo para luego mostrar ellos la información.
- *Thick*: es un tipo de computadora regular, utiliza algún explorador *Web* para conectarse a la Nube.

1.2.1.2. Centro de datos

Centro de Datos, es la colección de servidores donde las aplicaciones a las cuales se suscribe están almacenadas (Velte, 2009). Es la ubicación física en donde se concentran todos los recursos necesarios para el procesamiento de información digital en una organización. Requiere que la ubicación contenga redundancia de energía eléctrica, energía de respaldo (UPS) y reguladores de temperatura (cuarto frío). Suelen ubicarse en los sótanos de los edificios y en cuartos de servidores que se ubican en cualquier región del mundo y se accede a ellos por medio de conexión a *Internet*.

1.2.1.3. Servidores distribuidos

Consiste en la distribución de los servidores en diferentes locaciones físicas y la interconexión entre ellos, permitiendo que los servidores actúen como si estuvieran uno al lado del otro. Esto le da al proveedor del servicio más flexibilidad en opciones y seguridad (Velte, 2009).

La finalidad de los servidores distribuidos, es distribuir la carga de trabajo en diferentes locaciones. La comunicación entre los diferentes servidores se logra por medio de la infraestructura de telecomunicaciones distribuida;

diferentes aplicaciones son instaladas en los diferentes servidores permitiendo así la distribución de trabajo. Los usuarios acceden a cada recurso de software y/o *hardware* instalado en los diferentes servidores y lo perciben como si fuera un solo sistema.

Otra de las características de los servidores distribuidos es que tienen tolerancia a fallos, lo cual significa que, cada servidor tiene configurado un reemplazo el cual se activa en cuanto uno falla.

1.3. Virtualización

La virtualización, es el concepto esencial que permite crear servicios de Cloud-Computing. Prácticamente se puede afirmar que sin virtualización no será posible crear Cloud-Computing.

1.3.1. Concepto

La virtualización, es un término amplio que se refiere a la abstracción de los recursos de un computador. Este término se viene usando desde antes de 1960, y ha sido aplicado a diferentes aspectos y ámbitos de la informática. Sin embargo, el reciente desarrollo de nuevas plataformas y tecnologías de virtualización han hecho que se vuelva a prestar atención a este maduro concepto (López-Vallejo, 2008).

La virtualización permite la creación de recursos de *hardware*, software y datos de tal manera que quienes utilizan dichos recursos pueden percibirlos como si fuesen reales y no instancias de software. Recursos de *hardware* como dispositivos de red, memoria RAM y dispositivos de almacenamiento son comúnmente virtualizados para su utilización por el tipo de virtualización

actualmente muy comunes, el cual es la máquina virtual (VM por sus siglas en inglés). Las máquinas virtuales son la implementación de computadoras que ejecutan toda clase de programas tal y como lo hacen las computadoras reales.

Esto ha permitido que diferentes computadoras con diferente sistema operativo instalado en ellos pueda ejecutarse en una misma computadora física, cada computador virtualizado puede tener diferentes configuraciones de *hardware* como memoria, cantidad de CPUs –procesadores-, capacidad de almacenamiento, tarjetas de red, etc., la virtualización permite la reducción de costos en infraestructura de TI.

1.3.2. Paravirtualización

La paravirtualización permite a múltiples sistemas operativos ejecutarse en una computadora al mismo tiempo. Para esto utiliza recursos del sistema, como el procesador y la memoria volátil (Velte, 2009).

La paravirtualización no emula todos los recursos de *hardware* que utiliza una simple computadora (BIOS, drives, memoria, procesador, etc.). Esto ayuda a que sea un tipo de virtualización más eficiente con un mejor uso de recursos del sistema físico del cual únicamente utiliza el procesador, la memoria RAM y el disco duro.

Es decir, cuando se crea una máquina virtual con cierta cantidad de procesadores, memoria RAM y espacio de almacenamiento, la máquina virtual realmente utiliza estos recursos de la computadora física en donde se ejecuta, esto también indica que el limitante en cuanto a procesadores y memoria RAM a virtualizar estará dado por los recursos disponibles en el computador físico.

1.3.3. Cloud-Computing y virtualización

Los servidores distribuidos ejecutan diferentes servicios y el cliente no conoce el tipo de infraestructura del proveedor del servicio.

Los centros de datos de la infraestructura del proveedor por lo general virtualizan computadoras en las cuales instalan diferentes servicios. La virtualización en un centro de datos permite el ahorro de costos en *hardware*, esto también ayuda al respaldo de sistemas de *hardware* completos. Cuando un equipo falla, otro lo reemplaza, ambos poseen las mismas configuraciones técnicas e instaladas las mismas aplicaciones, facilitando de esta manera la recuperación en caso de fallas y el ahorro de costos en equipo.

En Cloud-Computing, la virtualización es una de las características presentes tanto en los clientes como en los proveedores de servicios, ambos lo utilizan para el ahorro de costos, sin embargo, el propósito de su implementación depende de las necesidades específicas de cada uno de ellos.

1.4. Pay-as-use

El software ha evolucionado continuamente desde que se logró codificar el primer algoritmo computacional, años atrás cuando pagar por la programación y codificación de un programa computacional era prácticamente un lujo para las empresas o entidades que lo necesitaran. Los costos de desarrollar un programa computacional han bajado considerablemente si se compara la época en la cual la computación apenas iniciaba y la época actual. Actualmente se tienen definidas metodologías, estándares, y prácticas de desarrollo, las cuales ayudan en cierta medida a bajar el costo final de las aplicaciones.

El pago por el desarrollo de una aplicación no representa un problema para las necesidades de la empresa, ya que si es un software a la medida, el costo se negocia con la empresa desarrolladora de la aplicación; si es una aplicación la cual puede conseguirse en el mercado, se paga el precio ya fijado por éste en el mercado.

Para el caso en el cual se paga por una aplicación en el mercado, en empresas grandes, la instalación, actualización o mantenimiento de la misma es un proceso que consume muchos recursos del departamento de IT, como lo es el recurso humano y el tiempo que invierten los mismos. Esto último, siempre representa atrasos en la producción del personal de los otros departamentos de la empresa ajenos a IT. Estos se ven afectados en la ejecución de sus actividades y tareas acordes a sus funciones dentro de la empresa, cuando el departamento de IT realiza actividades de mantenimiento en las aplicaciones que estos utilizan.

El mantenimiento al equipo de cómputo también resulta en atrasos de ejecución de tareas y perdidas como resultado de los atrasos. De aquí, que actualmente las aplicaciones corporativas desarrolladas en ambientes *Web* ha venido a reducir significativamente los problemas de actualización, instalación y mantenimiento que se traducen en pérdidas y atrasos al personal que hace uso de dichas aplicaciones.

El principal problema con este tipo de aplicaciones se presenta cuando la comunicación entre el cliente y la aplicación no puede ser establecida. La mayor parte del tiempo sucede por problemas en la infraestructura de red ya sea del cliente o del proveedor del servicio de la aplicación.

Sin embargo, actualmente este riesgo se gestiona mediante la contratación de diferentes proveedores de conexión a *Internet*, con lo cual se trata de reducir el tiempo de desconexión entre el proveedor de la aplicación y los usuarios de ésta.

El modelo *pay-as-use*, define que únicamente debe pagarse por un servicio, lo que de él se utilice. Los servicios tienen la particularidad que se utilizan por un período de tiempo y solamente se paga el costo del servicio utilizado durante el tiempo que se consumió. Se puede citar como ejemplo el servicio de telefonía disponible en la mayoría de hogares, en donde el minuto o segundo tiene un precio determinado.

Durante un período de tiempo, generalmente un mes, se contabilizan la cantidad de minutos consumidos, diferenciando las llamadas recibidas de las llamadas salientes realizadas ya que estas tienen un precio diferente. La suma de estos costos es enviado a las residencias por medio de una factura, la cual debe cancelarse en un tiempo prudencial a fin de evitar la cancelación del préstamo del servicio.

El modelo *pay-as-use*, prácticamente utiliza las características del modelo de pago del servicio telefónico mencionado anteriormente y las adapta para la entrega de artefactos por medio de contratos de servicio, todo esto relacionado a IT. Realmente es una adaptación innovadora y creativa, ya que pagar por el uso del software y/o *hardware* ayuda a reducir increíblemente los costos de operación de las empresas. El mayor problema que esto genera es la dependencia que en algún momento puede generarse entre el cliente y el proveedor del servicio.

Cloud-Computing, es actualmente el escenario natural donde el modelo *pay-as-use* se ha implementado, aun cuando los beneficios esperados de este modelo no han cumplido en su totalidad las expectativas de los clientes. Sin embargo, este modelo ha sido utilizado parcialmente y adaptado exitosamente para la entrega de artefactos a empresas que ven viable la no inversión en una infraestructura de *hardware* y/o software para su centro de datos físicos por los costos que esto representa, en lugar de ello, optan por suscribirse a algún proveedor que les presta el servicio de infraestructura de *hardware* y/o software.

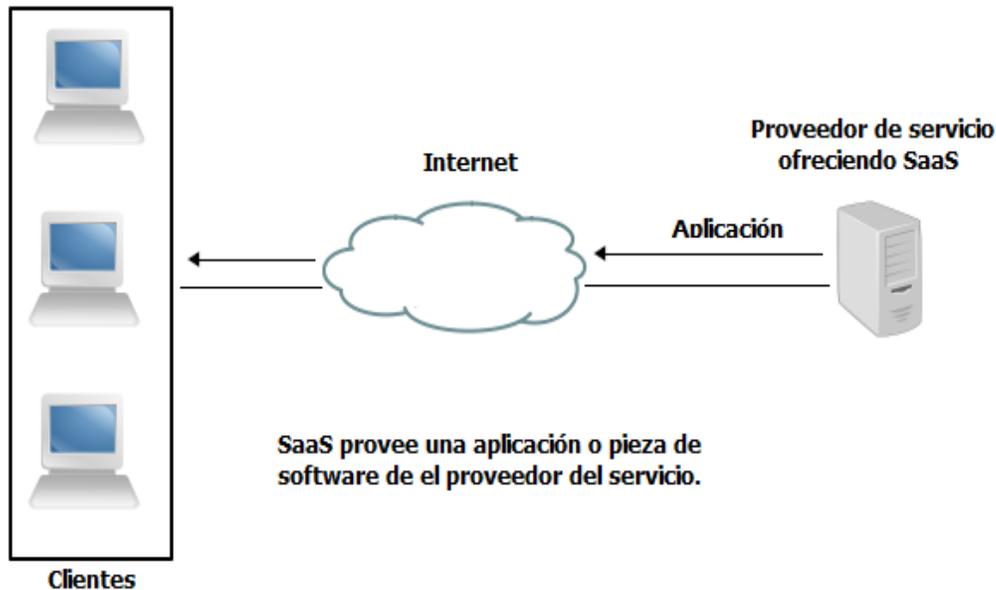
A continuación se describen los tipos de servicios que actualmente pueden adquirirse en Cloud-Computing mediante una suscripción con algún proveedor del servicio. Para cada tipo de servicio, no siempre un proveedor supe todos los diferentes servicios, es normal que una empresa pueda tener diferentes proveedores de servicio según el servicio que en su momento adquiera.

1.4.1. Software como Servicio (Software as a Service - SaaS)

SaaS, es un modelo en el cual una aplicación es alojada como un servicio para los clientes que acceden a él vía *Internet*. Cuando el software es alojado fuera del sitio, el cliente no tiene que mantener o darle soporte alguno. Por otra parte, queda fuera del alcance del cliente cuando el servicio de alojamiento decide hacer cambios en el software. La idea es que los clientes usen el software de la caja como es y no requieran hacerle muchos cambios o requerir la integración a otros sistemas.

El proveedor realiza todos los parches y actualizaciones que el software necesita, así como mantener la infraestructura corriendo (Velte, Velte & Elsenpeter, 2009: 11,12 y 13).

Figura 3. **Esquema de SaaS**

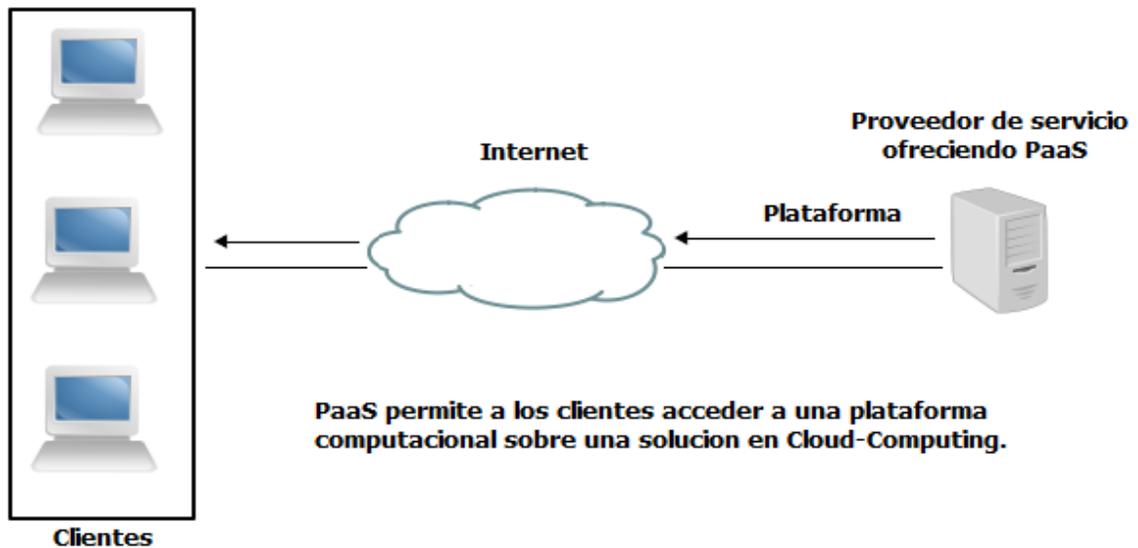


Fuente: Velte, Velte & Elsenpeter, 2009: 12.

1.4.2. **Plataforma como servicio (Platform as a Service - PaaS)**

PaaS provee todos los recursos necesarios para construir aplicaciones y servicios completos desde el *Internet*, sin tener que descargar o instalar software. Los servicios incluyen diseño de aplicación, desarrollo, pruebas, despliegue, y recepción. Otros servicios incluyen la colaboración de equipo, la integración de servicio de *Web*, la integración de base de datos, la seguridad, la escalabilidad, el almacenaje, declarar la dirección, y versionamiento (Velte, Velte & Elsenpeter, 2009: 13,14 y 15).

Figura 4. Esquema de PaaS



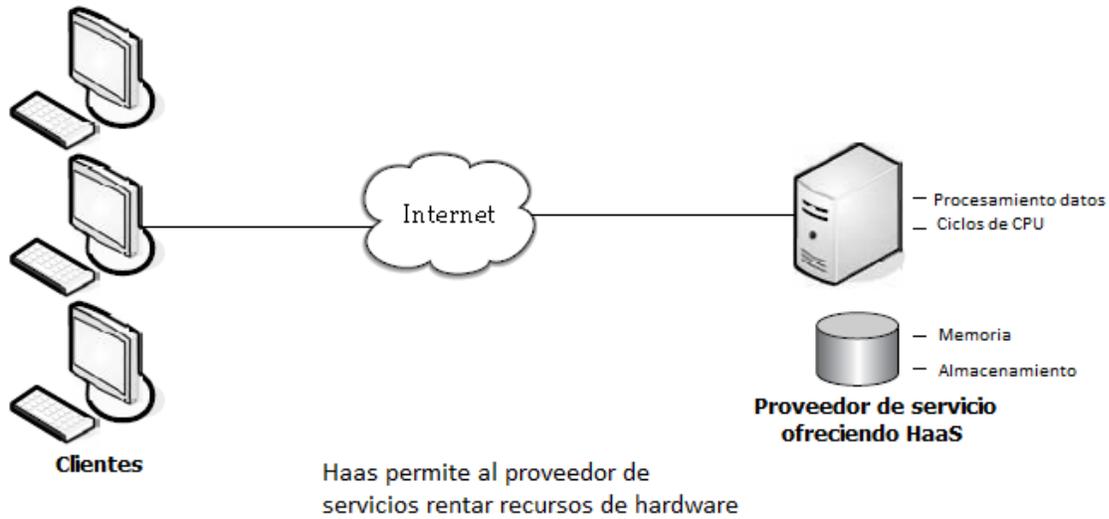
Fuente: Velte, Velte & Elsenpeter, 2009: 14.

1.4.3. Hardware como Servicio (*Hardware as a Service - HaaS*)

HaaS, es la nueva forma de servicio disponible en Cloud-Computing. SaaS y PaaS proveen aplicaciones a los clientes, HaaS no lo hace. Este modelo simplemente ofrece el *hardware* de modo que su organización puedan poner lo que ellos lo que desee. En lugar de comprar servidores, software, *racks*, y tener que pagar por el centro de datos y el espacio para ellos, el proveedor del servicio rentas esos recursos. Se pueden rentar recursos como:

- Espacio de servidor
- Equipo de red
- Memoria
- Ciclos de CPU
- Espacio de almacenamiento

Figura 5. **Esquema de HaaS**



Fuente: Velte, Velte & Elsenpeter, 2009: 15.

Además, la infraestructura puede escalar dinámicamente aumentada o disminuida, basada en las necesidades de recurso de la aplicación (Velte, Velte & Elsenpeter, 2009: 15,16).

2. ANTECEDENTES

2.1. Implementación de Cloud-Computing en las empresas

La computación en la nube o Cloud-Computing, es un modelo, pero sin embargo de rápida difusión en el ámbito empresarial. Este modelo que a grandes rasgos representa la adquisición de servicios especializados referentes a la computación, en lugar de que las empresas incurran en gastos de gestión de sistemas e infraestructura de cómputo para el cumplimiento de sus objetivos como empresa. Este modelo, a su vez, tiene aspectos positivos como negativos y habrá que hallar un balance para su posible implementación.

Grandes empresas a nivel mundial como Amazon o Google cuentan ya con una infraestructura para prestar servicios de computación en la nube. Cada una tiene su propio paquete de servicios para ofrecer a las empresas.

2.1.1. Amazon EC2

Amazon EC2 permite definir y configurar todos los aspectos de sus instancias, desde el sistema operativo hasta las aplicaciones. Una AMI (*Amazon Machine Image*) no es más que un entorno agrupado que incluye toda la información necesaria para configurar e iniciar su instancia. Sus AMI son su unidad de despliegue. Puede tener sólo una AMI o componer un sistema a partir de varias AMI de bloques de creación (por ejemplo, servidores *web*, servidores de aplicaciones y bases de datos). Amazon EC2 proporciona varias herramientas para hacer que la creación de una AMI sea más fácil, como AWS Management Console (Amazon.com Company, 2012).

Amazon EC2 cuenta con una infraestructura para el manejo de computación en la nube para desarrolladores, contando con un conjunto bastante considerable de herramientas tanto empresariales como para desarrolladores. Todo este conjunto de servicios.

Amazon EC2 presenta un auténtico entorno informático virtual, que permite utilizar interfaces de servicio *web* para iniciar instancias con distintos sistemas operativos, cargarlas con su entorno de aplicaciones personalizadas, gestionar sus permisos de acceso a la red y ejecutar su imagen utilizando los sistemas que desee. Para utilizar Amazon EC2, sólo tiene que: (Amazon.com Company, 2012)

- Seleccionar una imagen de plantilla preconfigurada para pasar a estar activo de inmediato. O bien crear una AMI (*Amazon Machine Image*) que contenga sus aplicaciones, bibliotecas, datos y valores de configuración asociados.
- Configurar la seguridad y el acceso a red en su instancia de Amazon EC2.
- Seleccionar los tipos de instancias y sistemas operativos que desee y, a continuación, iniciar, finalizar y supervisar tantas instancias de su AMI como sea necesario, a través de las API de servicio *web* o la variedad de herramientas de gestión proporcionadas.
- Determinar si desea una ejecución en varias localizaciones, utilizar puntos finales de IP estáticos o adjuntar almacenamiento de bloques continuo a sus instancias.

- Pagar sólo por los recursos que realmente consuma, como las horas de uso de instancias o la transferencia de datos.

Amazon presenta sus servicios de computación en la nube con las siguientes características:

- Elástico: permitiendo aumentar o reducir la capacidad de cada instancia, que es como llaman a una especie de máquina virtual, de inmediato y sin tener que incurrir en costos de infraestructura.
- Control total: teniendo un control total a todas las instancias a la disposición del cliente, a través de un usuario raíz cuyo acceso será como el de computadoras reales.
- Flexible: al permitir tener acceso a distintas instancias con distintas características, como capacidad, sistema operativo, almacenamiento, entre otros.
- Diseño pensado para el uso con otros Amazon *Web Services*: integrando servicios de almacenamiento, computación y consultas de datos.
- Seguro: con interfaces *Web* de acceso y mecanismos de seguridad para el control de acceso a la red.
- Económico: al poder disfrutar las ventajas de contratar *hardware* y software como servicio. (Amazon.com Company, 2012)

Todas estas características cumplen con la definición de computación en la nube y sus respectivos beneficios. (Velve, 2010).

También se ofrece distintos sistemas operativos para utilizarlos en sus respectivas instancias. Según el sitio de Amazon EC2 los sistemas operativos ofrecidos son (Amazon.com Company, 2012):

- Red Hat Enterprise Linux
- Open Solaris
- Fedora
- Windows Server 2003
- Windows Server 2008
- Amazon Linux AMI
- Gentoo Linux
- SUSE Linux Enterprise
- Oracle Enterprise Linux
- Ubuntu Linux
- Debian

También se ofrece servicios de gestores de bases de datos, procesamiento por lotes y alojamiento *Web*, mostrados en la siguiente tabla: (Amazon.com Company, 2012).

Tabla I. **Servicios de procesamiento computacional ofrecidos por Amazon EC2**

Bases de datos	Procesamiento por lotes	Alojamiento web
IBM DB2	Hadoop	Apache HTTP
IBM InformixDynamic Server	Condor	IIS/Asp.Net
Microsoft SQL Server Standard 2005/2008	Open MPI	IBM Lotus Web Content Management
MySQL Enterprise		IBM WebSphere Portal Server
Oracle Database 11g		

Fuente: elaboración propia.

A su vez se ofrecen entornos de desarrollo, servidores de aplicaciones y codificación y video streaming: (Amazon.com Company, 2012)

Tabla II. **Servicios de desarrollo y ambiente de desarrollo ofrecidos por Amazon EC2**

Entornos de desarrollo de aplicaciones	Servidores de aplicaciones	Codificación de vídeo y Streaming
IBM sMash	IBM WebSphereApplication Server	Wowza Media Server Pro
JBoss Enterprise ApplicationPlatform	Java Application Server	Windows Media Server
Ruby onRails	Oracle WebLogic Server	

Fuente: elaboración propia.

Referente a los precios, existe un servicio gratuito que ofrece Amazon, para los nuevos clientes de AWS pueden empezar de forma gratuita con Amazon EC2. Al registrarse, los nuevos clientes de AWS recibirán cada mes y durante un año los siguientes servicios de EC2:

- 750 horas de uso para la ejecución de instancias Linux/Unix Micro en EC2.
- 750 horas de Elastic Load Balancing más 15 GB de procesamiento de datos.
- 10 GB de Amazon Elastic Block Storage (EBS) más 1 millón de E/S, 1 GB de almacenamiento de fotos, 10.000 solicitudes para Obtener fotos y 1.000 solicitudes para Poner fotos.
- 15 GB de banda ancha entrante y 15 GB de banda ancha saliente sumadas por todos los servicios de AWS.

Los precios varían según las necesidades del cliente y puede ser completamente personalizable. Es así como puede pagarse por horas de uso de instancias, una cuota mensual o anual a largo plazo. Existen también instancias reservadas desde las cuales puede trabajarse con la tranquilidad de que pertenece al cliente y nadie más la podrá usar.

Asimismo, se puede optar por un plan más sofisticado en el cual los precios pueden fluctuar según la oferta y la demanda. Para cada una de las instancias existen cuatro regiones de las cuales uno puede escoger:

- Virginia del Norte, EEUU
- California del Norte, EEUU
- Irlanda, Unión europea
- Singapur

Cada una de estas locaciones tiene un precio distinto y se puede escoger la que se desee, con la tendencia a crecimiento y posibles nuevas locaciones. Así, por ejemplo, si se desea tener una instancia en Virginia del Norte, con una petición de alta capacidad de memoria del tipo extra grande, utilizando Windows tendrá un precio de \$0.62 por hora, esto por un servicio *On-demand*, que es un pago por utilización de equipo por hora, sin compromisos de largo plazo. Si se quiere una instancia con las mismas características, pero con locación en Singapur, el costo será de \$0.69 por hora.

Respecto a las instancias reservadas, puede por ejemplo pedirse una instancia de un CPU de alta potencia extragrande, con un plazo de tres años a un costo de \$ 2800,00 con un costo de uso de Linux a \$0,24 por hora.

Existe una extensa gama de configuraciones personalizables en donde puede escogerse una especie de máquina virtual, con la ejecución y almacenamiento en *Internet*. Puede ir desde una configuración estándar pequeña, hasta requerimientos para implementar un *clúster* en la nube.

Toda esta gama de configuraciones puede ser consultada en el enlace <http://aws.amazon.com/es/ec2/#pricing> (Amazon.com Company, 2012). También puede hacerse los cálculos sobre el coste que tendrá una configuración determinada por mes el cual se puede calcular en el siguiente enlace <http://aws.amazon.com/es/calculator/> (Amazon.com Company, 2012).

2.1.2. **Google Apps**

Google, por su lado, se centra más en software como servicios. Dichos servicios están segmentados en tres áreas generales:

- Pequeñas empresas
- Medianas empresas
- Grandes empresas

A diferencia de Amazon, *Google* no se orienta a la prestación de *hardware* como servicio (*HaaS*), sino software como servicio (*SaaS*) ofreciendo servicios de correo electrónico, calendario interactivo, documentos, entre otros. Este conjunto de software se orienta a la ofimática y puede ser contratado desde instituciones gubernamentales hasta educativas.

La idea principal de *Google Apps*, es la reducción de costos de TI en las organizaciones, lo que puede abarcar desde adquisición de equipo de cómputo, hasta pago a personal, reduciendo considerablemente los costos.

Hasta el momento de presentación de este trabajo, *Google Apps* contaba con las siguientes aplicaciones:

- *Gmail*: correo electrónico.
- *Google Calendar*: calendario.
- *Googledocs*: aplicaciones de oficina en la nube como procesador de texto, presentaciones, hojas de cálculo.
- Grupos de *Google*: gestión de grupos para publicación de contenidos.
- *Googlesites*: personalización de páginas *Web*.
- *Google videos*: videos en línea.

- *Ad Sense*: permite publicación de anuncios en las páginas *Web*.
- *AdWords*: publica anuncios de la empresa en otras páginas.
- *GoogleMaps*: consulta de mapas y publicación de indicaciones para llegar a sitios específicos.
- *Merchant Centre*: aplicación que permite publicar datos de productos de la empresa.
- *Alertas de Google*: envío de actualizaciones de temas específicos a correo electrónico.
- *Analytics*: aplicación que analiza el tráfico del sitio *Web* de la empresa.
- *Moderator*: aplicación sobre preguntas u opiniones de un grupo de usuarios.
- *Google noticias*: portal de noticias.
- *Blogger*: Gestor de blogs para publicación en *internet*.
- *Orkut*: red social análoga a Facebook.
- *Checkout*: herramienta de compra-venta por *internet*.
- *Álbumes web de Picasa*: gestión de fotografías en línea.
- *Places*: Inscripción de la empresa en Google para su publicación en mapas.
- *DoubleClickforAdvertisers*: administración, ejecución, publicación y revisión de publicidad online.
- *DoubleClickforPublishers*: administración de inventario de anuncios.
- *GoogleReader*: permite el acceso rápido a los blogs y *feeds* de noticias.
- *GoogleSketchUp*: creación de modelos en 3D.
- *GoogleEarth*: exploración virtual del mundo.
- *TranslatorToolkit*: herramientas para traductores.
- *FeedBurner*: analizador de suscripciones a los feeds de RSS.
- *GoogleFinance*: información, noticias y gráficos sobre empresas.
- *GoogleVoice*: administración de llamadas y buzón de voz. Disponible únicamente en Estados Unidos al momento de esta publicación.

- *iGoogle*: página personalizada de Google, permitiendo añadir noticias, juegos, entre otros.
- *GoogleWave*: elaboración de contenidos y debates.
- *Knol*: red de compartición de conocimiento.
- *Optimizador de sitios Web*: creación de sitios *web* eficaces.
- *YouTube*: permite compartir videos a nivel mundial. (Google, 2012).

Varias empresas de varios tamaños alrededor del mundo han optado por el Cloud-Computing y han escogido a *Google Apps* como su plataforma principal para la ejecución de sus procesos y administración de sus departamentos, reduciendo considerablemente sus costos de IT. Este servicio tiene un costo fijo de US\$ 50 anuales (Google, *Googleapps*, 2012), por cada cuenta de usuario que se inscriba. Es decir, depende del tamaño de la empresa.

En el sitio principal de *Google Apps* hay una lista considerable de todas las empresas que ya lo han puesto en marcha. Entre ellas se encuentra IRB Barcelona, GRUP SERHS, *Jazztel*, *Jaguar/LandRover*, EFE. Para obtener una lista completa de las organizaciones que lo han contratado, puede consultar el sitio <http://www.google.com/apps/intl/es/business/customers.html> (Google, 2012). Hasta el momento no se reporta ninguna empresa guatemalteca que haya implementado *Google Apps*.

Presentados los dos casos de las dos de empresas que prestan servicios de Cloud-Computing, Google y Amazon, se visualiza que Amazon presta un servicio más configurable, mientras que *Google* un servicio más orientado al usuario personal, grupal y a ofimática. Habrá que hacer un estudio introspectivo empresarial para determinar las necesidades principales de la propia organización y determinar qué tipo de servicio es el que se necesita, así como

el presupuesto para ponerlo en marcha y cambiar el paradigma en los empleados de cómo se ejecutarán los procesos y hacia dónde se computará y almacenará la información. Un cambio difícil, pero técnicamente realizable y justificable.

En este trabajo únicamente se presentaron dos servicios que ya están en cierto modo desarrollados y cuentan con infraestructura adecuada para implementarlos y que ejemplifican dos sub-modelos distintos del modelo que es Cloud-Computing y prestación de servicios. Existen otros servicios más pequeños pero a su vez más especializados que forman parte de esta nueva tendencia tecnológica y empresarial, en el presente trabajo únicamente se presentan dos servicios que ya están en cierto modo desarrollados y cuentan con infraestructura adecuada para implementarlos y que presentan dos sub-modelos distintos del modelo que es Cloud-Computing y prestación de servicios.

2.1.3. Implementación de Cloud-Computing

Para una organización, la implementación de Cloud-Computing puede ser una posibilidad rentable. Esto es posible debido a que este modelo cambia completamente el paradigma tradicional del departamento de IT de una empresa. Entra en juego un tercero que presta el servicio, a pedido, que le permite reducir costos innecesarios a la empresa.

Se puede ver, de nuevo, la analogía del Cloud-Computing con servicios como el de la energía eléctrica. El proveedor de energía eléctrica – en el caso del territorio guatemalteco – cobra una cuota mensual por prestación de servicio, que depende de la cantidad de energía consumida durante ese período de tiempo.

El proveedor de energía eléctrica tiene sus propias plantas o compra la energía a empresas transnacionales y tiene una infraestructura con la capacidad para distribuir la energía a una cantidad considerable de comercios, domicilios, edificios, instituciones, entre otros. Esto evita que cada persona tenga la necesidad de crear su propia energía eléctrica que, al ser generada con los métodos tradicionales y uso de combustibles fósiles, ocasionaría un gasto bastante grande y quizás innecesario.

Una de las ideas principales del Cloud-Computing tiene que ver con lo antes establecido. En las empresas se incurren en gastos grandes el hecho de tener una infraestructura capaz de manejar grandes cantidades de información y soporte a muchos usuarios, una red de alta disponibilidad, servidores que controlen todo movimiento de la empresa e incluso distribuidos en distintos lugares geográficos ya sea a nivel nacional o internacional, lo que conlleva la contratación de bastante personal de IT, entre programadores, administradores de red y base de datos, soporte técnico, mantenimiento, etc.

Por lo tanto entra en juego el Cloud-Computing, que ya posee una infraestructura para soportar una gran cantidad de usuarios, las aplicaciones ya están programadas o se presentan máquinas virtuales personalizadas listas para ser utilizadas lo que reduce grandemente los tiempos de implementación, contratación de personal y la adquisición constante de tecnología y servicios especializados. Todo se contrata como un servicio y ha pedido o a medida de uso, cuanto y cuando se requiera, esto alivia la carga de tener que gestionar los propios sistemas de información y tecnologías; salvo un elemento clave: un acceso permanente, eficaz y altamente disponible a *Internet*, que es la columna vertebral del tema de computación en la nube.

2.1.4. Implicaciones

Para la correcta implementación de un servicio de computación en la nube, debe aceptarse ciertas cuestiones. En primer lugar, toda la ejecución y almacenamiento de datos se encuentra en *DataCenters* que se encuentran fuera del edificio donde se encuentre ubicada la empresa y, en segundo lugar todo el contenido será prestado como un servicio a los clientes que accederán a él a través de un navegador *Web*, pagando por cada servicio a pedido sin necesidad de adquirir necesariamente *hardware* o *software*. (Coca, 2012).

2.1.5. Ventajas y desventajas del modelo de Cloud-Computing

Como todo servicio e implementación, la computación en la nube tiene tanto beneficios como desventajas, habrá que sacar un balance para la propia organización, y visualizar operaciones a largo plazo para su correcta ejecución y reducción de errores y vulnerabilidades. Habrá que recordar que lo que se busca siempre es la reducción en los costos y la optimización de recursos, por lo que no sólo se trata de seguir una moda o estar a la vanguardia tecnológica, sino el poder realizar un trabajo eficiente y eficaz buscando la efectividad.

Expuesto lo anterior, se puede evaluar las ventajas y desventajas que puede representar para una empresa contemporánea el uso de servicios en la nube, independientemente de *hardware* o *software*.

Primero las ventajas. Como se ha mencionado anteriormente, la reducción de costos resulta el valor más atractivo de Cloud-Computing pues permite prescindir de gastos en infraestructura de IT y en desarrollo de aplicaciones propias.

También está la cuestión del acceso remoto desde prácticamente cualquier lugar y cualquier dispositivo con acceso a *Internet*, ampliando la flexibilidad y la comodidad. Pero quizás lo más importante – en lo que a costos se refiere – es la contratación de servicios según las necesidades en el instante, pues no se gasta de más ni se subestima el crecimiento paulatino que pueda tener la empresa eventualmente, pues puede contratarse un ítem más en cualquier momento y estará listo de inmediato sin tener que esperar hasta meses en la implementación de tecnología, otra ventaja sobre todo en organizaciones que utilizan el sistema de licitaciones.

Ahora las desventajas. Como resulta obvio, se tendrá pura y entera dependencia hacia la o las empresas que estén prestando el servicio y cualquier problema que ellos tengan será, por ende, el problema del contratante. Es por ello que para implementarse este tipo de servicios se tendrá que contar con una infraestructura adecuada y que pueda estar disponible al menos el 99,9 por ciento del tiempo para su efectividad (Wikipedia, 2012).

Algo similar sucederá con la empresa que provea el servicio de *internet*, pues se dependerá al 99,999 por ciento de este servicio para realizar el trabajo (Wikipedia, 2012). Otra desventaja es que la información que quizás sea considerada confidencial o delicada será manejada por terceros, algo que no agrada a muchos gerentes; es por ello que habrá que implementar medidas de seguridad y políticas realmente claras que comprometan tanto al cliente como al que presta el servicio y se asegure también el tráfico de información a través de la red, con medidas como encriptación u otras similares.

Pero sin duda lo que se necesita es un desarrollo de esta tendencia en el país, para que alcance un nivel de madurez necesario y que se ejecuten las aplicaciones desde *Internet*. Hasta la fecha no hay muchos antecedentes desde

los cuales se pueda sacar conclusiones concretas y menos en el ámbito guatemalteco –donde de hecho no hay ninguno-, habrá que esperar que las empresas se aventuren a invertir en estos servicios y cambien su paradigma sobre las tecnologías de información y comunicación.

2.2. Servicios orientados a Cloud-Computing, cambios y oportunidades

Aunque empresarialmente hablando hay muy pocas implementaciones de servicios en la nube; es decir, no se ha implementado en las empresas sus procesos como servicios de Cloud-Computing; algunos servicios ya llevan algún tiempo en *Internet*.

2.2.1. Video

Ver videos en línea, es algo que ya lleva algún tiempo. Este tipo de red social, permite compartir videos alrededor de todo el mundo, haciendo públicos los videos permite la comunicación y difusión entre todo tipo de segmento de público. Sitios como YouTube, Megavideo, *Google* Video, entre otros, prestan este tipo de servicio.

2.2.1.1. Video streaming

Esta es una variante de servicio de video. Lo que hace un servicio de este tipo es ofrecer la posibilidad de una transmisión en vivo de un acontecimiento completamente en vivo, a través de un canal. El usuario puede transmitir conferencias, encuentros deportivos o políticos, etc. desde su propio espacio *Web*.

2.2.2. Hospedaje de archivos

El almacenamiento en la nube no es nada nuevo, sitios como Rapidshare o Megaupload fueron los primeros en prestar servicio de hospedaje de archivos. Actualmente existen muchos más sitios similares, en el cual hay servicio gratuito pero bastante limitado, con opción a creación de cuentas donde se dan beneficios extras como capacidad considerable y descarga ilimitada.

2.2.3. Música

Existen sitios donde puede compartirse música, al estilo de las aplicaciones P2P donde se puede compartir música y descargarla. La variante de estos sitios es que los usuarios pueden subir sus archivos de audio y compartirlos en cualquier lado, de este modo si uno desea escuchar un archivo de audio basta con conectarse a ese sitio y reproducirlo desde allí, esto sin ocupar espacio local o tener algún reproductor.

2.2.4. Imágenes

Pueden compartirse imágenes en *Internet*. Sitios como Picasa o Flickr proporcionan un espacio al usuario para subir sus fotografías y compartirlas en la red. Algunos sitios lo toman como una red social, en la cual puede hacerse pública la colección de fotografías de un usuario para que otros alrededor del mundo puedan hacer uso de ellas; otras únicamente se limitan a almacenarlas y mostrarlas.

2.2.5. Pagos y transferencias

También puede realizarse pagos por medio de *Internet*, no con tarjetas de crédito como tradicionalmente se había hecho, sino con sitios como PayPal

que permiten hacer compras en cualquier parte del mundo de una forma más segura. Este sistema permite brindar los datos de la tarjeta de crédito de un usuario a un sitio de confianza, y dejar que realice la transferencia de una forma seria y segura.

2.2.6. Anuncios publicitarios

Los anuncios publicitarios en la *Web* pueden reducir costos para las empresas, comparado con los precios en otros medios de comunicación como la televisión o el radio. Este modelo de negocio permitió el crecimiento y el éxito de *Google* como servicio y lo impulsó como una empresa millonaria y conocida a nivel mundial.

2.2.7. Sistemas operativos

Actualmente existen, sistemas operativos que pueden correr en la nube. Esto permite reducción de costos en licencias considerablemente caras y evita molestias al usuario como actualizaciones engorrosas y mantenimiento constante, además de que los recursos que se consumirán serán los que se encuentren en el centro de datos que presta el servicio, no de la computadora local o dispositivo con acceso a *Internet*.

2.2.8. Virtualización

También puede contratarse una máquina virtual como servicio. Como se ha visto anteriormente, Amazon EC2 puede brindar lo que ellos llaman una instancia, que no es más que una máquina virtual en la nube, que puede tener características personalizadas y funcionar desde el centro de datos y acceder a ella por medio de un navegador.

2.2.9. Ofimática

Servicios básicos para las tareas productivas en los negocios también son servicios que ya se prestan en la nube. El ejemplo más conocido son los *Google Docs*, que son un paquete para realizar tareas básicas como documentos, presentaciones, hojas de cálculo, formularios y diagramas. Esto también reduce costos en la adquisición de licencias o en uso de recursos para su ejecución.

Los anteriores son únicamente algunos ejemplos de servicios que ya llevan un tiempo funcionando en la nube y que han sido exitosos en su campo. Este modelo está en continua expansión y su éxito provendrá del crecimiento de la oferta y la demanda, en cuanto sea gradualmente aceptado como nuevo paradigma.

Actualmente hay servicios en la nube para muchísimas cosas, incluso para poner en marcha un call center. También hay sitios especializados en sus áreas, lo que permite tener una mayor confianza en la prestación de servicios; tal es el caso de *salesForce.com* que es un sitio que se especializa en la prestación de servicios en la nube para empresas. Sus servicios incluyen herramientas para marketing, ventas, servicio al cliente, colaboración empresarial (*SalesForce.com*, 2012).

2.3. Cambios y oportunidades

La computación en la nube representa un cambio considerable en la forma en que se estructuran las empresas y sobre todo sus departamentos de IT. Con la reducción de costos respecto a la adquisición de equipo para infraestructura y personal capacitado para la atención a toda la empresa, puede

representar una oportunidad sobre todo para medianas empresas que quieran partir desde cero. La reducción de costos es un hecho garantizado de la computación en la nube.

Esto resulta importante sobre todo para los emprendedores con ideas interesantes y un plan de negocios bien estructurado, pero que debe adecuarse a la realidad de su presupuesto. Suponiendo que la nueva empresa contrate un paquete de servicios de Cloud-Computing, no necesitará contratar servidores potentes ni costosos, tampoco equipo con demasiada capacidad; bastará con equipo que pueda soportar correr un navegador *web* y acceso permanente a *internet* para poder realizar sus actividades.

También permitirá acceder a esas aplicaciones desde cualquier lugar y desde otros dispositivos, incluso en otros países, una ventaja sobre todo si es necesario mudar la empresa de inmueble o si se quiere cambiar de equipo de cómputo.

Un cambio de esta naturaleza puede causar cierto escepticismo en algunos empresarios celosos de la naturaleza de la información que manejan, pues sus datos se operan y almacenan en *DataCenters* fuera de sus instalaciones y son manejados por personas externas a su empresa. Además está siempre presente la cuestión de la resistencia al cambio, inherente al ser humano.

Así que acompañado a este cambio y oportunidad están ciertas implicaciones como temas de seguridad informática, protección de la información, disponibilidad de acceso y apertura al cambio, además deberá preverse la flexibilidad que requiere funcionar con este nuevo paradigma, dado lo cambiante de las tecnologías y plataformas de los proveedores. Sólo de este

modo, teniendo confianza en ambas partes, se logrará el éxito de la implementación completa de servicios en la nube para empresas completas. El cambio resultará en una oportunidad que puede ayudar a las empresas a mejorar sus actividades y procesos.

3. RESPALDO DE INFORMACIÓN DIGITAL

El respaldo de información digital es un proceso común en cualquier institución educativa, de negocios, religiosa, de gobierno e individual. La importancia radica en el valor que puede representar la información en caso de que se pierda por cualquier acto voluntario o involuntario.

3.1. Introducción a los conceptos de respaldo de datos

Desde que se iniciaron a desarrollar los primeros sistemas de información, dicha información debió almacenarse de forma digital. Lo anterior derivó en la utilización de dispositivos de almacenamiento digital, que han tenido su propia evolución.

En primer lugar, se pueden mencionar las cintas magnéticas cuyo funcionamiento es similar a las cintas donde se grababan pistas de audio y que fueron populares durante 1980 y parte de 1990. También los discos flexibles, conocidos como *diskettes*; pero quizás los más utilizados sean los discos rígidos o discos duros, cuya arquitectura permitía guardar grandes volúmenes de datos. En la actualidad existen dispositivos como discos compactos, *DVDs*, memorias *flash*, discos de estado sólido o *solid state drives*.

Los arreglos de discos, RAID, son actualmente la mejor apuesta para la protección de información. Sin embargo, al igual que cualquier otro dispositivo electrónico, es susceptible a fallas. Actualmente son los dispositivos de almacenamiento más utilizados en los *datacenter* debido a su diseño que asegura un aumento en la eficiencia y seguridad en la información que en ellas

se almacena. El precio de estos dispositivos, aun cuando es alto, se compensa con su eficiencia, eficacia, seguridad y disponibilidad para la protección de la información.

La característica principal de todos estos dispositivos es que su funcionamiento es en la mayoría de los casos mecánico y en otros electrónicos, lo que ocasiona que tengan un tiempo de vida limitado y que estén constantemente vulnerables a fallos o daños ocasionados por diversas circunstancias como humedad, líquidos, golpes, descargas eléctricas, entre otras. Esto hace que peligre la información que en estos dispositivos se almacena.

En algunos casos la información puede ser tan importante que debe ser resguardada y protegida de una manera especial. Organizaciones militares, financieras o civiles pueden requerir que su información no solo tiene que estar segura contra ataques o espionaje, sino también protegerse contra algún tipo de fallo de un dispositivo de almacenamiento.

Por su naturaleza, estos dispositivos no pueden ser imperecederos o esperar que funcionen para siempre y jamás fallen, aunque el desarrollo de la tecnología puede hacerlos más resistentes y rápidos, pero sin lograr la perfección. Por lo antes expuesto, se han desarrollado técnicas de respaldo y recuperación, que van desde implementación de *hardware* hasta desarrollo de software y herramientas.

Todas estas técnicas van más orientadas a las bases de datos, que están siempre – o casi siempre – presentes en los sistemas de información, por lo que existen, por ejemplo, los sistemas distribuidos, las réplicas, servidores potentes, arreglos de discos, copias periódicas de la información en dispositivos

como cintas, etcétera. Estas técnicas son parte importantes en las políticas de protección de la información y pueden aplicarse por lo general varias de estas técnicas combinadas.

Quizás el mecanismo más simple y seguro de implementar es el respaldo de datos, en el cual las empresas hacen una copia de sus datos y se trasladan a otros dispositivos de almacenamiento que no son los discos duros en donde se procesan, sino otros más duraderos como discos compactos o cintas magnéticas. En el momento de que ocurra un fallo o un desastre, puede recuperarse parte, o en muy pocos casos la totalidad, de la información que se tenía en procesamiento al momento de la pérdida del dispositivo o del sistema completo.

3.1.1. Razones para hacer un respaldo de datos

Existen varias razones por las cuales se deba respaldar la información digital, se mencionan a continuación:

3.1.1.1. Respaldo de información

El respaldo de la información como un recurso valioso para todo individuo es una de las tareas que puede en algún momento convertirse en una tarea ardua según la importancia que esta tiene para él.

La misión de un proveedor de Cloud-Computing, es asegurar al cliente que su información siempre estará disponible y segura en todo momento. Para ello el proveedor aplicara todo tipo de técnicas que él considere apropiado para asegurarle al cliente la disponibilidad de su información. El cliente únicamente hace uso de su información cuando lo requiera.

3.1.1.2. Fallas en el medio

Los dispositivos de almacenamiento utilizados regularmente en las computadoras, como se explicó antes, funcionan en base a componentes electrónicos y mecánicos con riesgo latente a fallar por diversidad de razones, produciendo la pérdida total o parcial de toda la información que se pudo haber procesado durante un día o inclusive años de trabajo.

3.1.1.3. Operación incorrecta

La operación incorrecta puede resultar intencional o accidental y derivada de un mal manejo de la tecnología. Por ejemplo, la eliminación de archivos, eliminación de particiones lógicas, formateo, entre otros, pueden ser operaciones incorrectas de la tecnología y su consecuencia es la pérdida de información vital.

3.1.1.4. Ataques externos

Existen diversos elementos que pueden atacar un sistema como virus, gusanos, troyanos, usuarios malintencionados, que pueden hacer que los archivos se arruinen o se borren de nuestro dispositivo de almacenamiento. También puede mencionarse como pérdida algo que ocasione que el sistema operativo deje de funcionar obligando al usuario a formatear su disco y por ende perder su información. Un sistema puede sufrir un ataque externo en cualquier momento, aun cuando posea sofisticadas políticas de seguridad. El riesgo de un ataque externo únicamente se minimiza.

3.1.1.5. Incompatibilidad en el *hardware*

Este tipo de problemas pueden llegar a afectar el correcto funcionamiento del *hardware* asociado a la operación de los sistemas, lo que ocasiona que los datos que se almacenen en los dispositivos no sean correctos y la información no sea consistente, lo que puede tomarse como pérdida de información.

3.1.1.6. *Bugs* o problemas en *Drivers* o en aplicaciones

La existencia de *bugs* en los *drivers* relacionados al medio de almacenamiento o al sistema asociado a la información puede generar una pérdida debido a los problemas de lectura o escritura de datos. No solamente los *drivers* pueden tener estos *bugs*, sino también las aplicaciones desde las cuales se genera o procesa la información, su finalización repentina puede ocasionar pérdida de información que ya esté almacenada en el medio o dispositivo.

3.1.1.7. Desaparición del medio

Existe también la posibilidad de la pérdida del medio o dispositivo, ya sea por un desastre o por robo, así como daños materiales. Al encontrarse la información guardada en éstos medios, puede perderse al desaparecer el medio. (Descargar-Antivirus-Gratis.com, 2012).

Estas son sólo algunas de las razones por las cuales el usuario, departamento de IT o empresa se ve en la necesidad de respaldar su información. El autor del artículo anteriormente citado, muestra una serie de

razones en donde se expone que no solamente la combinación de dispositivos electrónicos y mecánicos puede exponer a un riesgo la información, sino está el factor del error humano o problemas a nivel de software o *firmware*.

La importancia de la información es algo que no puede medirse de forma objetiva, pues para un programador que necesita dar un entregable al día siguiente y pierde su información resultaría terrible; pero para un banco la pérdida de datos de transacciones hechas en un período de tan solo cinco minutos derivaría en pérdidas de miles o millones y los datos en sus sistemas carecería de consistencia. Es por ello que para cada empresa y cada usuario particular su información le resulta importante y no quisiera perder ni un solo *byte* por cualquier circunstancia.

3.1.2. Costo de pérdida de información

W. Curtis Preston se orienta al análisis empresarial del respaldo de datos y su recuperación en el momento de una eventualidad. Preston habla acerca de los costos que tendría perder información, sobre todo si se es una organización lucrativa y de prestigio.

3.1.2.1. Pérdida de clientes

Esta es una pérdida importante para las empresas, pues los datos de los clientes son imprescindibles y se pierde sus datos, probablemente se perderán para siempre los clientes de la empresa, o tomará un esfuerzo grande la recuperación de los clientes, lo que se traduce en pérdidas monetarias en cierta medida.

3.1.2.2. Pérdida de órdenes

La falla en los medios de almacenamiento también puede derivar en la pérdida de órdenes de productos o servicios realizados anteriormente, lo que podría ocasionar un retraso de horas, días o semanas en la entrega de dichos productos o servicios.

3.1.2.3. Moral

En el área de ventas, puede resultar una sensación de desaliento y desesperación para un buen vendedor el hecho de perder días de trabajo y ventas aseguradas, y se calcula que entre mejor sea el vendedor mayor impacto tendrá una pérdida de datos. Pero no solamente afecta a ventas, sino al empleado común, que al ver perdido su trabajo optará por retirarse a un lugar donde la infraestructura le permita trabajar de una mejor manera y mostrar sus capacidades sin tener la reputación de pérdidas.

3.1.2.4. Imagen

También está la cuestión de la situación competitiva de la empresa. La noticia de pérdida importante de datos puede esparcirse fácilmente y puede ser aprovechada por los competidores para su beneficio, además de presentar una pésima imagen, sobre todo si se tiene relación con entes reguladores o del Estado, como por ejemplo la pérdida de información importante dentro de un banco, llegará directo a la Superintendencia de Bancos y lo tomarán muy en cuenta.

3.1.2.5. Presupuesto

Solamente una historia de pérdida de datos le dará una mala reputación al departamento de informática sobre pérdida de información. Así que hay que deshacerse de una mala reputación, pues como el autor sugiere: “Eres tan bueno como tu última recuperación”, así que hay que lograr que las personas confíen en sus respaldos y recuperaciones, de lo contrario causará que cada quien haga sus propios respaldos – que es casi seguro que sean ineficientes e innecesarios – e incurran en gastos extras que no traerán consigo beneficios, sino todo lo contrario.

3.1.2.6. Tiempo

Como resulta obvio, se pierde tiempo. La pérdida trae consigo una pérdida de tiempo en recuperación, además de la imposibilidad de seguir adelante efectivamente, debido a que prácticamente se les paga por no hacer nada. (Preston, 2007).

Existen muchos factores por los cuales hay que hacer copias de seguridad, pero también existen aspectos clave en la implementación de sistemas de respaldo. Uno de esos aspectos es, por ejemplo, pensar en si de verdad se necesita personal especialmente contratado para hacer los respaldos y de controlar los insumos para hacerlo posible, pero si no se contrata personal para realizar los *backups*, el personal que ya está seguramente no le gustará contar con la carga de la responsabilidad de hacer esa tarea.

Y es que no se puede confiar plenamente en un sistema informático, o en el *hardware*, y seguramente los sistemas fallarán en el momento menos indicado, por eso la importancia de tener un respaldo que permita restaurar los

sistemas y recuperar lo mejor posible la información perdida. Todo en un tiempo casi inmediato, que es lo único aceptable para los gerentes y para los usuarios. Adicional, la recuperación garantiza menos pérdidas económicas.

Preston explica la dificultad y lo engorroso de realizar un trabajo de recuperación de datos, puede pasarse noches o fines de semana enteros tratando de recuperar información, pero actualmente se cuenta con una amplia variedad de soluciones para todos los problemas de seguridad. Es por ello que hay que tener mucho cuidado para elegir una herramienta, para no elegir algo equivocado o que pueda llevar al fracaso. El autor explica que hay que hacerse seis preguntas importantes (Preston, 2007):

- ¿Por qué?
 - ¿Por qué protegerse contra el desastre? ¿Realmente importa si se pierden los datos? ¿Qué implica las pérdidas? ¿Qué tipos distintos de datos se tiene, y cuál es el valor de cada tipo?

- ¿Qué?
 - ¿Qué se va a respaldar, la caja entera o solamente dispositivos o sistemas de archivos seleccionados? ¿Qué sistemas operativos se van a respaldar? ¿Qué más, aparte de las unidades normales o sistemas de archivos, debe incluirse en un respaldo?

- ¿Dónde?
 - ¿Dónde se producirá el respaldo? ¿Dónde es el mejor lugar para almacenar los volúmenes de respaldo?

- ¿Cuándo?
 - ¿Cuándo es el mejor momento para respaldar el sistema?
 - ¿Cuán a menudo debe realizarse un respaldo completo? ¿Cuándo debe hacerse un *backup* incremental?
- ¿Quién?
 - ¿Quién proveerá el *hardware*, software y servicios de instalación para poner el sistema en marcha?
- ¿Cómo?
 - ¿Cómo se va a lograr? Existe cierto número de maneras distintas para protegerse contra pérdidas. Investigar los distintos métodos, como el almacenamiento fuera de sitio, replicación, RAID (arreglos de discos) y los distintos niveles de protección que cada uno proporciona.

3.1.3. Tipos de *Backup*

Existen varios tipos de realizar un respaldo de datos, pues no resulta conveniente en todos los casos realizar una copia completa de todos los archivos, pues implica un costo de recursos como de esfuerzo especial.

3.1.3.1. Respaldo total

También denominado normal, crea una copia de todas las unidades, carpetas y archivos seleccionados, independientemente de su estado anterior, esto es, sin tener en cuenta si han sido modificados o creados desde la última copia de seguridad. Este tipo de copia de seguridad es el más recomendable

para guardar todo un sistema completo, al menos, una vez por semana, dependiendo lógicamente del volumen de información (Bricolaje, 2007).

3.1.3.2. Respaldos incrementales

Las copias incrementales guardan únicamente en el medio de respaldo seleccionado por el usuario aquellos ficheros que hayan sido modificados o creados desde la última copia incremental realizada.

La primera copia de este tipo, es idéntica por lo tanto a una copia total. Para reconocer los archivos creados o con cambios se utiliza en Windows el atributo *modificado*, que puede examinarse cuando se muestran las propiedades de un archivo (antiguamente en el modo MS-DOS, este atributo se identifica por la letra A de *Archive*). Este tipo de copia es el recomendado para *backups* más frecuentes, por ejemplo aquellos que se llevan a cabo a diario (Bricolaje, 2007).

3.1.3.3. Respaldos diferenciales

Un *backup* diferencial también copia los ficheros modificados o creados recientemente, pero en lugar de usar como referencia la última copia incremental, se usa la última copia completa. En este sentido, de la misma manera que la primera copia incremental se corresponde con una copia total, la segunda será diferencial. Este tipo de copias se restaura con mayor rapidez que una del tipo incremental, aunque para crearlas se necesita más tiempo (Bricolaje, 2007).

3.1.4. Estrategias de respaldo

Al entender bien los tipos de respaldo que existen, pueden establecerse ciertas estrategias, dependiendo del tamaño del sistema o de la empresa, para garantizar el éxito de un respaldo de datos.

En primer lugar se deben establecer ciertas cuestiones como la importancia de la información, la seguridad necesaria para protegerla. También es importante tomar en cuenta el tamaño y considerar que tiende a extenderse con el tiempo, en algunos casos a multiplicarse exponencialmente.

Al tener claros estos aspectos, puede elegirse una estrategia específica de respaldo de la información, e implementarla como está establecida u optar por alguna variante que resulte más beneficiosa para la organización.

Cuando se plantea la necesidad de hacer copias de seguridad de los datos es importante planificar una estrategia que contemple aspectos como la importancia de los datos que se guardarán (no tienen la misma trascendencia un documento de trabajo que un copia de respaldo de un programa), la periodicidad con la que se crearán la copias ola cantidad de medios (soportes) que se usarán y cuándo se utilizará cada uno de ellos.

No existe un criterio definitivo que indique la periodicidad con la que deben realizarse las copias de seguridad, aunque sí algunas consideraciones orientativas en este sentido, como el tiempo que ha sido necesario para crear los documentos (también su coste económico) o las consecuencias que suponen su pérdida. En general, si los documentos con los que se trabaja se modifican varias veces cada día, es recomendable realizarla copia como mínimo también a diario, y en algunos casos, creando copias de respaldo

provisionales en otras unidades de almacenamiento después de algunas horas de trabajo (Bricolaje, 2007).

El tiempo que deberán permanecer los datos guardados en un determinado soporte, es decir, el período de retención, está relacionado directamente con la rotación de medios elegida. Por ejemplo, si se ha decidido emplear una única cinta para realizar las copias, y se efectúa un *backup* a diario, lógicamente los datos guardados en la cinta nunca tendrán una antigüedad superior a este tiempo (Bricolaje, 2007).

Dado que el coste de los soportes para copias de seguridad será, con toda probabilidad, inferior al valor de los datos que contienen, es recomendable usar varias cintas, que se irán alternando a lo largo de copias sucesivas (Bricolaje, 2007).

Una estrategia extendida en empresas medianas y grandes consiste en almacenar tres copias diarias (por ejemplo, las correspondientes al lunes, miércoles y viernes inmediatamente anteriores), una copia de la semana anterior y otra del mes anterior (ésta última fuera del ambiente de trabajo) (Bricolaje, 2007).

Entre las estrategias de respaldo debe estar presente la rotación de medios, que permitirá tener la información segura en varios medios, aumentando el nivel de respaldo. A continuación se describen dichas estrategias (Bricolaje, 2007).

3.1.4.1. Esquema de tres cintas

El esquema de tres cintas comprende la utilización de tres unidades de almacenamiento – no necesariamente cintas magnéticas – y cuyo uso se distribuirá en los días de una semana, por lo general de lunes a viernes. Suponiendo una semana laboral de lunes a viernes, este esquema dicta que se debe realizar una copia total el primer día utilizando una cinta etiquetada con un número uno, los tres días posteriores se hará una copia incremental en una misma cinta etiquetada con el número dos, para que finalmente el último día laboral se haga otra copia total en la cinta etiquetada con un número tres. En este caso el tiempo de retención de la información digital no excederá de una semana de trabajo.

Una variante al esquema descrito anteriormente es la realización de *backups* incrementales de martes a viernes y hacer uno total el viernes, aumentando el tiempo de retención y da la posibilidad de utilizarlo como soporte de almacenamiento alternativo.

3.1.4.2. Esquema de seis cintas

En este esquema, se utiliza una cinta distinta cada día, utilizando adicionalmente dos cintas distintas el último día; una junto a las otras con los respaldos hechos anteriormente y otra como almacenamiento alternativo. Además, a diferencia del esquema de tres cintas, se realiza solamente una copia total a la semana. Las copias diarias pueden ser tanto incrementales como diferenciales, según las necesidades y características del sistema.

3.1.4.3. Esquema de 19 cintas

Principalmente utilizado en sistemas de grandes empresas y volúmenes grandes de información bastante importante. En este esquema la periodicidad es semanal para las copias incrementales, mientras que para las totales es mensual, así como las cintas utilizadas como almacenamiento secundario con una periodicidad de un año.

3.1.4.4. Esquema de 10 cintas

Este esquema ofrece un tiempo de retención de 12 semanas como máximo. Con este, se habrán usado todos los dispositivos de almacenamiento el mismo tiempo después de 40 semanas. Se hará una copia total los días viernes, utilizando una cinta distinta cada semana.

3.1.1. *Hardware para Backup*

En la actualidad, existen desarrollados un conjunto de dispositivos de almacenamiento, que pueden ser útiles tanto para el almacenamiento como para el respaldo de los datos. Tradicionalmente se piensa, al hablar de *backup*, en cintas magnéticas de gran capacidad; pero pueden ser útiles muchos otros dispositivos en términos de velocidad y seguridad. Pero, en realidad es cuestión de evaluar las necesidades y capacidades de cada empresa para poder tomar una decisión final.

W. Preston, establece que existen 8 factores de decisión, para establecer qué tipo de *hardware* es el necesario para realizar copias de seguridad dentro de la empresa. Los factores son los siguientes: confiabilidad, ciclo de trabajo,

velocidad de transferencia, flexibilidad, tiempo-de-datos, capacidad, amovilidad y costo. (Preston, 2007)

3.1.1.1. Confiabilidad

El primer aspecto a tomar en cuenta es la confiabilidad. Preston menciona que entre más piezas mecánicas tenga un dispositivo, más probable es que éste falle; pues los dispositivos mecánicos fallarían más a menudo que los dispositivos electrónicos. Es por ello que los dispositivos de cinta tienen más partes mecánicas que toda la computadora, lo que lo hace más susceptible a fallar que un dispositivo electrónico dentro de un servidor. Este riesgo siempre estará presente en las cintas magnéticas.

Deben buscarse mecanismos para determinar qué dispositivos son confiables, y no utilizar como principal referencia el tiempo medio entre fallos (MTBF, por sus siglas en inglés) que el fabricante propone, pues estos números están basados en situaciones ideales y ficticias para que resulte atractivo al cliente. Lo mejor es buscar asesoría o en foro de discusión que permitan consultar con personas que han tenido experiencias con respaldos y recuperación.

Lastimosamente, los dispositivos de cintas, aunque resulten baratos, también resultan menos confiables que unidades de disco; principalmente por contener demasiadas piezas mecánicas y por no poderse determinar en el momento de la copia, si los datos han sido copiados exitosamente o si se ha perdido información en el camino, sino hasta que se requiera una restauración.

La mejor opción según su confiabilidad son los dispositivos de disco, pues en los últimos años se ha reducido considerablemente su costo y ha

aumentado su capacidad, ofreciendo gran capacidad a un precio relativamente bajo. Además proveen la opción de implementar un arreglo de discos o RAID, que son más controlables que los dispositivos de cinta y permiten monitorear su estado en tiempo real y adicionalmente, proveen mecanismos de verificación y recuperación de datos, por su arquitectura.

3.1.1.2. Ciclo de trabajo

El ciclo de trabajo, es el tiempo para el cual ha sido diseñado el dispositivo para que funcione. Por ejemplo, un dispositivo que tiene un 50 de porcentaje de tiempo de ciclo de trabajo, se espera que pase trabajando unas 12 horas al día (Wikipedia, 2012).

Respecto a los discos, la arquitectura SCSI está diseñada para un alto ciclo de trabajo, es por ello que es utilizada en sistemas de bases de datos o en servidores de alta disponibilidad. En lo referente a los dispositivos de cinta, varían sus ciclos de trabajo según para qué fueron hechos, por ejemplo, los dispositivos de cinta para mainframe tienen un alto ciclo de trabajo.

Hay que tomar en cuenta que en algunos casos, una copia de seguridad será creada una vez y accedida muy de vez en cuando, lo que hace reducir la necesidad de ciclo de trabajo; pero otros casos requieren que los dispositivos sean usados muy a menudo, aumentando el nivel de ciclo de trabajo requerido para cumplir la tarea. Por ello entra en juego la política de respaldo utilizada y así determinar qué tanto se utilizará un dispositivo y con qué frecuencia.

Dado lo anterior, se recomienda buscar un dispositivo cuyo ciclo de trabajo coincida con el del sistema que se está utilizando, para no caer en errores como comprar un dispositivo que tenga un ciclo de trabajo bajo para un

sistema que requiera uno más alto, lo que provocará que falle constantemente, o el caso contrario que consistiría en comprar un dispositivo con un ciclo de trabajo alto para un sistema con un ciclo bajo, lo que incrementaría los gastos considerablemente y de forma innecesaria.

Por lo tanto, si se tiene un dispositivo que coincide con el sistema en funcionamiento, su tope de capacidad sería el mismo que el del sistema, asegurando tanto la confiabilidad del dispositivo, así como evitar gastos innecesarios.

3.1.1.3. Velocidad de transferencia

La velocidad de transferencia, es el indicador de cuánto tarda el dispositivo en leer, escribir y verificar los datos. Preston indica que la velocidad de transferencia que en realidad debe compararse es la “velocidad de transferencia nativa”, que es la velocidad de transferencia de los datos sin compresión. Distintos algoritmos de compresión, dan como resultado distintas velocidades de transferencia, por ello para evitar engaños o manipulaciones por parte de los fabricantes, se debe verificar la velocidad nativa.

3.1.1.4. Flexibilidad

Preston define la flexibilidad de los dispositivos en dos temas principales: responder a diferentes velocidades de datos y la posibilidad de ser utilizados de distintas maneras. Basándose en ese concepto, se puede concluir que los dispositivos menos flexibles son las cintas, seguido por los dispositivos ópticos, mientras que las unidades de disco son las más flexibles y por lo tanto conviene más su uso en términos de este factor. El orden en términos de flexibilidad de uso de dispositivos sería, unidades de disco, cintas y dispositivos ópticos.

3.1.1.5. Tiempo-de-datos

Este factor mide cuánto tarda cargar un volumen, buscar el lugar indicado en el volumen e iniciar a leer los datos. Con esto nuevamente la cinta está en desventaja, puesto que su acceso es secuencial y necesita un rebobinado, así como la localización exacta del archivo en cuestión tarda bastante tiempo; lo que no sucede con las unidades ópticas, pues su acceso es más fácil por su arquitectura y su tiempo de acceso es de menos de un segundo, si el disco está cargado en la unidad lectora. Finalmente acá la ventaja la tiene de nuevo las unidades de disco, pues sus mecanismos permiten un acceso directo y no secuencial, y su búsqueda es más fácil y rápida, en tiempos de nanosegundos.

3.1.1.6. Capacidad

Respecto a la capacidad, debe tomarse en cuenta el tamaño de los datos que requieren respaldarse. Uno de los propósitos, como se ha venido mencionando anteriormente, es evitar incurrir en gastos innecesarios como comprar un dispositivo de grande capacidad para un volumen de datos de muy bajo. Por ello hay que establecer las verdaderas necesidades para no desperdiciar recursos, aunque podría optarse por tener un dispositivo de amplia capacidad y hacer dentro de él varias copias pequeñas, con el riesgo de que si falla, se perderían todos los respaldos.

Quizás la mejor opción sería de nuevo los dispositivos de disco, pues su capacidad respecto al costo ha mejorado en los últimos años, además de presentar ventajas en otros factores antes mencionados, aumentando su valor.

3.1.1.7. Amovilidad

La amovilidad también puede llamarse “desmontabilidad”, que es la capacidad de poder ser quitado de su lugar. En esto llevan la ventaja las unidades de cinta, así como las unidades ópticas, pues su amovilidad es bastante grande y resulta muy sencillo cambiar una unidad por otra. Esta característica se hace notable sobre todo para aquellas empresas que necesitan que sus dispositivos de almacenamiento sean transportados y almacenados en bóvedas, mas no para sistemas de replicación. Todo dependerá de las características de la empresa y sus necesidades.

3.1.1.8. Costo

Sin duda alguna, el costo es algo de tomar muy en cuenta. Es principalmente lo que interesa a las personas que generalmente aprueban los presupuestos, aunque no ahonden en detalles técnicos o tecnológicos. Nuevamente todo depende de la situación de la empresa. Aunque se ha hablado de flexibilidad, tasa de transferencia, capacidad, entre otros; los dispositivos que tienen mejor valoración respecto a dichos aspectos resultan más costosos y más difíciles de obtener.

Además no solo debe tomarse en cuenta el costo del dispositivo en sí mismo, sino todo lo que implica decidirse por uno. Por ejemplo si se utilizan unidades de cinta, se debe incluir el costo de los dispositivos lectores, el lugar del almacenamiento de las cintas y la planificación que tomará implementarlas. Tampoco olvidarse del costo que conlleva comprar dispositivos baratos y de dudosa calidad, que al fallar se incurre en un nuevo gasto, aumentando el monto del costo, aplicando el dicho popular “lo barato sale caro”.

Dispositivos de almacenamiento en la actualidad hay una amplia variedad, con distinta arquitectura, tecnología, costo, capacidad y tamaño. Los más medios utilizados en el almacenamiento de copias de seguridad, generalmente son: cintas magnéticas, unidades ópticas, unidades de disco.

3.1.1.9. Cintas magnéticas

Estos dispositivos han sido tradicionalmente los utilizados como unidades de respaldo de información en las empresas. Su acceso es secuencial, no directo como los discos duros, su portabilidad es bastante beneficiosa y puede almacenar grandes volúmenes de información.

Uno de sus principales beneficios es que puede hacerse un *backup* completo en ellas, y como es un dispositivo completamente externo puede trabajar paralelo a la computadora y su funcionamiento se puede automatizar. Su costo respecto a almacenamiento es bajo.

Su desventaja es que su acceso es secuencial, por lo que la búsqueda de un archivo específico puede tomar demasiado tiempo, además de que no se tiene la seguridad de haber hecho una copia exitosa, hasta que en realidad se necesite utilizar la cinta para hacer una recuperación o bien hasta que un proceso de verificación rutinario detecte eventuales errores de escritura en el medio.

Existe variedad en el diseño, capacidad y velocidad en los dispositivos de cinta, esto quiere decir que al momento de tomar la decisión de que se va a utilizar un dispositivo de cinta para copias de seguridad, también habrá que pensar en qué tipo de dispositivo se va a adquirir. Algunas de las tecnologías de cintas son las siguientes (Galicia City, 1999):

- DLT: tecnología de línea digital, la información se almacena en cintas de media pulgada, acomodada en línea y un patrón de serpentina, permitiendo grabar de ambos lados de la cinta.
- QIC: cintas de un cuarto de pulgada, cuyo diseño es parecido al de cintas de *cassette*.
- TRAVAN: esta tecnología surge para estandarizar la arquitectura QIC, cuyo problema era la incompatibilidad respecto a los nuevos modelos que iban saliendo.
- DAT: cintas de 4 milímetros. con una arquitectura parecida a las cintas de video, su capacidad es mayor, pero su velocidad es inferior que las anteriores descritas.
- 8MM: tecnología creada para video, pero adaptada para almacenamiento de información, su principal característica es su alta capacidad de almacenamiento.

3.1.1.10. Unidades ópticas

Estos dispositivos poseen las ventajas de la amovilidad que proveen las cintas, aunque su capacidad de almacenamiento es considerablemente inferior. Una unidad óptica es aquella que necesita un lector laser para su acceso, así como su grabación.

Su principal ventaja es la portabilidad que ofrece, su tamaño físico es compacto, su precio es bajo, el acceso no es secuencial como una cinta, y su confiabilidad es mayor a las unidades de cinta. Existen distintas tecnologías

dentro de las unidades ópticas y su funcionamiento; como la forma en que se almacenan los datos, pues existen métodos magnéticos, ópticos y de cambios de fase, pero el resultado es el mismo: traducción a datos binarios en el disco por medio de cambios en la superficie, lo cual posteriormente se leerá como información.

Existen también varios formatos de disco, como CD, DVD, Blue-ray. Dentro de estos mismos, existe una serie de sub formatos, como CD-R, CD-RW, DVD-RAM, DVD-R, DVD-RW, DVD+R, DVD+RW. Las capacidades de estos dispositivos varían entre algunos MB y varios GB, pero su capacidad aún está muy limitada respecto a dispositivos como cintas o discos, por lo que su uso para *backup* es para copias de seguridad pequeñas, ideales para pequeñas y medianas empresas con un volumen no tan grande de información.

3.1.1.11. Unidades de disco

Si bien las unidades ópticas antes mencionadas son unidades de disco, en esta sección hará énfasis a las unidades de disco duro. Estas unidades son de arquitectura cerrada, lo que hace que se protejan mejor que otros dispositivos, además que tradicionalmente son utilizados en servidores y computadoras. Su variedad de tecnologías está relacionada con la conexión de los datos, estas son: IDE, SCSI, ATA, SATA, SAS, RAID. Si bien son más confiables que los dispositivos de cinta, estos dispositivos cuentan con piezas mecánicas que suponen un riesgo de fallas.

Sus ventajas son la velocidad de transferencia, su amplia capacidad, su seguridad y confiabilidad. Aunque anteriormente significaban un gasto significativo respecto a su capacidad de almacenamiento, recientemente se ha reducido su costo haciéndolo más accesible y con una mayor capacidad. Pero

su capacidad no solo se queda allí, su arquitectura permite realizar arreglos de discos, que es una forma de protección de datos y recuperación a fallos, lo que lo convierte en una muy buena opción de almacenamiento. Su desventaja mayor es que no es tan fácil de desmontar y carece de amovilidad; su funcionamiento no es tan independiente al de un computador, lo que limita su reemplazo o portabilidad.

3.1.1.12. Otros dispositivos

Existen otros dispositivos de almacenamiento de información, nuevas tecnologías han permitido la creación de dispositivos como las memorias *pendrive* (memorias USB), los discos de estado sólido, dispositivos SD, entre otros. Lo que no les ha permitido ser utilizados como dispositivos típicos de respaldo es su limitada capacidad, su alto costo y su poco desarrollo tecnológico y, mayormente si son nuevas tecnologías que no han sido probadas lo suficiente como para hacer en ellas sistemas confiables de respaldo.

3.2. Importancia del respaldo de datos de la información digital para las grandes y medianas empresas

Desde hace mucho tiempo se ha vuelto una necesidad para las empresas, dentro y fuera del ámbito guatemalteco, la utilización de sistemas informáticos para automatizar sus procesos, así como almacenar y procesar su información de una manera más rápida, confiable y segura. Como se ha mencionado anteriormente, la implementación de sistemas informáticos conlleva la necesidad de almacenar la información que se maneja en dispositivos.

Estos dispositivos han evolucionado en los últimos tiempos, y aunque han aumentado su capacidad, fiabilidad, entre otras características, sigue siendo un riesgo la pérdida de información por las razones ya anteriormente expuestas, como errores, robo, desastres, etc.

En lo que respecta a una empresa, es imposible pensar que el dueño o cualquier gerente estén dispuestos a perder información importante, en cualquier ámbito. Doctores que pierdan datos de sus pacientes como sus antecedentes médicos, proveedores de materias primas que pierdan los datos de las órdenes o pedidos ya extendidos, un sistema bancario que tenga importantes pérdidas financieras debido a transacciones cuyos datos fueron perdidos, organizaciones educativas que pierdan los datos de las notas de sus alumnos. Existe una infinidad de casos que se pueden plantear y en cualquiera la pérdida tendría consecuencias graves en las organizaciones.

Sobre todo porque la importancia de los datos es algo muy subjetivo y depende del ámbito de la empresa y a lo que se dedica, pero si algo es seguro es que si se le preguntara a cualquier persona qué por ciento de datos está dispuesta a perder, su respuesta será un definitivo 0 por ciento. La cuestión ahora sería plantearle qué tanto está dispuesto a invertir tanto en dinero como en esfuerzo para implementar un sistema de respaldo verdaderamente efectivo, que por supuesto le resultará más costoso entre mejor sea; su opinión cambiará acerca del porcentaje de pérdidas.

Principalmente para las grandes empresas, cuyo volumen de información es enorme. Para estas empresas se necesita una gran inversión en tema de protección de información. No solo invierten en respaldo de datos, sino también implementan mecanismos de alta disponibilidad, arreglos de discos, seguridad, replicación, distribución de carga, etc.

Todos los datos que manejan son lo más importante, pues es su trabajo completo, todas sus actividades están almacenadas en sus servidores, en sus dispositivos de almacenamiento, incluso dentro de sus computadoras personales. Pequeños detalles como la carta a enviar al proveedor, al jefe de recursos humanos, los contratos que deben imprimirse para la próxima semana, todo es importante e imprescindible, nadie quiere perder nada.

En algunas organizaciones, algunos respaldos de datos suelen hacerse sin siquiera la intervención del departamento de IT, en las oficinas las secretarías o los operadores quizás sin saberlo realizan respaldos de datos, al tener una copia de sus archivos importantes en sus “memorias USB”, hasta los grandes respaldos que se hacen de las bases de datos, programados según su estrategia y con dispositivos avanzados y sofisticados; servicios de traslado de cintas a bóvedas para proteger los datos y tener un *backup* fuera de sitio. En realidad no solo es importante el respaldo de datos, es una mera necesidad. El respaldo de la información forma parte de la política de seguridad de la información.

De igual manera es necesario en las PYMES, aunque quizás no manejen tal cantidad de información. En este caso, los mecanismos de respaldo pueden ser más sencillos, pero el problema es que no se le suele dar tanta importancia como en realidad debería. Dentro del contexto latinoamericano, en Argentina se realizó una encuesta entre 300 pequeñas y medianas empresas argentinas. El estudio muestra que sólo el 20 por ciento de ellas cuenta con un plan estandarizado de respaldo de sus datos (Fuzionia, 2010).

La encuesta evaluó el grado de cumplimiento de los resguardos básicos que deben hacerse a la hora de encarar una estrategia de protección: tener un plan de *backup* armado, probarlo al menos dos veces al año para asegurarse

que funciona correctamente, realizar un *backup* de la información crítica en forma diaria utilizando un medio magnético adecuado y uno integral al menos una vez por semana, y mantener dos copias del *backup*, una en un sitio externo y otra en las propias oficinas (Funziona, 2010).

Sólo el 7,4 por ciento realiza *backups* diarios de su información crítica. El 21 por ciento lo hace semanalmente, el 23,1 por ciento de manera bimestral, mensual o trimestral y el 48,5 por ciento, sin periodicidad fija. Sólo el 17 por ciento de las empresas consultadas tiene una copia del *backup* en un sitio alejado de sus oficinas. Apenas el 8 por ciento de las firmas consultadas ha probado sus sistemas de *backup* una vez en los últimos doce meses, y sólo el 1,5 por ciento lo ha hecho en más de una oportunidad en el último año. El 5 por ciento lo hizo alguna vez, y el 85,5 por ciento restante no lo ha hecho nunca (Funziona, 2010).

Entre más pequeña sea una empresa, seguramente menor cantidad de información maneja, dada la naturaleza de sus actividades; y generalmente se le da menor importancia no a la información en sí, sino a la pérdida de esa información y se adoptan menos medidas de seguridad para evitarlo o para tener la capacidad de recuperación.

Y es que en algunos casos todo es cuestión de costos, puede que algunos dueños de pequeñas o medianas empresas piensen que respaldar sus datos requiere un presupuesto bastante alto y que es un lujo, pero incluso puede hacerse en dispositivos ópticos como CDs o DVDs que están al alcance de cualquiera, todo es cuestión de adoptar una política de protección de datos y una estrategia de *backup* efectiva y que se acople a sus necesidades.

4. BANCO DE DATOS VIRTUAL

El Banco de Datos Virtual pretende la adopción de un nuevo paradigma el cual no es más que adaptar el paradigma actual al avance de las innovaciones tecnológicas.

4.1. Adaptación más que un nuevo paradigma

Los métodos de *backup* de datos utilizados actualmente desde un hogar con una única computadora hasta una empresa trasnacional con miles de computadoras, pueden todos ellos tomarse como parte de un paradigma. Todos los métodos utilizan un dispositivo físico conectado al computador o a la red, en la cual se almacenan los datos. El dispositivo tiene la particularidad de residir en las instalaciones físicas en donde el *backup* de datos se realiza. Los datos son trasladados a un medio físico de almacenamiento, el cual también suele luego depositarse en alguna locación física dentro de la compañía o en el mejor de los casos, fuera de ella.

El paradigma agrupa todo este conjunto de ideas que dictan las prácticas de respaldo de datos que la mayor parte de las empresas actualmente utilizan. Sin embargo, la idea principal que hace que esto pueda llamarse paradigma, es el hecho de que la información respaldada se encuentra en todo momento al alcance físico del usuario.

Para un usuario, respaldar la información significa prevenir la pérdida de la misma para uso futuro y la pronta disposición de la información respaldada en el momento que lo requiera. Aun cuando esto también es posible con los

grandes avances en la ciencia de la computación y específicamente en las aéreas de redes y almacenamiento. Los usuarios tienden a inquietarse si no tienen la información respaldada a su alcance físico en un medio físico.

Para grandes empresas que cuentan con oficinas distribuidas en diferentes partes del mundo, la información digital se vuelve en uno de los principales activos que buscan proteger a toda costa. El respaldo de la información es por lo general depositado en empresas que se dedican al resguardo de los medios de almacenamiento digital mediante cajas de seguridad.

Actualmente existen grandes empresas que se dedican al respaldo de información de diferentes empresas. Se puede apreciar que la confianza que depositan las empresas en este tipo de compañías se debe al contrato de confidencialidad, seguridad y disposición que en su momento se haya firmado. No es común que las empresas deseen confiar su información a terceros, sin obtener garantías de que esta no será proporcionada a alguien ajeno a la empresa dueña de la información.

El almacenamiento de datos en los servidores de empresas que se dedican al respaldo de información no es un nuevo paradigma, es la adaptación del modelo de respaldo de datos existentes a las nuevas e innovadoras formas de almacenamiento de datos que la Cloud-Computing ofrece.

Actualmente las velocidades de conexión a *Internet* son lo bastante rápidas como para confiar en replicar los datos de la empresa a algún servidor de un tercero y descargarla en cuanto se necesite. Los medios de almacenamiento soportan capacidades de almacenamiento lo suficientemente grandes como para depositar grandes cantidades de Terabytes en ellos.

Tecnologías como la SAN/NAS permiten tener cantidades de almacenamiento increíblemente grande. *Failover* permite asegurar en el mejor de los casos un 99,99 por ciento de disponibilidad de los sistemas y asegurar al cliente la prestación ininterrumpida de servidores de aplicaciones, datos y/o *hardware*.

Todo este conjunto de tecnologías y algunas otras más, combinándolas, permiten crear un sistema de respaldo de información el cual podría ser un análogo entre la combinación de un sistema de bancos y el paradigma de respaldo de datos. Sin embargo, no se crearía ningún paradigma nuevo, únicamente se adaptaría el paradigma ya existente a la tendencia tecnológica actual: Cloud-Computing. Se adaptarían los beneficios que ofrece Cloud-Computing al paradigma de respaldo de datos y en el proceso, se reducirían costos de operación, logística, mantenimiento y protección, que actualmente requiere.

4.2. Reducción de costos

Es un hecho que todas las empresas actuales buscan reducir sus costos de operación y así tener más capital para invertir en nuevos proyectos. Un Banco de Datos Virtual, reduce costos.

Los medios físicos donde se almacenan datos, los dispositivos utilizados para almacenar información, el espacio físico para almacenar los medios con información digital conlleva un costo asociado a él. Para empresas grandes, el costo es periódico y fijo. Mientras que para empresas pequeñas el costo únicamente es variable según la necesidad que tengan de respaldar los datos. Para empresas que utilizan cajas de seguridad en entidades bancarias, el costo de respaldo de datos se incrementa en gran manera.

El uso del servicio de respaldo de información por medio de un banco de datos, únicamente requiere del costo asociado a la conexión permanente de *Internet* y el costo del contrato por el pago del servicio. Para que un banco de datos sea atractivo para las empresas, el costo por el almacenamiento de información debe estar de acuerdo a la cantidad de *Gigabytes* de información digital almacenada en los servidores del banco de datos.

Empresas como Rsync.net ofrecen actualmente el servicio de *backup* de datos *online*, el costo es cobrado de acuerdo a la relación GB/Mes y agrupado según la necesidad de replicación de datos o no. (<http://www.rsync.net>)

Tabla III. **Precios de *backup* de datos a la fecha 04 febrero 2012**

Servicio	Replicación	Costo
Respaldo GB/Mes	No	USD 0.80
Respaldo GB/Mes	Si	USD 1.40

Fuente: Página oficial de Rsync.net.

Rsync.net únicamente ofrece el servicio de respaldo de información. Sin embargo, dado que este servicio es un análogo a un *backup* de datos físico, se aprecia que el costo de almacenar 1GB/Mes de información digital representa un ahorro en costos de medios de almacenamiento digitales, dispositivos de almacenamiento y espacios físicos para el depósito de los medios de almacenamiento.

Un banco de datos representa una inversión en reducción de costos para el respaldo de información.

4.3. Disponibilidad

La disponibilidad se utiliza para expresar numéricamente el porcentaje de tiempo que un servicio está disponible para su uso. Cuando se quiere implementar un servicio en Cloud-Computing, éste debe de tener un nivel de no inferior a 99 por ciento (Wikipedia, 2012), pues todas las operaciones de una empresa dependerán del sistema informático y si este se encuentra fuera del sitio de la misma se requerirá con mayor razón.

En lo que respecta a un servicio de banco de datos, un cliente requerirá que sus respaldos estén disponibles a cualquier hora del día, cualquier día del año, pues un imprevisto requerirá la intervención inmediata y la recuperación de datos. Asimismo, se requerirá el servicio en cualquier momento para realizar una copia, y sobre todo el cliente requerirá que el servicio esté disponible en los momentos en los que las copias estén programadas.

Y es que no solamente se requerirá la disponibilidad del servicio de banco de datos, también se dependerá del proveedor del servicio de *internet*, cuyo caso será tanto o más crítico que el del proveedor del servicio de almacenamiento, porque todo el proceso requerirá conexión permanente a *internet*.

En un caso ideal, uno de los requerimientos es la utilización de un ancho de banda grande con uso de fibra, asegurando una transferencia de alta velocidad, así como contar con un canal directo al servidor; es decir, no compartir el ancho de banda con el acceso a *internet* de la empresa, sino una dirección IP pública dedicada única y exclusivamente a la transferencia de los datos para los respaldos.

Pero, tal y como sucede con los dispositivos en el lugar, entre más disponibilidad se requiera, mayor será el costo del servicio y la implementación. Así que es importante definir niveles mínimos de disponibilidad, así como tiempos de espera para la restauración de un sistema.

4.4. Seguridad

Rsync.net únicamente se dedica al respaldo de información digital en sus servidores y cobra una tarifa mensual por Gigabyte almacenado en sus servidores. No garantiza la seguridad de la información depositada en sus servidores.

Para las empresas medianas y grandes, la seguridad de su información será siempre prioridad. El cliente siempre querrá tener la certeza de la capacidad de los servidores para el almacenamiento de datos, anchos de banda para sus servidores u otros equipos, que se le asegure que su información estará libre de corrupción, accesos indeseados, borrado y otros tipos de daño o alteraciones maliciosas que pueda sufrir.

Casos como el de *WikiLeaks* no deberían presentarse en un sistema de Banco de Datos Virtual (Wikipedia, 2012). El acceso a la información debe ser garantizada única y exclusivamente al cliente. La eliminación, modificación deberá ajustarse a políticas que busquen garantizar la integridad de la información y la violación de esta política únicamente podrá ser rota por el cliente exclusivamente. El Banco de Datos Virtual debe garantizar que el único responsable de la información almacenada en sus servidores es el cliente.

El acceso a la información se protegerá por medio de cuentas de usuario autorizadas por el cliente y gestionadas con el Banco de Datos Virtual. La

eliminación y modificación de la información en la cuenta del cliente, se registrará por una serie de privilegios que se le otorgarán a un usuario administrador de las cuentas de acceso del cliente. Sin embargo, la pérdida de la información por desastres naturales que escapan de las manos de los clientes, corren por cuenta del Banco de Datos Virtual. El Banco de Datos Virtual podrá valerse de replicación de datos de sus centros de datos ubicados en diferentes países para garantizar la integridad de la información de sus clientes.

Garantizar la seguridad de la información de los clientes incrementará el costo del servicio de respaldo de información de datos. El costo asociado a la seguridad, también debe cobrarse en términos de GB/Mes y es el Banco de Datos Virtual quien decidirá el costo.

4.5. Banco de Datos Virtual

Un banco, es una entidad la cual es un intermediario para la captación de recursos en forma de depósitos y el préstamo de servicios financieros. El principal recurso de un banco financiero es el dinero. Las actividades de captación, almacenamiento y devolución del dinero de los clientes son las únicas que serán análogas al Banco de Datos Virtual. Un banco financiero adicionalmente ofrece el servicio de préstamos de dinero, el cual no se emulará en el sistema de un Banco de Datos Virtual. Cada cliente es identificado en el banco financiero por medio de una cuenta, en la cual realiza sus depósitos y retiros.

Un Banco de Datos Virtual será un intermediario para la captación de recursos en forma de depósitos. El principal recurso del Banco de Datos Virtual será la información digital, específicamente bytes. El servicio de préstamo que ofrece un banco de datos financiero no podrá emularse en un Banco de Datos

Virtual debido a que esto implicaría brindar la información de los clientes a terceros. El compartir información de los clientes a terceros violaría las políticas de seguridad, esto jamás debe permitirse bajo ninguna circunstancia más que por un medio legal.

La disponibilidad garantizará el servicio de depósito y retiro de información de los clientes en su cuenta en el Banco de Datos Virtual. Por medio de la conexión a *Internet*, el cliente podrá acceder en cualquier momento y en cualquier lugar a su información. El Banco de Datos Virtual deberá contar con una infraestructura la cual incluya *failover*, redundancia, *mirror* y replicación entre diferentes centros de datos que posea dispersos alrededor del mundo.

También garantiza para el caso de ejecutivos que viajen a diferentes partes del mundo por citas de negocios, el que puedan en la noche al llegar al hotel, conectar sus computadores al *Internet* y respaldar su información. En el caso inverso, los clientes puedan acceder a su información en diferentes partes del mundo en el cual cuenten con conexión a *Internet* y un computador para visualizar y descargar la información que requieran.

La seguridad permitirá garantizarle al cliente que su información estará segura todo el tiempo de cualquier eventualidad. Se permitirá el almacenamiento, visualización y descarga de información en la cuenta del cliente del Banco de Datos Virtual únicamente por medio de cuentas de acceso previamente registradas. El banco de datos deberá colocar en el contrato de arrendamiento del servicio, una cláusula de confidencialidad en el cual se compromete a no proporcionar la información depositada en sus servidores más que al cliente, y en caso de ser requerido por un tercero, la solicitud de entrega de datos deberá ser gestionada por medio de una demanda legal y una orden judicial dictada por un juez competente. La entrega de datos por medio de una

demanda legal a un tercer interesado en un país, deberá regirse al sistema fiscal y judicial del país del cual es originara la empresa dueña de la información.

Al igual que un banco financiero cuenta con una agencia central y una serie de agencias distribuidas a lo largo del país o países en los cuales opera; un Banco de Datos Virtual podrá análogamente tener diferentes *DataCenters* ubicados en diferentes partes del mundo. Los clientes podrán almacenar su información en un *DataCenter* ubicado en Europa, visualizar su información en un servidor ubicado en EEUU y descargar la información de algún servidor ubicado en Sur-América. El cliente visualizará al Banco de Datos Virtual como una única entidad.

El Banco de Datos Virtual adicional al respaldo de información digital, también prestará el servicio de seguridad, disponibilidad e integridad de la información digital que los clientes depositen en sus cuentas. El costo de estos servicios deberá ser cobrado en una única factura mensual y ser presentado como un único servicio.

Aun cuando algunos ejecutivos piensen en cobrar por cada servicio independientemente, el concepto de Banco de Datos Virtual que presentamos los agrupa a todos ellos como un único servicio. La seguridad, disponibilidad, integridad y respaldo de información es el servicio el cual podrá ser facturado a los clientes según el depósito mensual de Gigabytes en los servidores del Banco de Datos Virtual.

5. IMPLEMENTACIÓN DE UN BANCO DE DATOS VIRTUAL

La implementación del Banco de Datos Virtual que se describe a continuación muestra el *hardware* y software necesario para hacerlo. Esta implementación tiene como objetivo demostrar que es posible realizarlo y puede ser utilizada como guía para todo aquel empresario innovador que estuviese interesado y cuente con los recursos para desarrollarlo.

5.1. *DataCenter*

La infraestructura de un *DataCenter*, está compuesta en un 80 por ciento de servidores SAN/NAS, el resto está dividido en la infraestructura de red de comunicaciones y eléctrica. A continuación se describe el *hardware* utilizado para la demostración de una posible metodología de implementación de un Banco de Datos Virtual.

5.1.1. SAN / NAS

Una SAN (*Storage Area Network*) o Red de Área de Almacenamiento, es una red diseñada específicamente para conectar distintos dispositivos de almacenamiento, vistos como nodos y que en conjunto almacenan información.

Esta arquitectura habilita el acceso rápido y confiable a los servidores a recursos de almacenamiento, externos o independientes, sin importar su ubicación geográfica física.

En una SAN, un dispositivo de almacenamiento no es propiedad exclusiva de algún servidor. La mayor parte de las veces, los dispositivos son compartidos por servidores en red como recursos peer (Machuca, 2008).

Una analogía válida, es que por ejemplo las LAN o Redes de Área Local sirven para la conexión de clientes a servidores, las SAN permiten conectar los servidores a unidades de almacenamiento, así como también a otros servidores, o el caso de conexión de unidad a unidad de almacenamiento.

5.1.1.1. Características principales de una SAN

Las características que hacen atractivas a las SAN son las siguientes (Mundo CISCO, 2009):

5.1.1.1.1. Rendimiento

Permiten acceso concurrente por dos o más servidores lo que proporciona un mejor rendimiento.

5.1.1.1.2. Disponibilidad

Se puede hacer una copia exacta de los datos a una distancia de 10 Kilómetros lo que las hace más seguras.

5.1.1.1.3. Escalabilidad

Como las LAN/WAN puede usar muchas tecnologías. Lo que permite fácil reubicación, seguridad migración y duplicación de datos.

5.1.1.1.4. Seguridad

La seguridad en las SAN ha sido desde el principio un factor fundamental, desde su creación se notó la posibilidad de que un sistema accediera a un dispositivo que no le correspondiera o interfiriera con el flujo de información, es por ello que se ha implementado la tecnología de zonificación, la cual consiste en que un grupo de elementos se aíslen del resto para evitar estos problemas.

La zonificación puede llevarse a cabo por *hardware*, software o ambas, siendo capaz de agrupar por puerto o por WWN (*World Wide Name*), una técnica adicional se implementa a nivel del dispositivo de almacenamiento que es la presentación, consiste en hacer que una LUN (*Logical Unit Number*) sea accesible sólo por una lista predefinida de servidores o nodos.

Por otro lado, está la tecnología NAS, que permite compartir la capacidad de almacenamiento de un dispositivo o servidor con clientes a través de una red. Esta tecnología utiliza un propio sistema operativo, dedicado exclusivamente al almacenamiento y control del mismo.

Se utiliza principalmente para grandes cantidades de almacenamiento, reduciendo costos y aprovechando mejor los recursos que se tienen al alcance. Regularmente los dispositivos conformados en NAS son arreglos de discos (RAID).

Una tecnología SAN/NAS, aprovecha los beneficios de ambas convirtiéndose en un híbrido entre las dos, permitiendo almacenar grandes volúmenes de información y distribuyendo el almacenamiento en varios dispositivos, que pueden estar dispuestos geográficamente en lugares distintos.

El Banco de Datos Virtual, puede aprovechar estos beneficios, aportando las características antes mencionadas, cumpliendo con requerimientos como seguridad y disponibilidad, además de encontrarse en lugares distintos para una distribución de la carga de operaciones. Aun disponiéndose los dispositivos en forma distribuida, permite un almacenamiento centralizado viéndose todo el *storage* como uno solo, accediendo y administrándolo desde un navegador y éste utiliza a su vez un propio sistema operativo optimizado y dedicado al almacenamiento, separando la ejecución de datos en otro servidor o servidores.

5.1.2. iSCSI

iSCSI, es un protocolo de almacenamiento en redes basado en el protocolo IP y su función principal es permitir la unión entre las instalaciones de almacenamiento de datos. El protocolo permite la transmisión de comandos SCSI a través de redes IP y facilita el envío de datos a través de intranets y administrar el almacenamiento a través de largas distancias. El protocolo es muy popular en servidores SAN.

El protocolo iSCSI, es el más utilizado para la transmisión de datos a través de grandes distancias que separaran a los *DataCenters*. Es decir, para la transmisión de datos entre un *DataCenter* ubicado en Europa a otro ubicado en América, el protocolo iSCSI será el responsable de la transmisión a través de la red WAN configurada entre ambos *DataCenters*.

La transmisión de datos entre servidores ubicados en diferentes cuartos e inclusive dentro del mismo cuarto en un centro de datos se realizara por medio de fibra óptica. La fibra óptica actualmente, es el medio de transmisión de datos más veloz que existe en sistemas de comunicaciones y el mismo garantizara

que las peticiones de escritura y lectura en el *DataCenter* sea lo más rápido posible.

5.2. Infraestructura

La infraestructura de un Banco de Datos Virtual consistirá en varios *DataCenters* los cuales estarán distribuidos en diferentes partes del mundo. El mayor recurso de *hardware* presente en los *DataCenters* será el de servidores de almacenamiento especializados. Los servidores de almacenamiento especializados consistirán en su mayoría en *rack's* de servidores SAN/NAS.

La cantidad de servidores SAN/NAS que se tengan instalados en un *DataCenters*, dependerá exclusivamente de las políticas del Banco de Datos Virtual en cuanto a la distribución de su infraestructura de *hardware*. Si cada servidor SAN/NAS tiene capacidad máxima de 20TB de información y se desea que cada centro de datos posea una configuración mínima de capacidad de almacenamiento de 1PB, el centro de datos deberá poseer alrededor de 410 servidores SAN/NAS. La locación física donde se instale el centro de datos deberá ser un edificio el cual posea una infraestructura eléctrica redundante y generadores de energía de *backup* para prevenir el riesgo del suministro de energía eléctrica al edificio y principalmente a los servidores.

La infraestructura física de un centro de datos deberá estar instalada en instalaciones las cuales posean contratos de seguridad policiaca privada para resguardar la integridad de las instalaciones de posibles sabotajes. Las instalaciones y principalmente los cuartos donde se instalen los servidores, deben estar condicionados a temperaturas favorables, principalmente cuando la temperatura del país donde se instale el centro de datos no provea el clima necesario para evitar el sobrecalentamiento de los servidores.

Un *DataCenter* contendrá diferentes cuartos en los cuales se instalaran los servidores y el acceso a cada cuarto debe ser estrictamente restringido únicamente para personal autorizado. Cada entrada a los cuartos de servidores debe ser registrada en una bitácora.

Países como Finlandia, Suecia y Suiza, se ofrecen actualmente como países ideales para la instalación de *DataCenter* por su favorable clima. De ellos, Finlandia es el país que ha atraído a grandes compañías de la industria tecnológica para la instalación de sus *DataCenters*. *Google* y *HP* tienen *DataCenters* instalados en Finlandia. El clima y el precio de la electricidad juegan un papel importante para la selección de la instalación de un *DataCenter* en determinados países. Los costos de un *DataCenter* serán superiores o inferiores según el país donde se instale. La estabilidad política también es importante al momento de seleccionar el país para la instalación de un *DataCenter*, esto debido a que un país con alta inestabilidad política podría incrementar los costos del *DataCenter*.

Idealmente la infraestructura del Banco de Datos Virtual debe poseer distintos *DataCenters* instalados en diferentes países y continentes. En las ideales condiciones, se debe instalar dos *DataCenter* por cada continente.

Tabla IV. **Países preferentes para instalar *DataCenter*. Consultado el 05 marzo de 2012**

Continente	País
Oceanía	Australia
Oceanía	Nueva Zelandia
Asia	China
Asia	Hong Kong
América	EEUU
América	Brasil
Europa	Finlandia
Europa	Suecia
África	Sur África
África	Marruecos

Fuente: <http://www.greenbiz.com/blog/2009/10/12/whats-greenest-country-your-next-data-center>.

5.3. Componentes

La implementación técnica de un Banco de Datos Virtual debe estar compuesta por al menos los siguientes componentes:

5.3.1. Sistema operativo de Cloud-Computing

El sistema operativo de Cloud-Computing, es el responsable de administrar los recursos disponibles del servidor físico y utilizarlos para virtualizarlos, permitiendo compartir estos recursos físicos con las máquinas virtuales. Estas máquinas virtuales en el Banco de Datos Virtual son el interfaz que utilizan los clientes para administrar sus datos. Este interfaz adicional

garantiza que el cliente es el único responsable de la administración de su información.

5.3.2. Plataforma de virtualización

Este componente permite transformar la infraestructura de IT en un Cloud privado, permitiendo ofrecer infraestructura de IT fácilmente como servicio accesible. Para un Banco de Datos Virtual, este componente es el responsable de ofrecer las máquinas virtuales con un determinado espacio de disco para almacenamiento de datos según las necesidades de los clientes.

Las máquinas virtuales podrán tener instalado un sistema operativo Microsoft Windows o alguna distribución de Linux. Tanto la máquina virtual en el Banco de Datos Virtual como la terminal del cliente deberán tener instalado una aplicación P2P que permitirá la sincronización de datos entre ambos equipos.

5.3.3. *DataCenter*

El centro de datos es de los tres componentes, el que tiene mayor importancia debido a que es prácticamente el repositorio de datos de los clientes. La plataforma de paravirtualización virtualiza el espacio de almacenamiento del *DataCenter* y segmenta la capacidad de almacenamiento según las necesidades de los clientes, asigna un tamaño de disco duro para que se pueda almacenar la información en ella.

Este capítulo muestra de forma técnica como se puede construir un Banco de Datos Virtual utilizando diferentes herramientas de software ya existentes en el mercado de IT, así como también diferentes tecnologías de *hardware*.

La robustez de un Banco de Datos Virtual estará limitada por el recurso financiero del que se disponga para financiar su creación, operación y administración.

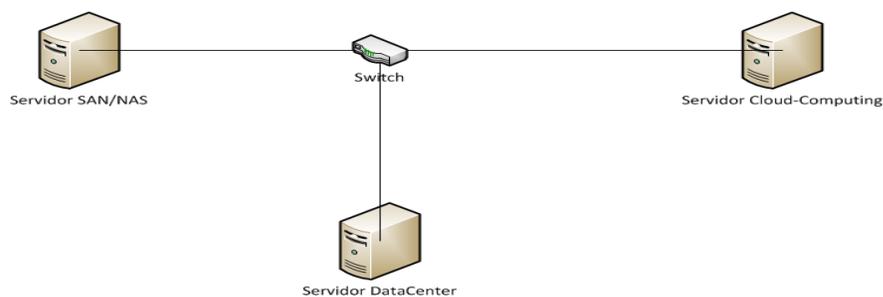
5.4. Implementación

La combinación *hardware* y software a utilizar en la implementación del Banco de Datos Virtual que se detallara a continuación, es una versión ligera del mismo y la combinación de mayor infraestructura de *hardware* ayudaría a una más robusta implementación de un Banco de Datos Virtual.

5.4.1. Hardware

El *hardware* mínimo a utilizar en una implementación para efectos de demostración de un Banco de Datos Virtual y que se describe según cada componente es el siguiente:

Figura 6. **Esquema del *hardware* mínimo de la implementación del Banco de Datos Virtual**



Fuente: elaboración propia.

5.4.1.1. Servidor Cloud-Computing

- Procesador AMD Phenom II 6 núcleos
- 8 GB memoria RAM
- 1 TB disco duro
- Tarjeta de red 1000 Mbit/s

5.4.1.2. Servidor SAN/NAS

- Procesador AMD Phenom II 6 núcleos
- 2 GB memoria RAM
- 2 TB disco duro
- Tarjeta de red 1000 Mbit/s

5.4.1.3. Servidor *DataCenter*

- Procesador AMD Phenom II 6 núcleos
- 8 GB memoria RAM
- 2T disco duro
- Tarjeta de red 1000 Mbit/s

5.4.1.4. Switch Lan

- Switch NetGear GS105E 10/100/1000

5.4.2. Software

El software recomendado para la implementación de un Banco de Datos Virtual, es el desarrollado por la empresa VMWare, Inc. Aun cuando existen otras alternativas en el mercado, VMWare es la única empresa que adicional a brindar un software robusto, también brinda soporte, capacitación en tiempo real para la implementación de sus productos.

VMWare, Inc. es una empresa que provee todo tipo de software el cual ayuda a construir la plataforma de Cloud-Computing. Es la empresa líder en el mercado de plataformas de virtualización. Una gran lista de prestigiosas empresas que confían en sus aplicaciones para la construcción de sus *DataCenters* sirven como impulso para confiar en la utilización de sus aplicaciones para la construcción del Banco de Datos Virtual (VMWare).

El software que se utiliza para la construcción del Banco de Datos Virtual según su función es la siguiente:

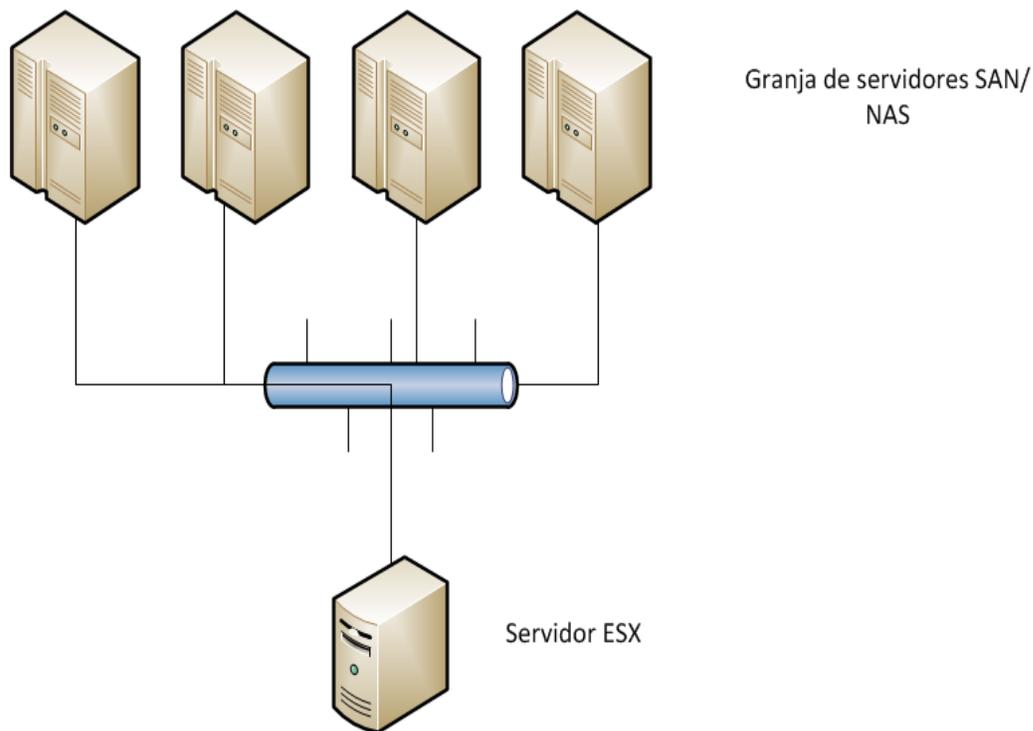
5.4.2.1. VMWare ESXi

VMWare ESXi, es el software núcleo de un *DataCenter* y Cloud-Computing. VMWare ESXi permite en cuestión de minutos instalar la infraestructura de un *DataCenter*. La configuración de un *DataCenter* podrá tomar un par de horas y el tiempo de configuración dependerá de las políticas y el recurso de infraestructura disponible. Sin embargo, el interface de administración de VMWare ESXi está diseñado de tal manera que configurar un *DataCenter* sea un proceso fácil de ejecutar.

Está compuesto de un sistema operativo autónomo que proporciona el entorno de gestión, administración y ejecución al software hipervisor, y los servicios y servidores que permiten la interacción con el software de gestión y administración y las máquinas virtuales (Wikipedia).

Las características sobre servidores compatibles y recomendados para alojar al sistema VMWareESXi se puede consultar en la página oficial <http://www.vmware.com/resources/compatibility/search.php>.

Figura 7. **Servidor ESXi administrando una granja de servidores SAN/NAS en un DataCenter**



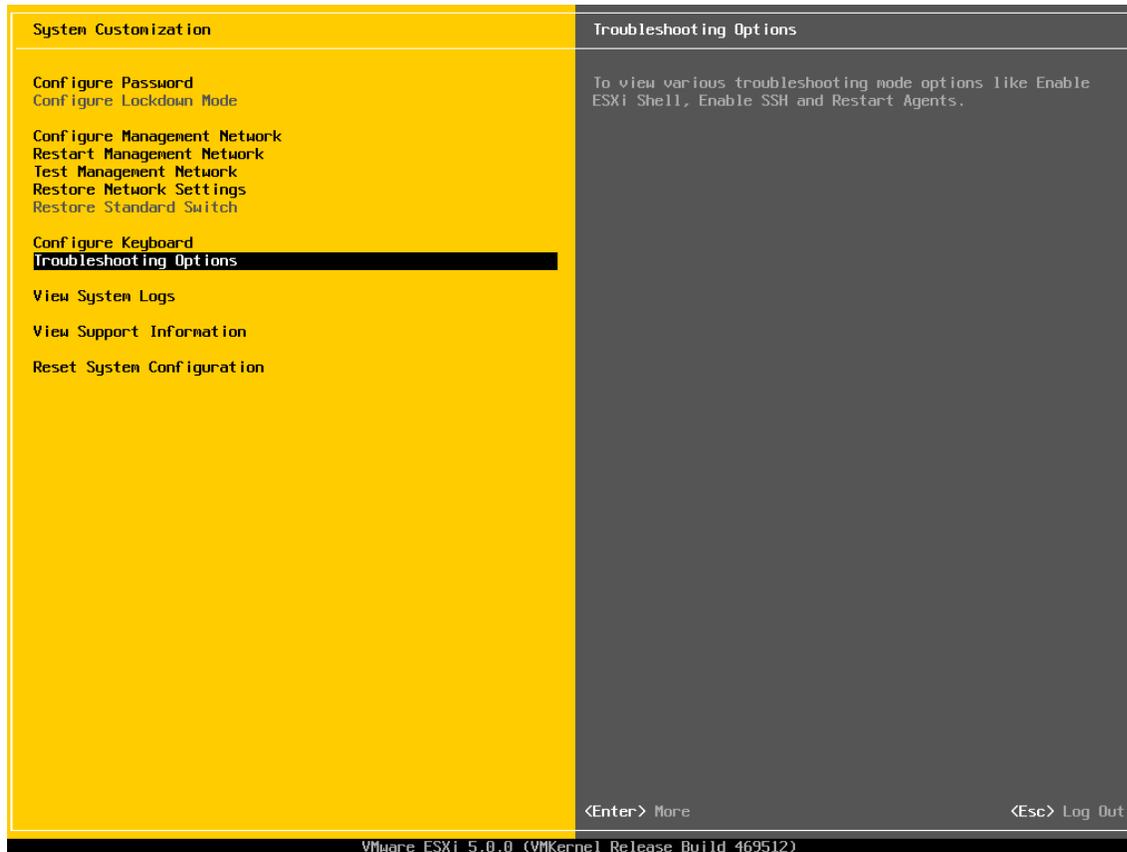
Fuente: elaboración propia.

VMWare ESXi debe ser instalado en un servidor con capacidad de procesamiento robusto. La capacidad de los procesadores y la memoria RAM juegan un papel muy importante, al momento de seleccionar el servidor en el cual se decidirá instalarlo. Debido a que su función principal es la de gestionar y proveer recursos, la capacidad de almacenamiento no es crucial en este servidor, ya que idealmente debe configurarse para que por medio de iSCSI acceda a los recursos del servidor SAN/NAS y se los provea al servidor de Cloud-Computing. Esto evidencia que la inversión para el servidor donde se instalara VMWare ESXi deberá ser alta para que el servicio de Cloud-Computing resultante sea lo suficientemente robusto para atender las peticiones de los clientes al momento de acceder a sus datos.

VMWare ESXi en el proceso de implementación se instala en el servidor *DataCenter* de la figura 6. Es el enlace entre el recurso de almacenamiento disponible en el servidor SAN/NAS y el servidor de Cloud-Computing. Gestiona los recursos del servidor SAN/NAS y los pone a disposición del servidor de Cloud-Computing. Durante el proceso de instalación del servidor VMWare ESXi, encontrara un asistente que lo guiara de forma sencilla durante todo el proceso el cual no toma más de 10 minutos.

Al final de la instalación, únicamente debe configurar la dirección IP por medio de la cual accederá a configurar el servidor, esta dirección IP es la misma que utiliza el servidor VMWare vSphere para conectarse al mismo. El interfaz para realizar todas las configuraciones de la Cloud-Computing privada se deben realizar por medio de VMWare vSphere.

Figura 8. **Consola de administración de VMWare ESXi**

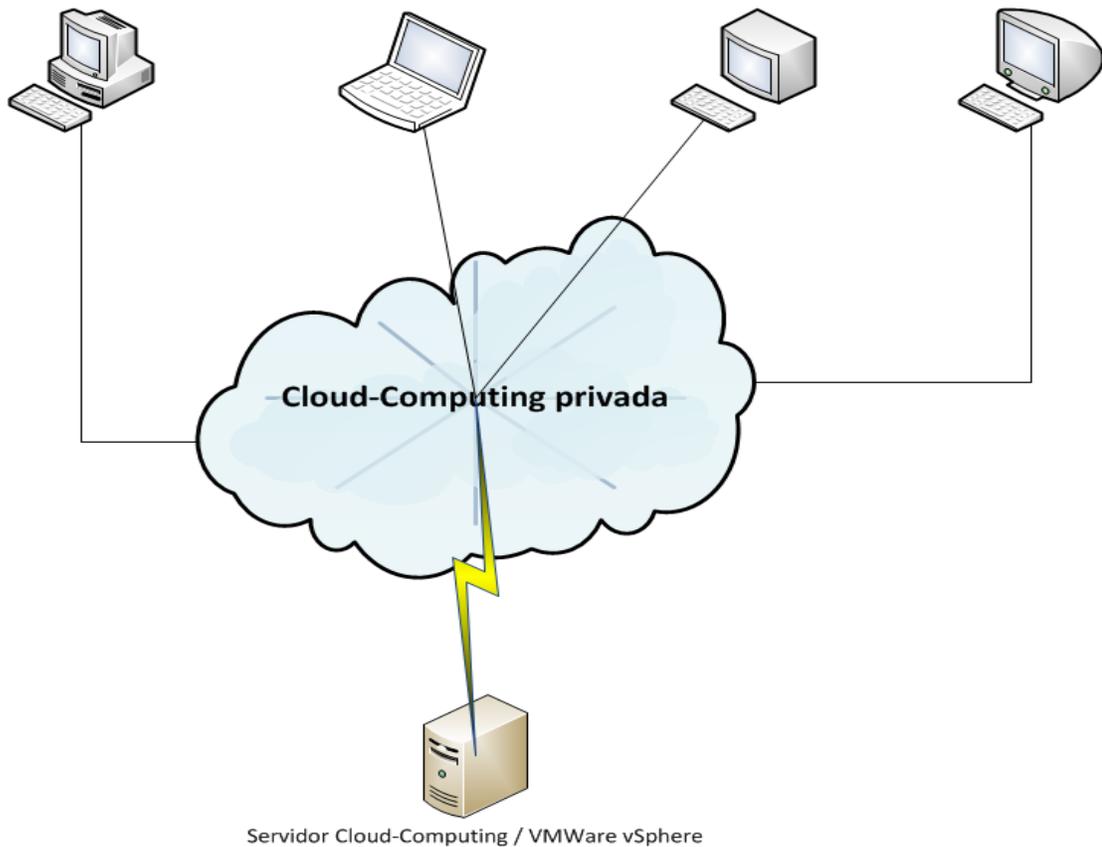


Fuente: imagen de instalación sistema ESXi.

5.4.2.2. **VMWare vSphere**

VMWare vSphere cuyo nombre formal es VMWare Infrastructure 5. Es el primer sistema operativo diseñado para Cloud-Computing. Es la más completa y robusta plataforma de virtualización. La mayor cantidad de sistemas de Cloud-Computing existentes actualmente están creados sobre VMWare vSphere. Su función principal es transformar la infraestructura de IT en una Cloud-Computing privada (VMWare).

Figura 9. **VMWare vSphere convierte una infraestructura de IT en una Cloud-Computing privada**



Fuente: elaboración propia.

Adicionalmente, VMWare vSphere está compuesto por varios componentes los cuales permiten crear la infraestructura de Cloud-Computing a partir de un o unos *DataCenters* ya existentes, los componentes son: ESXi, VMWare vCenter Server, VMWare vCenter Server Heartbeat y VMWare vSphere Client. Cada uno de estos componentes cumple una función específica que permiten incrementar la funcionalidad del Cloud-Computing privada de tal manera que se convierta más robusta con cada componente que se le agregue.

VMWare vSphere al igual que VMWare ESXi debe ser instalado en un servidor con capacidad de procesamiento robusto. Se recomienda una configuración de *hardware* similar en ambos servidores dada la carga de trabajo que deberán desempeñar.

VMWare vSphere, en el proceso de implementación se instala en el servidor Cloud-Computing de la figura 6. Es el enlace de los clientes por medio del interfaz de las máquinas virtuales y sus datos almacenados en el servidor SAN/NAS.

5.4.2.3. OpenFiler

OpenFiler es un sistema operativo basado en Linux que provee archivos utilizando NAS y bloques de datos utilizando SAN. Es un gestor de almacenamiento basado en software (<http://en.wikipedia.org/wiki/Openfiler>).

Una de las principales características de OpenFiler, es que es un sistema libre. La descarga e implementación de OpenFiler es libre y gratuita, mas no lo es el soporte u otros componentes que dan valor agregado al sistema.

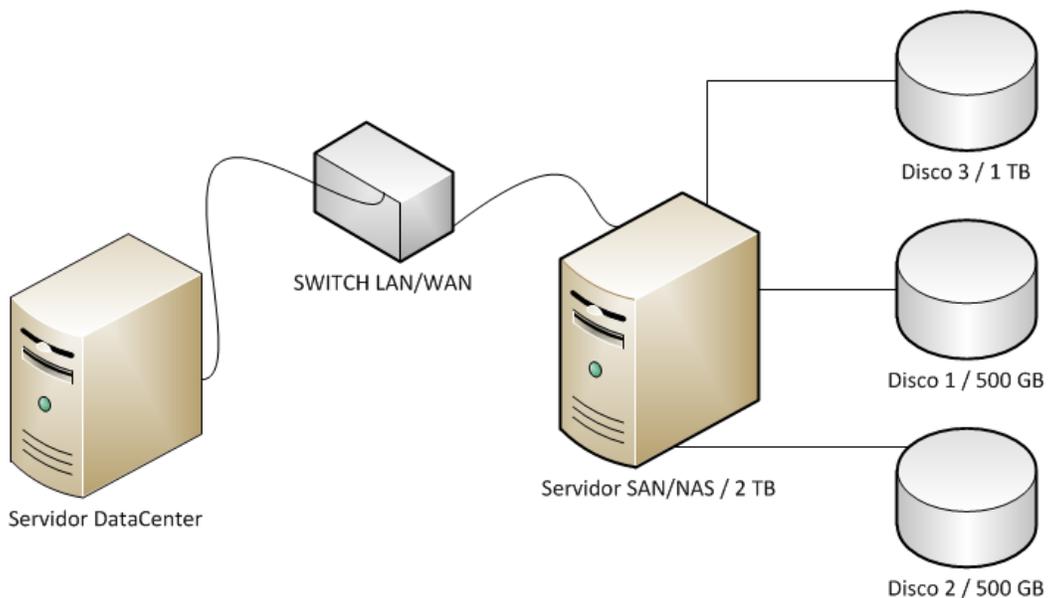
Más que un sustituto de los servidores físicos SAN/NAS, OpenFiler, es una alternativa confiable.

OpenFiler en el proceso de implementación se instala en el servidor SAN/NAS de la figura 6. Es un servidor que contiene diferentes discos los cuales sumandos sus diferentes capacidades de almacenamiento dan un total de 2TB. La información de los clientes del Banco de Datos Virtual, así mismo como las máquinas virtuales en la cual interactúan con su información toman el espacio de almacenamiento de este servidor. Por medio de VMWare vSphere

se configura el servidor SAN/NAS para crear el *DataCenter*. En VMWare vSphere se configura el servidor VMWare ESXi para que tome como principal fuente de almacenamiento de datos o *DataCenter* al servidor SAN/NAS.

Una vez creado el *DataCenter*, al momento de crear una máquina de datos virtual, VMWare vSphere tomara el espacio de almacenamiento por medio de VMWare ESXi y lo depositara en el servidor SAN/NAS. La capacidad de almacenamiento para cada cliente según la capacidad de almacenamiento que solicite se realiza en el servidor VMWare vSphere.

Figura 10. **Servidor SAN/NAS con varios discos de diferentes capacidades de almacenamiento**



Fuente: elaboración propia.

5.4.3. Factibilidad técnica de implementar un Banco de Datos Virtual

El concepto de Cloud-Computing no es nuevo, tampoco lo es un *DataCenter*. Actualmente existen diferentes tipos de recursos con diferentes tipos de proveedores tecnológicos que permiten crear *DataCenters* e infraestructuras de Cloud-Computing. Adicional a esto, la capacidad de almacenamiento de los servidores SAN/NAS se ha incrementado con el transcurso de los años, permitiendo de esta manera ofrecer mayores capacidades de almacenamiento con menores costos. La combinación de estos factores hace factible construir robustos *DataCenters* y emular los servicios de un banco financiero para prestar un servicio de Banco de Datos Virtual.

El apartado 5.4.2 muestra técnicamente como implementar un Banco de Datos Virtual con software y *hardware* existentes en el mercado y con una inversión relativamente accesible. Para efectos de demostración puede recurrirse a lo mostrado en el anterior apartado. Sin embargo, para una solución comercial de un Banco de Datos Virtual, debe considerarse incrementar los costos de inversión para la construcción de *DataCenters* lo suficientemente robustos que puedan en su momento contener la información de millones de usuarios.

Cuando se considere no necesario contar con un *DataCenter* propio, se puede recurrir al arrendamiento de uno. Existen diferentes proveedores de *DataCenters*, los cuales ofrecen *DataCenters* certificados y ubicados en diferentes países en diferentes continentes. En una inversión típica de un banco financiero, el banco utiliza el dinero de sus cuentahabientes para realizar préstamos a terceros o invertir con el fin de ganar intereses. En un Banco de

Datos Virtual, el arrendamiento de un *DataCenter* no debe ser permitido. El arrendamiento de un *DataCenter* expone los datos de clientes a terceros y deja fuera del control de la información digital de los clientes al Banco de Datos Virtual.

Como se ha citado anteriormente, la inversión que requiere un Banco de Datos Virtual es fuerte, el pago de licencias por el software para la construcción de los *DataCenters* en Cloud-Computing no es la mayor inversión de capital. La inversión que consumirá mayores recursos financieros serán los *DataCenters* en sí, dado que se deberá poseer una cantidad considerable de *DataCenters*, el Banco de Datos Virtual necesitara de fuerte inversión financiera en la etapa de la construcción de los *DataCenters*, la cual debe ser absorbida por los inversionistas.

Aun cuando la inversión en la implementación de un Banco de Datos Virtual es enorme y puede de alguna manera obtenerse, vender el concepto a los potenciales clientes es el mayor de los obstáculos que puede afrontarse. La campaña publicitaria para cambiar el paradigma de respaldo de datos tradicional utilizado por la mayoría de las medianas y grandes empresas debe enfocarse en más que prestar un servicio, crear una necesidad. Sin embargo, la implementación de un Banco de Datos Virtual es técnicamente factible con las tecnologías existentes actualmente en el mercado. Un Banco de Datos Virtual ayudaría a muchas empresas a reducir sus costos de respaldo de información.

A un Banco de Datos Virtual también se le podría conocer como un sistema ecológico, ya que permitirá almacenar información en los *DataCenters* y no en medios de almacenamiento físicos que terminan tirándose a la basura.

CONCLUSIONES

1. Implementar un Banco de Datos Virtual requiere gran inversión en infraestructura tecnológica e inmueble, y éste requerimiento se agrava si su fin es para empresas que requieren grandes volúmenes de información. En Guatemala sería difícil la construcción de una granja de servidores para almacenamiento de datos, no solo porque no existe ningún antecedente, sino que además tampoco se presta la situación del país. La forma de instalar uno a corto plazo, sería que una empresa internacional viniese a implementarlo.
2. A pesar de ser una tendencia actual, el Cloud-Computing, no ha tenido auge importante en el país, de hecho no lo ha sido del todo a nivel mundial, pero promete ser la forma de operaciones de IT en las empresas en un futuro no muy lejano.
3. Es evidente que el respaldo de datos en cualquier tipo de empresa es importante, una necesidad inherente a su existencia y utilización de sistemas que automatizan sus procesos y aumenten la productividad y efectividad. Pero si se desarrolla por completo el Cloud-Computing en las empresas y todos los procesos se ejecutan en la nube, esto significa que los datos se encontrarán en los *DataCenters* del proveedor del servicio y por lo tanto, la protección y respaldo de la información correrá por cuenta del proveedor y el servicio de respaldos *online* será innecesario o muy poco utilizado.

4. La reducción de costos utilizando Cloud-Computing, no es una especulación, es un hecho demostrado. Por ello más y más empresas se unen a esta tendencia y adoptan la idea, y no solo reduce costos de servicios y compra de equipo, también es amigable al medio ambiente.
5. La utilización de servidores de almacenamiento en línea y servicios de *backup online* es una idea bastante aceptada y utilizada por muchas personas. Así que el uso de un Banco de Datos Virtual podría resultar en una idea atractiva para las empresas, tomando en cuenta la reducción de costos que esto representa, algo que los gerentes siempre están buscando, sobre todo las pequeñas y medianas empresas.
6. El mecanismo de respaldo de datos de información sigue siendo la utilización de cintas y una estrategia de *backup*. Pero existen mecanismos que permiten proteger los datos y los sistemas como la alta disponibilidad, *failover*, sistemas distribuidos, *mirroring*, entre otros, reduciendo el impacto de la necesidad de respaldo de información.
7. En Guatemala, sería factible la implementación de un Banco de Datos Virtual como servicio de respaldo de información, pero se tendría que buscar una empresa que preste el servicio de almacenamiento en el extranjero; así como la búsqueda de un ISP que maneje un ancho de banda considerablemente alto y haya una alta disponibilidad del servicio de *Internet*.

RECOMENDACIONES

1. A los gerentes de IT, buscar soluciones que permitan la reducción de costos y gastos innecesarios, así como reducir el consumo de energía como una forma de ayudar al medio ambiente.
2. También se les insta a exigir la protección de los datos y garantizar la confidencialidad de la información, evitando lagunas o ambigüedades en los contratos con empresas que presten servicios de almacenamiento de datos.
3. A los emprendedores de negocios informáticos interesados en utilizar la idea de Banco de Datos Virtual como servicio de respaldo de información digital, podrían actuar como intermediarios entre una empresa internacional con la infraestructura necesaria para implementarlo y las empresas locales interesadas en contratar ese tipo de servicio, el hecho de que no haya infraestructura para hacerlo en Guatemala no deja de ser una oportunidad de negocio.
4. A cualquiera que pretenda contratar servicios en la nube, asegurar que el proveedor de *internet* preste un servicio de calidad y que tenga una alta disponibilidad, pues aunque el ISP no forma parte del modelo de Cloud-Computing, es una parte importante por el papel que juega el *internet* al utilizar esta tendencia tecnológica.

BIBLIOGRAFÍA

1. *Amazon Elastic Compute Cloud* [en línea].
<http://aws.amazon.com/es/ec2/>. [Consulta: 05 de marzo de 2012].
2. *Compute Cloud Functionality* [en línea].
<http://aws.amazon.com/es/ec2/#functionality>. [Consulta: 03 de marzo de 2012].
3. Estrategias de *backup* [en línea].
<http://www.spidergim.com/spidergim/bricolaje/store/ESTRATEGIAS%20DE%20BACKUP.pdf>. [Consulta: 03 de marzo de 2012].
4. COCA, Juan R. *Cloud Computing* en TIC [en línea].
http://www.tendencias21.net/EI-cloud-computing-se-consolida-como-modelo-empresarial-para-las-TICs_a3078.html. [Consulta: 03 de marzo de 2012].
5. *Backup* de datos [en línea]. <http://descargar-antivirus-gratis.com/backup-datos.php>. [Consulta: 04 de marzo de 2012].
6. Funziona. *BACKUP* en la Pyme. [en línea].
<http://www.funziona.com.ar/backup.php>. [Consulta: 04 de marzo de 2012].

7. Galicia City. Las unidades de *cintadebackup*. [en línea].
<http://www.galiciacity.com/servicios/hardware/chcinta.htm>.
[Consulta: 04 de marzo de 2012].
8. *Google Apps for Business*. [en línea].
<http://www.google.com/apps/intl/es/business/more.html>.
[Consulta: 05 de febrero de 2012].
9. *Google apps*. [en línea]. http://www.google.com/apps/intl/es-419/business/index.html#utm_source=google&utm_medium=ha&utm_campaign=latamapps-central-bk&utm_content=008&term=google%20apps. [Consulta: 05 de febrero de 2012].
10. *Google Fiber*. [en línea]. <http://www.google.com/appserve/fiberrfi/>.
[Consulta: 05 de febrero de 2012].
11. *Informe de Vigilancia Tecnológica Madrid "Green IT: tecnologías para la eficiencia energética en los sistemas TI"*. Madrid : Fundación madrid para el Conocimiento, 2008, 146 p.
12. Salesforce. *CRM on demand*. [en línea]. <http://www.salesforce.com/>.
[Consulta: 05 de marzo de 2012].
13. MORALES MACHUCA, Carlos Andrés. *Storage area network*. Bogotá, Colombia : Universidad Autónoma de Colombia, 2008, 16 p.

14. Mundo CISCO. *¿Qué es una SAN?* [en línea]. septiembre de 2009.
<http://www.mundocisco.com/2009/09/que-es-una-red-san.html>.
 [Consulta: 04 de febrero de 2012].
15. PRESTON, W. Curtis. *Backup & Recovery*. EEUU: O'Reilly Media, 2007, 768 p.
16. VELTE, Anthony T.; VELTE, Toby J.; EISENPETER, Robert. *Cloud Computing. A practical approach*. EEUU: McGraw-Hill, 2009, 352 p.
18. VMWare, *Company*. [en línea]. <http://www.vmware.com/company/>.
 [Consulta: 08 de marzo de 2012].
19. -----, *Products*. [en línea].
<http://www.vmware.com/products/vsphere/mid-size-and-enterprise-business/overview.html>. [Consulta: 08 de marzo de 2012].
20. *Wikipedia. Alta disponibilidad*. [en línea].
http://es.wikipedia.org/wiki/Alta_disponibilidad. [Consulta: 05 de marzo de 2012].
21. -----, *Ciclo de trabajo*. [en línea].
http://es.wikipedia.org/wiki/Ciclo_de_trabajo. [Consulta: 05 de 03 de 2012].

22. -----*.Openfiler*. [en línea]. <http://en.wikipedia.org/wiki/Openfiler>. [Consulta: 08 de marzo de 2012].
23. -----*.VMware ESX*. [en línea]. http://en.wikipedia.org/wiki/VMware_ESX. [Consulta: 06 de marzo de 2012].
24. -----*.Wikileaks*. [en línea]. <http://en.wikipedia.org/wiki/WikiLeaks>. [Consulta: 05 de marzo de 2012].

6.