

LOS CENTROS DE DATOS EN EL CLOUD COMPUTING

RAFAEL ENRIQUE MONTERROZA BARRIOS

RAFAEL REYES VELÁSQUEZ

MINOR EN COMUNICACIONES Y REDES

PROGRAMA DE INGENIERÍA DE SISTEMAS

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR

CARTAGENA DE INDIAS, D.T. Y C.

2011

LOS CENTROS DE DATOS EN EL CLOUD COMPUTING

RAFAEL ENRIQUE MONTERROZA BARRIOS

RAFAEL REYES VELÁSQUEZ

Trabajo de monografía para obtener el título de Ingeniero de Sistemas

Director:

Ing. Gonzalo Garzón

MINOR EN COMUNICACIONES Y REDES

PROGRAMA DE INGENIERÍA DE SISTEMAS

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR

CARTAGENA DE INDIAS, D.T. Y C.

2011

Nota de aceptación

Jurado

Cartagena, Julio de 2011

Señores

COMITÉ CURRICULAR

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR

Ciudad

Respetados señores:

Con todo el interés me dirijo a Uds. Para presentar a su consideración, estudio y aprobación la monografía titulada **LOS DATA CENTERS EN EL CLOUD COMPUTING**, como requisito para obtener el título de Ingeniero de Sistemas.

Atentamente,

RAFAEL ENRIQUE MONTERROZA BARRIOS

Cartagena, Julio de 2011

Señores

COMITÉ CURRICULAR

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR

Ciudad

Respetados señores:

Con todo el interés me dirijo a Uds. Para presentar a su consideración, estudio y aprobación la monografía titulada **LOS DATA CENTERS EN EL CLOUD COMPUTING**, como requisito para obtener el título de Ingeniero de Sistemas.

Atentamente,

RAFAEL REYES VELÁSQUEZ

Cartagena, Julio de 2011

Señores

COMITÉ CURRICULAR

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR

Ciudad

Respetados señores:

Por medio de la presente me permito hacer entrega de la monografía titulada **LOS DATA CENTERS EN EL CLOUD COMPUTING**, para su estudio y evaluación, la cual fue realizada por los estudiantes RAFAEL ENRIQUE MONTERROZA BARRIOS y RAFAEL REYES VELÁSQUEZ y de la cual acepto ser su director.

Atentamente,

ING. GONZALO GARZÓN

AUTORIZACIÓN

Yo, RAFAEL ENRIQUE MONTERROZA BARRIOS, identificado con la cédula de ciudadanía número 73.575.145 de Cartagena, autorizo a la Universidad Tecnológica de Bolívar, para hacer uso de mi trabajo de monografía y publicarlo en el catálogo on-line de la Biblioteca.

RAFAEL ENRIQUE MONTERROZA BARRIOS

AUTORIZACIÓN

Yo, RAFAEL REYES VELASQUEZ, identificado con la cédula de ciudadanía número 8.870.312 de Cartagena, autorizo a la Universidad Tecnológica de Bolívar, para hacer uso de mi trabajo de monografía y publicarlo en el catálogo on-line de la Biblioteca.

RAFAEL REYES VELÁSQUEZ

RESUMEN

En el desarrollo del presente documento se hace referencia a los conceptos, características y requerimientos de dos tecnologías muy importantes en el mundo tecnológico actual: Cloud Computing y Data Centers.

En el primer capítulo, se hace una revisión general de los conceptos de Cloud Computing y Data Centers. Para el primero, se abarcan su definición, sus características básicas, los modelos de servicio y los modelos de implementación de forma general, para que el lector se forme una idea de esta tecnología. Con respecto a los Data Centers, se presenta su definición, las bases de diseño de estos espacios, los servicios que integran los sistemas de Data Centers, su historia y evolución y su clasificación según el nivel de disponibilidad que ellos pueden brindar.

El segundo capítulo abarca más profundamente todos los conceptos de Cloud Computing, sus requerimientos y características, ventajas y desventajas, definiendo amplia y precisamente cada uno de estos aspectos con el objeto de contextualizar al lector en esta nueva tecnología.

El tercer capítulo hace una recopilación de los conceptos relacionados con los Data Centers; entre ellos encontramos: sus características y requerimientos desarrollados para cada tier correspondiente dentro de la clasificación realizada por The Uptime Institute. Se describe además algunos de los elementos de diseño que diferencian un tier de otro y por último se presenta una solución para llevar un Data Center a lograr un alto nivel de disponibilidad siguiendo pautas no mencionadas en las normas.

El capítulo cuarto hace referencia a describir la interrelación que existe entre los Data Centers y el Cloud Computing con el objetivo de definir el tier mínimo con que se debe diseñar un data center para que sea apto para alojar servicios de Cloud Computing. Adicionalmente se trata el tema de la virtualización en los Data Centers, pieza fundamental en la consecución del objetivo de lograr un Data Center adecuado para las nuevas aplicaciones en la nube. Por último, se desarrolla un modelo de Data Center que cumple con los requerimientos exigidos.

Por último, se presentan las conclusiones y recomendaciones que se desprenden del proceso investigativo y de la recopilación del material bibliográfico y sus aportes para el futuro de este tipo de tecnologías.

TABLA DE CONTENIDO

	Pág.
GLOSARIO	15
INTRODUCCIÓN	21
OBJETIVOS	22
OBJETIVO GENERAL	22
OBJETIVOS ESPECÍFICOS	22
1. CONCEPTOS BÁSICOS DE CLOUD COMPUTING Y DATA CENTERS	23
1.1 CLOUD COMPUTING (COMPUTACIÓN EN LA NUBE)	23
1.1.1 DEFINICIÓN DE CLOUD COMPUTING	23
1.1.2 CARACTERÍSTICAS BÁSICAS	24
1.1.2.1 Auto-Servicio por demanda	24
1.1.2.2 Acceso ubicuo a la red	24
1.1.2.3 Fondo común de recursos	24
1.1.2.4 Rápida elasticidad	24
1.1.2.5 Servicio medido	24
1.1.3 MODELOS DE SERVICIO	25
1.1.4 MODELOS DE IMPLEMENTACIÓN	26
1.1.4.1 Nube privada	26
1.1.4.2 Nube en comunidad	26
1.1.4.3 Nube pública	26
1.1.4.4 Nube híbrida	26
1.2 DATA CENTERS (CENTROS DE DATOS)	26
1.2.1 DEFINICIÓN DE DATA CENTER	27
1.2.2 HISTORIA DE LOS DATA CENTERS	27
1.2.3 BASES DEL DISEÑO DE DATA CENTERS	28
1.2.4 SERVICIOS DE DATA CENTERS	31

1.2.4.1 Servicios de infraestructura IP	31
1.2.4.2 Servicios de aplicación	32
1.2.4.3 Servicios de seguridad	33
1.2.4.4 Servicios de almacenamiento	33
1.2.4.5 Infraestructura de servicios para continuidad del negocio	33
1.2.5 CLASIFICACIÓN DE DATA CENTERS	34
1.2.5.1 Data Center Tier I	35
1.2.5.2 Data Center Tier II	35
1.2.5.3 Data Center Tier III	35
1.2.5.4 Data Center Tier IV	35
2. CARACTERÍSTICAS Y REQUERIMIENTOS DEL CLOUD COMPUTING	38
2.1 CAPAS DEL CLOUD COMPUTING	38
2.1.1 SOFTWARE AS A SERVICE (SaaS)	40
2.1.1.1 Beneficios del modelo SaaS	41
2.1.1.2 Obstáculos a la implementación del modelo SaaS	41
2.1.2 PLATFORM AS A SERVICE (PaaS)	42
2.1.3 INFRAESTRUCTURE AS A SERVICE (IaaS)	43
2.1.4 DATA STORAGE AS A SERVICE (dSaaS)	44
2.2 REQUERIMIENTOS DEL CLOUD COMPUTING	45
2.2.1 ESCALABILIDAD	45
2.2.2 DISPONIBILIDAD	47
2.2.3 UBIQUIDAD	47
3. CARACTERÍSTICAS Y REQUERIMIENTOS DE LOS DATA CENTERS MODERNOS	48
3.1 CLASIFICACIÓN Y CARACTERÍSTICAS DE LOS DATA CENTERS	48
3.1.1 DATA CENTER TIER I: BASICO	48
3.1.2 DATA CENTER TIER II: REDUNDANTE	50
3.1.3 DATA CENTER TIER III: CONCURRENTEMENTE MANTENIBLE	52
3.1.4 DATA CENTER TIER IV: TOLERANTE A FALLAS	54

3.2 SOLUCIONES A LAS EXPECTATIVAS NO CUMPLIBLES ACERCA DE LA DISPONIBILIDAD DE UN DATA CENTER	56
4. MODELAMIENTO PROPUESTO DE UN DATA CENTER ORIENTADO A BRINDAR SERVICIOS DE CLOUD COMPUTING	59
4.1 ARQUITECTURA DEL DATA CENTER	60
4.1.1 CONSIDERACIONES GENERALES	60
4.1.2 UBICACIÓN FÍSICA	61
4.1.3 CONSIDERACIONES ARQUITECTÓNICAS	61
4.2 REQUERIMIENTOS DE TELECOMUNICACIONES E INFRAESTRUCTURA ELÉCTRICA	63
4.2.1 SISTEMAS ELÉCTRICOS EN LOS DATA CENTER	64
4.2.1.1 Acometidas Principales	64
4.2.1.2 Sistemas de UPS	65
4.2.1.3 Sistemas de Tierras	65
4.2.2 CONSIDERACIONES DE DISEÑO A NIVEL ELÉCTRICO	68
4.3 VIRTUALIZACIÓN	68
4.4 COMO PUEDE UNA EMPRESA COLOMBIANA MIGRAR SU DATA CENTER PARA BRINDAR SERVICIOS EN LA NUBE A NIVEL PRIVADO	70
4.4.1 ELECCIÓN DE LOS PROVEEDORES DE ACCESO	71
4.4.2 INFRAESTRUCTURA DE RED	72
CONCLUSIONES Y RECOMENDACIONES	76
REFERENCIAS BIBLIOGRÁFICAS	79

LISTA DE GRÁFICOS

	Pág.
Figura 1. Correlación entre los distintos servicios de data center	30
Figura 2. Mapa conceptual de cloud computing	36
Figura 3. Mapa conceptual de data centers	37
Figura 4. Arquitectura en capas del cloud computing	39
Figura 5. Esquema típico de un sistema eléctrico Tier I	49
Figura 6. Esquema típico de un sistema eléctrico Tier II	52
Figura 7. Esquema típico de un sistema eléctrico Tier III	54
Figura 8. Esquema típico de un sistema eléctrico Tier IV	56
Figura 9. Especificación del aterrizaje de un Data Center	66
Figura 10. Virtualización Tipo 1	69
Figura 11. Virtualización Tipo 2	70
Figura 12. Arquitectura de networking de un data center Tier III	73
Figura 13. Estructura de la capa Core de un Data Center	73
Figura 14. Diagrama lógico de red	74

LISTA DE TABLAS

	Pág.
Tabla 1. Modelos de Servicio Principales de Cloud Computing	25
Tabla 2. Disponibilidad de los Data Center según su nivel	35
Tabla 3. Dimensionamiento del conductor de unión vertical (TBB)	67

GLOSARIO

AAA: el acrónimo AAA corresponde a un tipo de protocolos que realizan tres funciones: Autenticación, Autorización y Contabilización (Authentication, Authorization and Accounting en inglés). La expresión *protocolo AAA* no se refiere pues a un protocolo en particular, sino a una familia de protocolos que ofrecen los tres servicios citados.

BGP: El protocolo de gateway fronterizo (BGP) es un ejemplo de protocolo de gateway exterior (EGP). BGP intercambia información de encaminamiento entre sistemas autónomos a la vez que garantiza una elección de rutas libres de bucles.

CACHING: La más importante aproximación técnica para reducir el retardo y, como resultado, mejorar las prestaciones del sistema, por medio del almacenamiento temporal de los datos frecuentemente accedidos para ganar tiempo.

CLOUD: (Nube), término Usado para referenciar el internet y las interconexiones globales que lo definen como la red de redes.

CLUSTER: Es un conjunto de equipos de cómputo que se basan en hardware y software común especializado que permite que se comporten como si fuese un único equipo.

COMPUTING: Reúne o asocia los conceptos de informática, lógica de coordinación y almacenamiento.

CRM: (CUSTOMER RELATIONSHIP MANAGEMENT) Software para la administración de la relación con los clientes. Sistemas informáticos de apoyo a la

gestión de las relaciones con los clientes, a la venta y al marketing. Con este significado CRM se refiere al sistema que administra un data warehouse (*almacén de datos*) con la información de la gestión de ventas y de los clientes de la empresa.

EIGRP: (Enhanced Interior Gateway Routing Protocol, Protocolo de enrutamiento de gateway interior mejorado) es un protocolo de encaminamiento híbrido, propiedad de Cisco Systems, que ofrece lo mejor de los algoritmos de vector de distancias y del estado de enlace.

ENTRANCE ROOM: Cuarto de entrada. En un data center, es el sitio donde se hace la interconexión del data center con los proveedores de servicios externos como Internet, telefonía y otros servicios externalizados.

EXTRANET: Una extranet es una red privada que utiliza protocolos de Internet, protocolos de comunicación y probablemente infraestructura pública de comunicación para compartir de forma segura parte de la información u operación propia de una organización con proveedores, compradores, socios, clientes o cualquier otro negocio u organización.

FIREWALL: Es un sistema o grupo de sistemas que impone una política de seguridad entre la organización de red privada y el Internet. Es un mecanismo para restringir acceso entre la Internet y la red corporativa interna. Típicamente se instala un firewall en un punto estratégico donde una red (o redes) se conectan a la Internet.

HDA: Horizontal Distribution System. Área de un data center donde se terminan las conexiones de los equipos ubicados en los rack y se hace luego la conexión con el MDA.

HSRP: (Hot Standby Router Protocol). Es un protocolo propiedad de CISCO que permite el despliegue de routers redundantes tolerantes a fallos en una red. Este

protocolo evita la existencia de puntos de fallo únicos en la red mediante técnicas de redundancia y comprobación del estado de los routers.

INTRANET: Es una red de ordenadores privados que utiliza tecnología Internet para compartir dentro de una organización parte de sus sistemas de información y sistemas operacionales.

IT: Tecnologías de la Información, Es el área de la gestión de la tecnología y abarca gran variedad de áreas que incluyen cosas tales como procesos, programas informáticos, sistemas de información, hardware, lenguajes de programación, y las construcciones de datos.

JAULA DE FARADAY: Es una caja metálica que protege lo que se encuentre en su interior de las interferencias electromagnéticas. Se emplean para proteger de descargas eléctricas ya que en su interior el campo eléctrico es nulo.

kPa: Kilo Pascal. Unidad de medida de presión.

MDA: Main Distribution Area. Es el área de un data center donde se encuentran los equipos principales de distribución (core – agregación) hacia las demás áreas del mismo.

MODELO OSI: El modelo de interconexión de sistemas abiertos, también llamado OSI (en inglés *open system interconnection*) es el modelo de red descriptivo creado por la Organización Internacional para la Estandarización. Es decir, es un marco de referencia para la definición de arquitecturas de interconexión de sistemas de comunicaciones.

MULTICAST: Multidifusión (inglés *multicast*) es el envío de la información en una red a múltiples destinos simultáneamente.

NIST: (Instituto Nacional de Estándares y Tecnología de Estados Unidos). Es una agencia federal que trabaja la tecnología con la industria para desarrollar y aplicar la tecnología, medidas y estándares.

OSPF: (Open Shortest Path First). es probablemente el tipo de protocolo IGP más utilizado en grandes redes. Puede operar con seguridad usando MD5 para autenticar a sus puntos antes de realizar nuevas rutas y antes de aceptar avisos de enlace-estado.

PDU: Power Distribution Unit. Es un elemento eléctrico para la conexión de los equipos de cómputo y de red que se ubican en un rack. (Regleta eléctrica).

QOS: Calidad de Servicio (*Quality of Service*, en inglés) son las tecnologías que garantizan la transmisión de cierta cantidad de información en un tiempo dado (*throughput*). Calidad de servicio es la capacidad de dar un buen servicio. Es especialmente importante para ciertas aplicaciones tales como la transmisión de vídeo o voz.

RACK: Es un marco estandarizado o recinto para el montaje de módulos de múltiples equipos de computo.

REST: (REPRESENTATIONAL STATE TRANSFER). La razón fundamental detrás de la arquitectura de diseño Web puede ser descrita por un estilo arquitectónico que consiste en el conjunto de restricciones aplicadas a los elementos dentro de la arquitectura.

SAN: (STORAGE AREA NETWORK). Una red de área de almacenamiento, es una red de almacenamiento dedicada que proporciona acceso a consolidarse, el almacenamiento a nivel de bloque.

SNMP: El Protocolo Simple de Administración de Red o SNMP es un protocolo de la capa de aplicación que facilita el intercambio de información de administración entre dispositivos de red. Es parte de la familia de protocolos TCP/IP. SNMP permite a los administradores supervisar el funcionamiento de la red, buscar y resolver sus problemas, y planear su crecimiento.

SOAP: (*Simple Object Access Protocol*) es un protocolo estándar que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos XML.

SSH: (Secure SHell, en español: intérprete de órdenes segura) es el nombre de un protocolo y del programa que lo implementa, y sirve para acceder a máquinas remotas a través de una red. Permite manejar por completo la computadora mediante un intérprete de comandos, y también puede redirigir el tráfico de X para poder ejecutar programas gráficos si tenemos un Servidor X (en sistemas Unix y Windows) corriendo.

SSL: Secure Socket Layer. Es un protocolo criptográfico que provee seguridad en las comunicaciones a través de Internet.

STP: (Spanning Tree Protocol). Es un protocolo de red de nivel 2 de la capa OSI, (nivel de enlace de datos), Su función es la de gestionar la presencia de bucles en topologías de red debido a la existencia de enlaces redundantes, necesarios en muchos casos para garantizar la disponibilidad de las conexiones.

TIA: (The Telecommunications Industry Association). es la principal asociación comercial que representa el mundo de la información y las comunicaciones las industrias de tecnologías a través de la elaboración de normas , los asuntos de gobierno , las oportunidades de negocios, inteligencia de mercado, la certificación y en todo el mundo cumplimiento de la normativa ambiental.

TIER: Para determinar el nivel de disponibilidad que tiene un Data Center, existe lo que se denomina clasificación Tier, Creada por el Uptime Institute Inc, en el 1995. Este instituto ha establecido cuatro niveles Tier, desde el nivel de menor disponibilidad o Tier I al de mayor disponibilidad o Tier IV.

TUI: (The Uptime Institute). Ofrece educación, publicaciones, consultoría, certificaciones, conferencias y seminarios, la investigación independiente y de liderazgo para la industria de centros de datos empresariales y los datos profesionales de los centros.

VIRTUALIZACIÓN: Se denomina a un conjunto de software y hardware que simula la ejecución de equipos o sistemas operativos distintos a los reales.

VLAN: (*Virtual LAN*, 'Red de Área Local Virtual') es un método de crear redes lógicamente independientes dentro de una misma red física. Varias VLANs pueden coexistir en un único conmutador físico o en una única red física.

VPN: Una red privada virtual, RPV, o VPN de las siglas en ingles de *Virtual Private Network*, es una tecnología de red que permite una extensión de la red local sobre una red pública o no controlada, como por ejemplo Internet.

VRRP: (Virtual Router Redundancy Protocol). Es un protocolo de redundancia diseñado para aumentar la disponibilidad de la puerta de enlace por defecto dando servicio a máquinas en la misma subred.

INTRODUCCIÓN

En Colombia actualmente existen pocas empresas de telecomunicaciones que brinden servicios tipo Cloud Computing y el conocimiento de las características que los Centros de Datos deben tener para brindar este tipo de servicios es escaso. Esta falencia se puede convertir en un factor que retarde la implementación de este tipo de servicios en el país y en la región lo cual también retrasa el avance tecnológico en materia de telecomunicaciones y servicios para los usuarios lo que conlleva una entrada tardía a dicha ola tecnológica.

En Colombia pocas empresas brindan servicios de Data Center y menos aún las hay con la capacidad de expandir sus servicios al nivel del Cloud Computing, por las exigencias que este tipo de servicios requiere.

Con el desarrollo de esta investigación se pretende brindar un punto de partida a nivel de presentación de los estándares mínimos requeridos dirigido tanto a los operadores de telecomunicaciones como a los usuarios de los servicios, para que sepan cuáles son las características y requerimientos mínimos necesarios para la construcción de Data Centers que puedan brindar soluciones de Cloud Computing en Colombia y puedan arrancar con este tipo de proyectos que generarían más empleo y desarrollo tecnológico al país. Por lo tanto, es importante caracterizarlos y la diversificación de servicios para el mejoramiento tanto a nivel de usuarios como de prestadores de servicios.

OBJETIVOS

OBJETIVO GENERAL

Establecer mediante un modelado los parámetros de diseño mínimos requeridos por un Data Center en ambientes de Cloud Computing.

OBJETIVOS ESPECÍFICOS

- ✦ Definir los conceptos de Data Center y Cloud Computing para contextualizar y armonizar el tema con el fin de reconocer sus características.
- ✦ Establecer los requerimientos, características e importancia del Cloud Computing para establecer una caracterización básica tanto en lo funcional como en lo tecnológico.
- ✦ Establecer los requerimientos, características e importancia de los Data Centers para establecer una caracterización básica tanto en lo funcional como en lo tecnológico.
- ✦ Con base en la caracterización de los Data Centers y de Cloud Computing, modelar una solución de Data Center en el ambiente de Cloud Computing.

1. CONCEPTOS BÁSICOS DE CLOUD COMPUTING Y DATA CENTERS

1.1 CLOUD COMPUTING (COMPUTACIÓN EN LA NUBE)

El cloud computing es un concepto que data del año 2006 y que está emergiendo a nivel mundial como una solución tecnológica para muchas empresas y usuarios que quieren contar con servicios de última generación pero no tienen los recursos suficientes para implementar dentro de sus premisas la infraestructura necesaria para desplegarlos dentro de una nueva estrategia de prestación de servicios en la red y su uso globalizado. Es por esto que se aprovecha la Internet y las redes de datos en general para instalar y compartir en ellas software, infraestructura y demás servicios que van a ser accesibles para los usuarios desde cualquier lugar, pagando éstos solo por lo que usan. Este, principalmente, es el concepto clave que denota el cloud computing. Entonces, ¿qué es el cloud computing? Este término se acuña utilizando dos palabras claves de nuestra era informática: “cloud”, que significa “nube” y representa la Internet o red de redes y “computing”, que significa computación y reúne varios conceptos asociados a la tecnología informática, el almacenamiento, entre otros.

1.1.1 DEFINICIÓN DE CLOUD COMPUTING

Si se realiza una búsqueda del término “cloud computing” en un buscador en Internet podremos encontrar muchas definiciones y conceptos expuestos por muchísimos autores a nivel mundial. Sin embargo, al hacer una búsqueda más profunda y eligiendo entidades reconocidas a nivel mundial, podemos elegir como una definición estandarizada y aceptada la expuesta por el Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST, por sus siglas en inglés), que plantea que: “cloud computing es un modelo para habilitar el acceso a un conjunto de servicios computacionales (Ej. Redes, servidores, almacenamiento, aplicaciones y servicios) de manera conveniente y por demanda, que pueden ser rápidamente aprovisionados y liberados con un esfuerzo y una interacción con el proveedor de servicio mínimos” [1]. De la anterior definición, y tal como lo estipuló la entidad en mención, el cloud computing posee 5 características básicas, 3

modelos de servicio principales y 4 modelos de implementación que son los más utilizados.

1.1.2 CARACTERÍSTICAS BÁSICAS

- 1.1.2.1 **Auto-servicio por demanda:** Los servicios pueden ser solicitados por el usuario a través de Internet directamente. Esto implica que el operador de los servicios no interviene en el proceso de suscripción o cancelación del servicio solicitado; además, el usuario solo paga por el tiempo de uso del servicio.
- 1.1.2.2 **Acceso ubicuo a la red:** Los servicios están desplegados en la nube (Internet, Intranet o Extranet) y son accesibles desde cualquier medio con acceso a la red. Además toda infraestructura de cloud computing cuenta con respaldos de toda la información y los servicios ubicados en distintos puntos a nivel mundial, lo que garantiza que siempre estén disponibles aunque uno de los puntos falle totalmente.
- 1.1.2.3 **Fondo común de recursos:** Los servicios se encuentran en la nube para ser utilizados por múltiples usuarios en diferentes lugares del mundo. Esto genera una independencia de la ubicación de los recursos aprovechando la naturaleza de Internet.
- 1.1.2.4 **Rápida elasticidad:** La cantidad o calidad de los servicios ofrecidos en la nube puede aumentar o disminuir rápidamente dependiendo de las necesidades de los usuarios. También esta característica implica que se puede actualizar los servicios y/o recursos computacionales de manera transparente para el usuario.
- 1.1.2.5 **Servicio medido:** Cada recurso que consume un usuario y que deba ser facturado, debe ser medido con exactitud, no solo para la tarificación sino con fines de control. Bajo esta premisa, el proveedor deberá contar con la infraestructura necesaria para que los usuarios puedan ver en tiempo real cual es su consumo de los recursos utilizados y el costo de los mismos.

1.1.3 MODELOS DE SERVICIO

Se han aceptado comúnmente 3 modelos de servicios como los principales en los ambientes cloud computing, aunque existen otros más que son menos utilizados. A continuación presentamos un cuadro ilustrativo de los modelos de servicio de cloud computing y algunos ejemplos significativos [2]:

Tabla 1. Modelos de Servicio Principales de cloud computing

Modelo / Servicio	Definición	Consideraciones	Ejemplos
IaaS Infraestructura como Servicio Infrastructure as a Service	Externalización de recursos como redes, almacenamiento, capacidad de procesamiento y otros recursos fundamentales de TI permitiendo a la organización ejecutar cualquier software desde sistemas operativos hasta aplicaciones. Una de las principales características de este modelo es el uso de la virtualización.	Disponibilidad de los recursos de TI externalizados. Se debe minimizar el impacto ante el evento de una interrupción del proveedor de la nube	<ul style="list-style-type: none"> • Amazon EC2 • Akamai • Microsoft SkyDrive • GoDaddy • Rackspace
PaaS Plataforma como Servicio Platform as a Service	Externalización de las capacidades para desplegar aplicaciones propias o adquiridas, creadas con lenguajes de programación soportados por la plataforma del proveedor disponible en la nube.	<ul style="list-style-type: none"> • Seguridad de la información • Privacidad y responsabilidad legal. • ¿Quién es el dueño de los datos? 	<ul style="list-style-type: none"> • Google App Engine • Microsoft SQL Azure Database • Amazon Simple Queue Service
SaaS Software como Servicio Software as a Service	Externalización del uso de las aplicaciones del proveedor ejecutándose en la nube. Las aplicaciones son accesibles utilizando cualquier medio o cliente.	<ul style="list-style-type: none"> • ¿Quién es el dueño de la aplicación? • ¿Dónde residen las aplicaciones? 	<ul style="list-style-type: none"> • Google Docs • Google Maps • Microsoft Office Live • Yahoo! Mail • Gmail

Los anteriores son los modelos de servicio más importantes en ambientes cloud computing, pero no son los únicos. A continuación se reseñan algunos otros menos utilizados o implementados:

- Almacenamiento como servicio (DaaS – Data storage as a Service)
- Comunicaciones como servicio (CaaS – Communications as a Service)
- Hardware como Servicio (HaaS – Hardware as a Service)

1.1.4 MODELOS DE IMPLEMENTACIÓN

1.1.4.1 Nube privada: La infraestructura de la nube es operada por una única organización. Puede ser administrada por la misma organización o por un tercero y puede ubicarse en las instalaciones de la organización o fuera de ella.

1.1.4.2 Nube en comunidad: La infraestructura de la nube es compartida por varias organizaciones y sus servicios están orientados a soportar las necesidades de una comunidad específica que tiene un interés compartido (ej.: misión, requerimientos de seguridad, políticas y consideraciones de cumplimiento, entre otras). Puede ser administrada por la una o varias organizaciones de la comunidad o por un tercero y puede ubicarse en las instalaciones de la organización o fuera de ella.

1.1.4.3 Nube pública: La infraestructura de la nube está disponible para el público en general o para un segmento muy grande de la industria y es propiedad de una organización que vende los servicios prestados por la nube.

1.1.4.4 Nube híbrida: La infraestructura de la nube está compuesta de dos o más nubes, ya sean estas privadas, públicas o de una comunidad, que permanecen independientes como organización pero que se unen mediante tecnologías estandarizadas o propietarias y permiten la portabilidad de las aplicaciones y los datos entre ellas (ej.: balanceo de cargas entre nubes).

1.2 DATA CENTERS (CENTROS DE DATOS)

Los Data Centers (Centros de Datos) son sistemas complejos que reúnen gran variedad de tecnologías que están en constante evolución. Diseñar y mantener la red de un Data Center requiere habilidades y conocimientos básicos y avanzados en diferentes temas como redes, enrutamiento, seguridad, servidores, aplicaciones, entre otros.

En este aparte se estudiarán los conceptos básicos de los Data Centers, sus características más representativas y algunas de sus aplicaciones más importantes.

1.2.1 DEFINICIÓN DE DATA CENTER

Un Data Center (Centro de Datos) es un sitio con la infraestructura física y tecnológica necesaria para el procesamiento de gran cantidad de información, utilizando como recursos principales servidores interconectados entre sí por redes de datos y con salas debidamente equipadas para su operación [3].

Un Centro de Datos acoge recursos de cómputo de misión crítica en ambientes controlados y bajo una administración centralizada, lo que habilita a las compañías a operar de acuerdo con sus necesidades de negocio [5].

1.2.2 HISTORIA DE LOS DATA CENTERS

Los Data Centers tienen su origen en los enormes centros de cómputo de los primeros años de la industria computacional. Los primeros sistemas de cómputo eran complejos de operar y de mantener y requerían ambientes especiales en los cuales trabajar. Dentro de las características de estos primeros sitios se encontraban bastidores (racks) en los cuales montar los servidores y demás elementos activos, guías de cableado, pisos y techos falsos y los equipos de cómputo de esa época requerían grandes fuentes de energía y sistemas de enfriamiento de alta capacidad para mantenerlos en condiciones óptimas. También se requerían medidas de seguridad avanzadas dado que la gran mayoría de estos centros eran operados por la milicia.

Durante la revolución tecnológica que trajo consigo la era de los microcomputadores, más exactamente en los años 80, las computadoras empezaron a distribuirse en casi todos los campos de la industria y en la mayoría de los casos, no se tenían en cuenta los requerimientos de operación. A medida que las operaciones de IT fueron creciendo, se requería mayor control sobre los recursos de cómputo disponibles. Ya en la década de los 90, durante el nacimiento de la computación cliente-servidor, las microcomputadoras se convirtieron en servidores los cuales debieron ubicarse en los antiguos centros de cómputo y se aprovechó las tecnologías y estándares emergentes de cableado y de equipos de red que ayudaron a que el concepto de Centro de Datos o Data Center, aplicado al diseño de centros de cómputo, empezara a generar buena acogida.

Ya en el año de 2007, el diseño de Data Centers, su construcción y operación se convirtió en una disciplina ampliamente conocida. Se crearon documentos con los estándares para el diseño de Data Centers; dentro de ellos, son reconocidos los promulgados por la Telecommunications Industry Association (TIA), entidad reconocida a nivel mundial en esta materia y quienes mantienen vigentes hasta la fecha todas estos estándares y métricas específicas para el diseño, operación y control de grandes centros de Datos [6].

1.2.3 BASES DEL DISEÑO DE DATA CENTERS

Los Data Centers se crearon principalmente para cumplir con algunos objetivos básicos de los usuarios de los mismos (las empresas) como son: soportar las operaciones de negocio de forma eficiente, disminuir el costo total de operación y mantenimiento y el despliegue rápido de aplicaciones y consolidación de recursos de cómputo. Para cumplir con estas metas de negocio, se determinaron los siguientes elementos básicos de diseño con que todos los Data Centers deben cumplir:

- **Disponibilidad:** Un centro de datos exige una alta disponibilidad, esto es, que desde la etapa de diseño del sistema y su implementación se asegure un cierto grado de continuidad operacional durante un período de medición dado. Disponibilidad se refiere a la habilidad de la comunidad de usuarios finales para acceder a los servicios del Data center, realizar nuevos trabajos, actualizar o alterar datos existentes. Si un usuario no puede acceder al sistema se dice que está no disponible.

- **Escalabilidad:** En un Data Center la escalabilidad se define como, la propiedad del sistema para extender el margen de operaciones, sin perder calidad en los servicios prestados; el poder hacer más grande su capacidad de almacenamiento y procesamiento. La escalabilidad de un Data Center requiere una planeación cuidadosa desde el principio de su desarrollo e implementación.
- **Seguridad:** La seguridad en un centro de datos juega un papel muy importante. Por un lado encontramos la seguridad física, donde el acceso del personal al sitio es usualmente restringido a unos pocos. Para garantizar esto, se dispone de cámaras de video, guardias de seguridad permanentes y además control de incendios, alarmas. Sabemos que algunos de los servidores del Data Center son usados para tareas básicas del personal como uso del Internet, intranets, email. Debido a esto, la seguridad lógica de la información también debe ser celosamente resguardada, haciendo uso de herramientas como, firewalls, vpn y control de acceso para la detección de intrusos en la red.
- **Desempeño:** Es la capacidad que tiene el Data Center de poder responder a las actividades demandadas por los usuarios finales, como tiempos de respuesta, capacidad de almacenamiento, conectividad ininterrumpida, accesos remotos, velocidad de descarga y subida de información.
- **Administración:** Posibilidad de ejecutar tareas como configurar, administrar y monitorear los niveles de servicio del Data Center, crear y mantener archivos y estructuras de directorio, instalación y actualización de software de aplicación cuando se requiera y sean solicitadas por el cliente.

Estos elementos o criterios básicos de diseño se aplican a todas las áreas funcionales o servicios de Data Centers, entre las que se cuentan:

- **Servicios de infraestructura:** enrutamiento, conmutación, arquitectura de granjas de servidores, etc.
- **Servicios de aplicación:** Balanceo de carga, caching, independización del proceso de manejo de SSL, etc.

- **Servicios de seguridad:** filtrado e inspección de paquetes, detección de intrusos, prevención de intrusos.
- **Servicios de almacenamiento:** Arquitectura SAN, copias de seguridad.
- **Continuidad de negocio:** selección de sitio, interconexión de Data Centers, replicación de datos.

Existen estándares bien definidos para el diseño e implementación de centros de datos para que estos cumplan con unas características físicas básicas. Entre dichas características podemos mencionar:

- Debe contar con doble acometida eléctrica
- Accesos anchos preferiblemente con puertas corredizas
- Altura suficiente entre piso y techo
- Medidas de seguridad apropiadas en caso de incendios o desastres naturales
- Aire acondicionado con control de temperatura
- Pisos y techos falsos para la acometida de cables
- Generadores y cuartos de distribución eléctrica
- Control de acceso físico a las áreas sensibles y restringidas
- Alarmas y sistemas de video vigilancia

Debe contar con un almacén para los equipos y demás elementos que no se encuentren en uso en un momento determinado [4].

1.2.4 SERVICIOS DE DATA CENTERS

Los servicios de Data Center no solo se relacionan unos con otros sino que también, en algunos casos, dependen unos de otros.

Figura 1. Correlación entre los distintos servicios de data center

Infraestructura de Continuidad del Negocio	
Seguridad del Data Center	
Infraestructura de Aplicaciones	
Infraestructura de Almacenamiento	Infraestructura de Red (IP)

La figura anterior muestra la correlación entre los distintos servicios de Data Center; la infraestructura de red y la de almacenamiento son los pilares sobre los que se edifican los demás servicios y proveen los fundamentos de la red y, por ende, de cualquier otro servicio de la arquitectura. Una vez tenemos la infraestructura básica operativa, se puede implementar granjas de servidores para soportar las aplicaciones. Toda esta infraestructura debe protegerse mediante los servicios de seguridad, tanto física como lógica. Finalmente debemos contar con planes de contingencia ante posibles fallos para poder garantizar altos niveles de redundancia que garanticen la continuidad de la operación de los servicios soportados por el Data Center.

1.2.4.1 Servicios de infraestructura IP: En esta categoría se incluyen todas las características necesarias para que la infraestructura de red del Data Center entre en operación. La infraestructura IP se organiza de la siguiente manera:

- Capa 2
- Capa 3
- Servicios inteligentes de Red

Las características de la capa 2 proveen la conectividad entre granjas de servidores y entre los dispositivos que prestan los servicios (Ej.: VLANs); también facilitan el acceso al medio físico, deben proveer una convergencia rápida de la red, eliminación de bucles de cableado

(mediante STP). Debe garantizarse que la convergencia STP sea rápida y escalable.

La capa 3 debe proveer rápida convergencia a nivel del enrutamiento de la red, redundancia de rutas por defecto. El objetivo principal de esta capa es mantener alta disponibilidad de los servicios de red. Debe soportar enrutamiento estático, dinámico (Ej. BGP, OSPF, EIGRP), redundancia de puertas de enlace predeterminadas mediante protocolos como HSRP o VRRP.

Por último, los servicios inteligentes de red abarcan características que permiten que los servicios de aplicación operen por toda la red eficientemente. Entre los más comunes tenemos QoS (Calidad de Servicio) y Multicast. Además, existen otros también importantes como son: VLANs privadas, enrutamiento basado en políticas (policy-based routing)

1.2.4.2 Servicios de Aplicación: incluyen características que mejoran la operación de las aplicaciones en toda la red y usan inteligencia de red para optimizar el entorno de las aplicaciones. Dentro de estas características podemos encontrar: Balanceo de carga (load balancing), caching, terminación de SSL, entre otros.

El balanceo de carga permite que las peticiones hacia los servidores sean distribuidas de manera equitativa entre todos los servidores que hacen parte de la granja y permiten también monitorear el estado de los servidores para garantizar alta disponibilidad de los servicios.

Normalmente, el balanceo de carga ocurre en las capas 4 y 5 del modelo OSI. Los mecanismos de balanceo de carga deben permitir la adición o sustracción de servidores en tiempo real para permitir la flexibilidad de la operación sin interrupciones.

El caching ofrece la capacidad de brindar contenido estático al usuario que lo requiere directamente desde un repositorio específicamente diseñado para almacenar copias de esta información que se encuentra en los servidores y así disminuir la carga de procesamiento y tráfico que se dirige a los servidores primarios.

Por último, los mecanismos de terminación SSL permiten que el procesamiento de las solicitudes SSL en el Data Center sean realizadas por dispositivos especialmente diseñados y dedicados a este propósito. Todo el proceso de encriptación y des encriptación así como el manejo de las llaves y certificados digitales es realizado fuera de la granja de servidores y así éstos no tendrán esta carga adicional de procesamiento para brindar los servicios.

1.2.4.3 Servicios de seguridad: Los mecanismos de seguridad incluyen tanto la seguridad lógica como la seguridad física del centro de cómputo. Dentro de las tecnologías a usar a nivel lógico podemos encontrar:

- Listas de control de acceso
- Firewalls
- Sistemas de detección de intrusos
- Administración segura (SNMP v3, SSH, AAA, etc.)

1.2.4.4 Servicios de almacenamiento: aquí se incluyen las capacidades de consolidación de discos conectados directamente a los servidores, mediante la utilización de arreglos de discos conectados directamente a la red. De esta manera se provee un mejor aprovechamiento de la capacidad implementada y se centraliza el acceso a la información. Dos servicios adicionales relacionados con este concepto son la capacidad de consolidar distintas SANs más pequeñas e independientes (SAN: Storage Area Network) en una SAN más grande y robusta y la capacidad de virtualizar el almacenamiento de manera que varios servidores puedan utilizar el mismo conjunto de arreglos de discos de manera simultánea.

La consolidación de SANs aisladas en una única SAN centralizada se logra mediante la combinación de tecnologías como VSAN y Switches SAN avanzados.

1.2.4.5 Infraestructura de servicios para continuidad del negocio: mediante los mecanismos provistos por esta infraestructura se logra dotar a los

centros de datos de la robustez necesaria para brindar alta disponibilidad mediante la redundancia de los demás servicios soportados.

La implementación de la infraestructura de continuidad de negocio se basa en tres tecnologías fundamentales:

- **Selección de sitio:** mecanismo que permite la detección automática de las fallas en la red a nivel de aplicación y que estas reaccionen adecuadamente redireccionando las peticiones a otra granja de servidores o sitio disponible.
- **Extensión de la SAN:** esto se refiere al proceso de replicar una SAN existente en una ubicación diferente, ya sea en el mismo Data Center o en otra localización geográfica, con el objetivo de replicar los datos y toda la información almacenada de manera simultánea y automatizada entre los dos sitios.
- **Interconexión de Data Centers:** son alternativas de comunicación y conectividad entre los distintos sitios y extensiones de SAN que especifican como debe realizarse la comunicación entre dichos elementos.

1.2.5 CLASIFICACIÓN DE DATA CENTERS

A medida que los Data Centers fueron evolucionando, fue necesaria una estandarización de los parámetros de diseño y operación de los mismos, tendiente a diferenciarlos plenamente entre sí y a proveer un mecanismo para determinar los niveles de disponibilidad dependiendo de las capacidades de cada nivel. Es por esto que The Uptime Institute, Inc. (TUI) se dio a la tarea de recopilar en lo que ellos llamaron Tiers (niveles) los distintos estados en la evolución de los Data Centers durante los últimos 40 años [7].

Durante los últimos 40 años, los data centers han evolucionado pasando a través de al menos 4 estados diferentes, los cuales componen el sistema de clasificación de TUI. El primer nivel (Tier I) hizo su aparición a principio de la década de 1960; Tier II, en la década de 1970; Tier III a finales de la década de 1980 y Tier IV en el

año de 1994 con el proyecto United Parcel Service Winward [7] que fue el primer Data Center que utilizó equipos de cómputo con fuentes de energía duales. El TUI participó activamente en el desarrollo del nivel III de Data Centers y fue pionero en la creación del Tier IV.

A continuación se presentan las características más relevantes de cada Tier:

1.2.5.1 Data Center Tier I: Es considerado el Data Center Básico y consta solo de una única fuente de distribución de energía, sin componentes redundantes.

1.2.5.2 Data Center Tier II: cuenta con una sola fuente de distribución de energía y un único sistema de enfriamiento, pero cuenta con algunos componentes redundantes en su infraestructura.

1.2.5.3 Data Center Tier III: cuenta con múltiples sistemas de distribución de energía y múltiples sistemas de enfriamiento, pero solo uno de ellos está activo durante la operación. Posee componentes redundantes en toda la infraestructura y es posible hacer mantenimiento en caliente.

1.2.5.4 Data Center Tier IV: está compuesto de múltiples sistemas de distribución de energía y sistemas de enfriamiento, todos activos al mismo tiempo, con redundancia total en los componentes de la infraestructura y es totalmente tolerante a fallos.

A continuación se presentan los niveles de disponibilidad que se pueden garantizar en cada nivel de Data Center:

Tabla 2. Disponibilidad de los Data Center según su nivel

Tier	% Disponibilidad	% Parada	Tiempo de parada al año
Tier I	99.671%	0.329%	28.82 horas
Tier II	99.741%	0.251%	22.68 horas
Tier III	99.982%	0.018%	1.57 horas
Tier IV	99.995%	0.005%	52.56 minutos

De esta tabla se desprende que para las operaciones de misión crítica para los clientes, es necesario invertir en Data Centers que se puedan clasificar mínimo en Tier III, dada la alta disponibilidad que se puede ofrecer a partir de dicho nivel.

Para resumir el contenido del presente capítulo se presentarán mapas conceptuales de los dos términos tratados en el mismo.

Figura 2. Mapa Conceptual de Cloud Computing

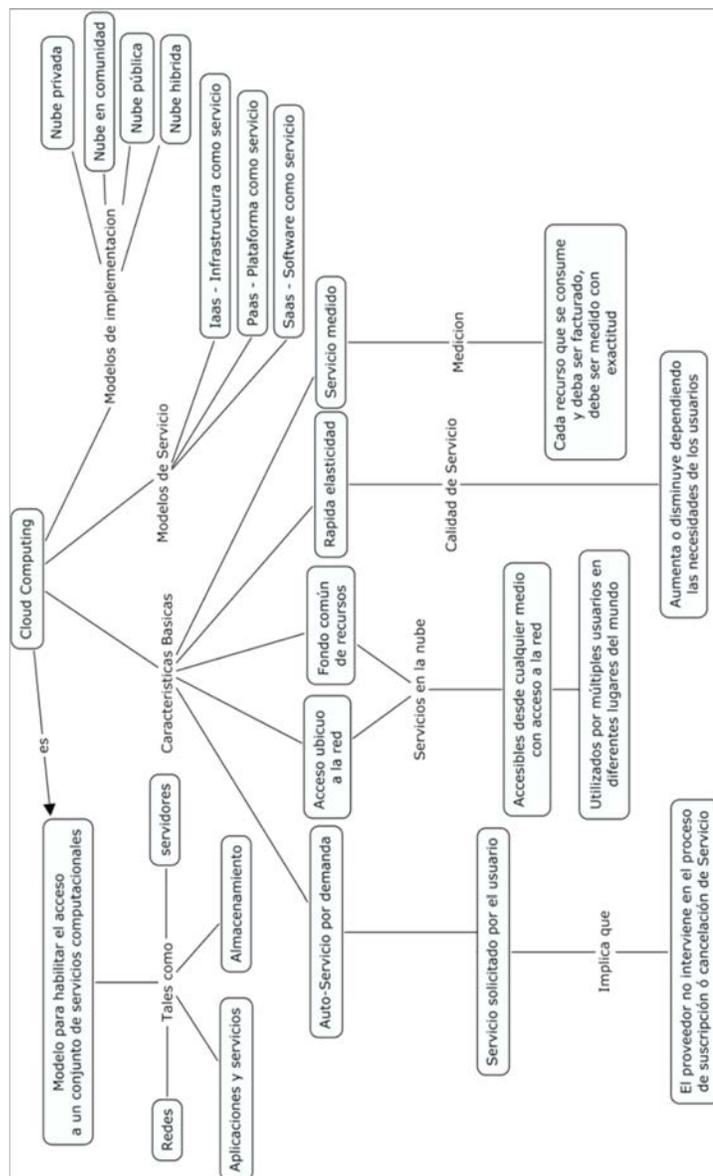
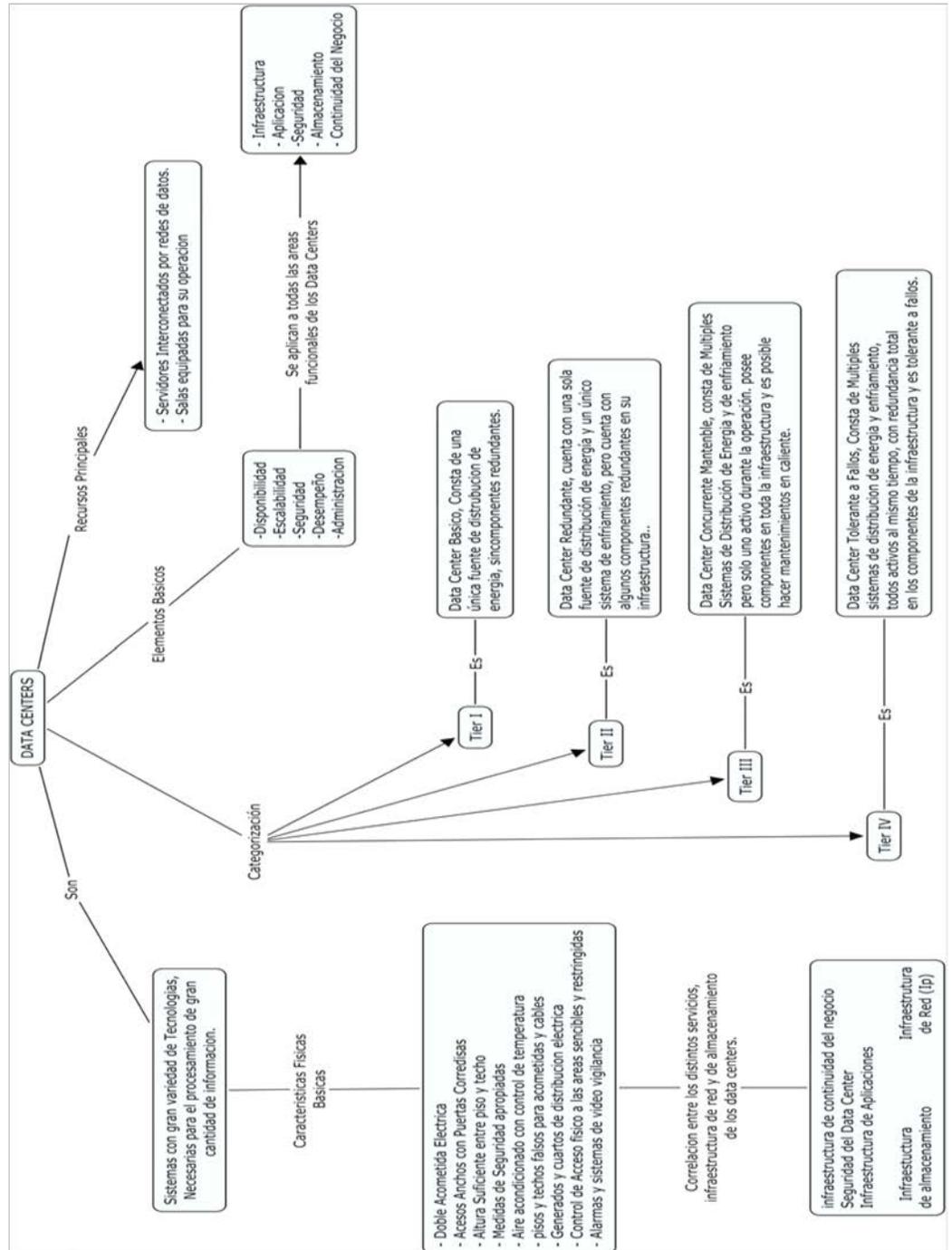


Figura 6. Mapa Conceptual de Data Centers



2. CARACTERÍSTICAS Y REQUERIMIENTOS DE CLOUD COMPUTING

Como se estudió en el capítulo anterior, el cloud computing posee unas características básicas y una clasificación bien definidas lo cual permite que se pueda determinar los requerimientos y características avanzadas que ayudarán a modelar los tipos de data centers adecuados que puedan suplir dichos requerimientos.

A continuación se profundizará en los conceptos fundamentales del Cloud computing así como en sus características y requerimientos para así complementar todo el conocimiento que se requiere para modelar este tipo de sistemas desde la perspectiva de la conectividad de los mismos.

2.1 CAPAS DEL CLOUD COMPUTING

El Cloud Computing se puede visualizar como una colección de servicios que se pueden presentar como una arquitectura en capas, como se puede observar en la figura 4.

Los servicios ofrecidos mediante Cloud Computing conforman lo que se conoce como SaaS, y se encuentra en la capa más alta de la pila. Esta capa es la que interactúa directamente con el usuario final de los servicios.

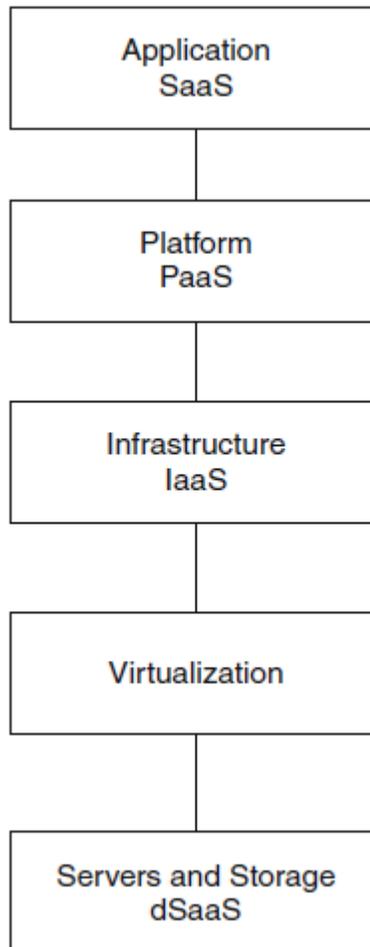
Platform as a Service (PaaS) se refiere a recursos de cómputo como servicio, incluyendo tanto recursos de hardware como los sistemas operativos y las aplicaciones de soporte a los servicios de la primera capa; es la capa sobre la que se implementan los servicios de la capa inicial. Se puede pensar en PaaS como IaaS con un conjunto de aplicaciones de software y de sistemas operativos adaptados a los requerimientos de las aplicaciones específicas que se brindan a los usuarios.

Más abajo en el esquema de capas encontramos Infrastructure as a Service (IaaS), que no es más que los recursos físicos y/o lógicos (virtualización) de hardware con capacidad de procesamiento, ancho de banda y almacenamiento garantizados para suplir los requerimientos de las capas superiores.

La virtualización hace parte de la IaaS y es uno de los pilares de las tecnologías de cloud computing debido a que con ella se logra aprovechar al máximo los recursos físicos existentes y se pueden brindar más cantidad de servicios minimizando los costos.

Por último encontramos la capa de dSaaS (data-storage as a service) que brinda todo el soporte de almacenamiento de datos tanto de los usuarios como de las capas intermedias [8].

Figura 4. Arquitectura en capas del cloud computing



2.1.1 SOFTWARE AS A SERVICE (SaaS)

SaaS es el modelo en el cual una aplicación está alojada como un servicio para los clientes, quienes acceden a él a través de Internet o de una Intranet/Extranet. De esta manera se provee acceso a las aplicaciones sin que el cliente tenga que estar pendiente de su mantenimiento y/o actualización.

El proveedor del sistema realiza todo el mantenimiento necesario a las aplicaciones para mantener las mismas en óptimas condiciones operativas y adicionalmente siempre mantiene las aplicaciones actualizadas a la última versión disponible, sin que estos cambios causen traumatismos a los usuarios.

A nivel de costos, el modelo de SaaS brinda a los clientes la flexibilidad de pagar solo por el uso que le da a la aplicación, pudiéndose medir este uso en tiempo, en capacidad de almacenamiento, entre otros. Para los proveedores de estas aplicaciones, el modelo SaaS le brinda la capacidad de mantener segura la propiedad intelectual de sus creaciones y además de mantener un flujo constante de dinero por el uso que los clientes le dan al servicio.

A continuación se presentan algunos tipos de aplicaciones típicas que se pueden desplegar bajo el modelo SaaS:

- CRM (Customer Relationship Management)
- Sistemas de Video conferencia
- Administración de servicios de IT (Helpdesk)
- Aplicaciones contables
- Administradores de contenido Web

El tipo de aplicaciones más propensas a desplegarse bajo el modelo SaaS incluyen aquellas con poca o ninguna interacción con otros sistemas o aplicaciones, sistemas que realizan tareas sencillas.

Las aplicaciones SaaS difieren de aquellas implementadas bajo el modelo distribuido en que las primeras están orientadas al acceso bajo ambiente web utilizando solo un navegador de Internet. Esto permite el acceso a aplicaciones de terceros, comerciales o de software libre, desde cualquier parte donde se tenga una conexión a la red [9].

2.1.1.1 Beneficios del modelo SaaS: Uno de los más grandes beneficios ofrecidos por el modelo SaaS es el costo del servicio, ya que normalmente es más económico acceder a una aplicación mediante SaaS que adquirir la misma aplicación para uso interno en una organización. Podemos encontrar, además, otros beneficios de este modelo entre los que podemos mencionar los siguientes:

- **Familiaridad con la Web:** los clientes que acceden a estos servicios están muy familiarizados con la navegación web y el uso de Internet. Por esto, la curva de aprendizaje de este tipo de aplicaciones se hace bastante pequeña.
- **Equipo de IT más pequeño:** no se requiere mantener en la organización mucho personal dedicado a IT ya que las aplicaciones y los servidores están alojados externamente, lo que trae consigo ahorros en salarios y en espacio físico.
- **Personalización:** las aplicaciones SaaS son fácilmente personalizables por el cliente, a diferencia de las antiguas aplicaciones internas lo que permite obtener exactamente lo deseado por la organización.
- **Mercadeo global:** Un proveedor de aplicaciones bajo el modelo de venta directa, normalmente debe gastar muchos recursos en mercadeo y publicidad para hacer llegar su aplicación a su mercado objetivo. SaaS, en cambio, solo con la publicación del servicio en la Web, este es accesible a todo el mundo y se da a conocer muy rápidamente.
- **Seguridad:** las aplicaciones SaaS normalmente son desplegadas al cliente utilizando mecanismos de acceso seguro como SSL, mecanismos que son bastante confiables y altamente conocidos por los usuarios de Internet.

2.1.1.2 Obstáculos a la implementación del modelo SaaS: Como todo, el modelo SaaS puede presentar obstáculos para su implementación y uso. El primero de todos podría ser el hecho que muchas organizaciones requieren plataformas computacionales y aplicaciones muy específicas que podrían no encontrarse o podrían no poderse implementar bajo el modelo SaaS. Debido a esto, estas organizaciones se decidirán por comprar directamente el software necesario del fabricante, en lugar de acceder al mismo en la red.

Otro factor que obstaculiza la implementación de SaaS es el hecho que al adquirir un servicio de acceso a una aplicación de un proveedor, normalmente no es posible migrar los datos ni la propia aplicación hacia otro proveedor o al tratar de hacerlo, el proveedor original cobrará un recargo por hacer el cambio.

Por último, el software libre y las plataformas de hardware abierto son un reto a superar por parte de SaaS ya que muchas organizaciones optan por estas alternativas que pueden salir muy económicas, e incluso, gratis de implementar y ofrecen un desempeño muy alto para las necesidades de la organización.

2.1.2 PLATFORM AS A SERVICE (PaaS)

PaaS, como se vio en el aparte anterior, provee todos los recursos que se requieren para la construcción de aplicaciones y servicios completamente desde Internet, sin necesidad de descargar o instalar ningún software.

Dentro de esta capa encontramos sistemas de diseño y desarrollo de software, servicios de hosting de aplicaciones web, etc. PaaS normalmente provee soporte para la creación de interfaces de usuario usando HTML y/o javascript.

Algunas de las tecnologías que se pueden encontrar en PaaS incluyen: SOAP (Simple Object Access Protocol), REST (Representational State Transfer).

PaaS se encuentra implementado en uno de tres grandes sistemas distintos, como son:

- **Sistemas de desarrollo de módulos agregables (Add-ins):** Esto permite que las aplicaciones SaaS ya existentes puedan extenderse y complementarse o personalizarse mediante módulos.
- **Ambientes independientes:** Estos ambientes no incluyen dependencias técnicas, financieras o de licenciamiento específicas para una aplicación o aplicaciones SaaS y son usados para propósitos generales.
- **Ambientes de solo distribución de aplicaciones:** Se encuentran sistemas de soporte al alojamiento de aplicaciones como sistemas de seguridad para hosting, sistemas de escalamiento por demanda, entre otros [10].

2.1.3 INFRASTRUCTURE AS A SERVICE (IaaS)

Esta capa se encarga de proveer tanto los recursos de software y de hardware de bajo nivel para el soporte de las capas superiores, que constan principalmente de aplicaciones y servicios. Esta capa podría dividirse en dos subcapas que son: Sistemas Operativos y Hardware.

Un concepto clave que hace parte de IaaS es la virtualización, tanto para proveer sistemas operativos diferentes bajo una misma plataforma de hardware o para proveer máquinas virtuales sin ningún sistema operativo para que el cliente pueda instalar el que desee sin ninguna restricción.

Otro mecanismo que podemos encontrar para ofrecer servicios de IaaS es el HaaS, Hardware as a Service, en el que los operadores arriendan equipos físicos protegidos por todos los mecanismos de seguridad física, lógica y de respaldo eléctrico necesarios para operar ininterrumpidamente, y el cliente puede instalar en ellos lo que requiera para su operación. Con este mecanismo los clientes evitan tener que comprar equipos, racks, equipos de red y cableado para comunicarse con estos equipos dado que el proveedor en la nube lo hace por él.

Bajo el modelo IaaS, se logra brindar servicios de hardware a las capas superiores de la arquitectura de cloud computing, así como también brindar estos servicios directamente a los usuarios. Entre los servicios consumibles por los usuarios se encuentran:

- Espacio para alojar servidores
- Equipos de red
- Servidores virtuales
- Servidores físicos
- Espacio de almacenamiento

Entre los recursos necesarios para implementar IaaS tenemos:

- **Equipos de cómputo:** incluye servidores y sistemas de almacenamiento en red.
- **Red:** incluye equipos como firewalls, balanceadores de carga, switches routers y todo lo necesario para la interconexión física y lógica de los servidores y equipos. Es la base de los data centers pues ofrece la infraestructura de comunicaciones necesaria para toda la operación.

- **Conectividad a Internet:** recurso fundamental para poder distribuir los servicios a los clientes o usuarios.
- **Plataformas de virtualización:** Permite a los clientes y los administradores del sistema aprovechar los recursos físicos al máximo, pudiendo operar varias máquinas virtuales soportadas por una única máquina física de alta capacidad.
- **Sistemas de contabilización y facturación:** estos sistemas son importantes para poder medir los recursos utilizados por los clientes y poder realizar el cobro correspondiente por los mismos [11].

2.1.4 DATA STORAGE AS A SERVICE (dSaaS)

Mediante este modelo, el proveedor ofrece servicios de almacenamiento en sus servidores para el uso de los clientes, los cuales no pueden adquirir sus propios sistemas de almacenamiento por los altos costos de los mismos. Este modelo también brinda todo el soporte de almacenamiento de datos y de aplicaciones para las capas superiores de la arquitectura en capas del Cloud computing.

En esta capa se ofrecen servicios muy importantes para el mantenimiento de los datos como son replicación, copias de seguridad, recuperación ante desastres.

Una de las ventajas más grande de dSaaS es el ahorro de costos para el cliente. El cliente solo paga por la capacidad ya sea por uso (datos almacenados), por disponibilidad (capacidad total contratada) o por datos transferidos y no tiene que pagar costos por infraestructura.

Uno de los principales usos de este modelo es la implementación de sistemas de copia de seguridad para los clientes, quienes cuentan con la capacidad contratada para asegurar sus datos locales y evitar la pérdida de los mismos.

Para implementar este tipo de sistemas se puede hacer uso de tecnologías de almacenamiento como SAN (Storage Area Network) que consiste en una red especialmente diseñada para interconectar dispositivos de almacenamiento en cluster, que se acceden remotamente desde los dispositivos que requieren

acceder a los datos. Estas redes se implementan con gran velocidad de transmisión, a través de enlaces de fibra óptica, y se utilizan sistemas de almacenamiento de alto desempeño, normalmente basados en discos SCSI de alta velocidad de giro, aunque también se utilizan arreglos de discos tipo SATA, sistemas de cinta, entre otros. Los sistemas SAN no ofrecen al sistema que los accede el concepto de archivo sino que implementan acceso directo por bloques. El sistema operativo que accede a la SAN deberá saber implementar los mecanismos de acceso de alto nivel para el almacenamiento de la información.

2.2 REQUERIMIENTOS DEL CLOUD COMPUTING

El cloud computing, a nivel funcional, requiere que la infraestructura subyacente cumpla con ciertos requerimientos mínimos que garanticen la operación y el funcionamiento apropiados y óptimos de los servicios que la nube presta, tanto en ambientes privados como públicos.

Podemos mencionar como requerimientos del cloud computing los siguientes:

- **Escalabilidad:** se requiere que la infraestructura física sobre la que opera la nube pueda ser escalada (o reducida) sin afectar los servicios prestados por la nube. Debe haber suficiente flexibilidad para que la infraestructura crezca o decrezca de acuerdo con la carga real del sistema.
- **Disponibilidad:** se requiere que los servicios prestados en la nube se mantengan operativos con el mínimo de interrupciones.
- **Ubiquidad:** es la capacidad de que los servicios prestados por la nube puedan ser accedidos desde cualquier parte del mundo, sin que esto implique una degradación en la experiencia de interacción del cliente con dichos servicios.
- **Seguridad:** Los servicios prestados a los clientes deben garantizar un alto nivel de seguridad y confidencialidad de la información administrada por los mismos.

2.2.1 ESCALABILIDAD

También denominado flexibilidad, es la capacidad que debe tener la infraestructura de cloud computing para crecer o decrecer al ritmo de la demanda de los servicios y/o de la carga sobre los mismos. Se puede hablar de tres tipos básicos de escalabilidad:

- **Escalabilidad de carga:** se refiere a cuando un sistema tiene la capacidad de hacer un buen uso de los recursos disponibles bajo diferentes niveles de carga sobre los servicios. Por ejemplo, evitando consumos de energía improductivos.
- **Escalabilidad de espacio:** el sistema tiene la habilidad de mantener un consumo de recursos del sistema como capacidad de memoria o de procesamiento en niveles moderados bajo altos niveles de carga. Por ejemplo, el uso de memoria virtual por el sistema operativo para evitar que la memoria principal se use al límite. También se puede tratar en este tipo de escalabilidad la posibilidad de agregar capacidad de almacenamiento en caliente sobre los sistemas para permitir el crecimiento en dicha área sin afectar los servicios.
- **Escalabilidad estructural:** se refiere a cuando un sistema se acoge a los estándares para permitir un crecimiento ordenado en las capacidades de almacenamiento o tratamiento de información. Por ejemplo, el tipo de datos escogido para una aplicación puede influir en la cantidad de datos que la aplicación puede manejar.

Podemos realizar escalamiento de los sistemas de dos maneras:

- **Escalamiento vertical:** cuando se agrega más poder de cómputo a la infraestructura (Ej. más procesador, más memoria, más ancho de banda).
- **Escalamiento horizontal:** cuando se agregan software o hardware como un todo a la infraestructura, por ejemplo, al agregar más servidores a un cluster o a un sistema de balanceo de cargas o agregar otra instancia de una base de datos para balancear las cargas.

Para el cloud computing, la escalabilidad es sumamente importante debido a que una de sus características fundamentales es permitir el auto aprovisionamiento de los servicios por parte de los usuarios, y se requiere que cuando los usuarios requieran los servicios, haya recursos disponibles en la infraestructura para poderlos desplegar.

2.2.2 DISPONIBILIDAD

También denominado estabilidad, se refiere a la minimización del tiempo en que los servicios suministrados por la nube no están disponibles para los usuarios. Se requiere que los servicios estén disponibles 24x7x365 independientemente de las condiciones de operación.

Este concepto se relaciona con la tolerancia a fallas, que es la capacidad del sistema de recuperarse automáticamente y en el menor tiempo posible de cualquier fallo que se presente en cualquier elemento de la infraestructura, tanto si es a nivel físico en los equipos como si es a nivel del proveedor de servicios externos (ej. conectividad a Internet).

Para obtener altos niveles de disponibilidad se requiere mantener redundancia total en todo el sistema, que garantice la continuidad de la operación incluso en condiciones críticas.

2.2.3 UBIQUIDAD

Uno de los requerimientos más importante del cloud computing es que los servicios que se prestan bajo este modelo sean alcanzables y utilizables desde cualquier lugar del mundo, sin que el usuario perciba ninguna diferencia cuando lo utiliza, digamos, en Japón y luego en Colombia. Esta característica del cloud computing se logra haciendo replicación de toda la infraestructura de soporte de las nubes en diferentes partes para así lograr que el acceso regional sea de la misma calidad en todas partes.

Adicionalmente, la ubicuidad brinda la posibilidad de personalizar los servicios según la región en que se brindan, con total portabilidad de los datos.

3. CARACTERÍSTICAS Y REQUERIMIENTOS DE LOS DATA CENTERS MODERNOS

Debido a que los Data Centers alojan en su interior recursos de cómputo de misión crítica, requieren que sean diseñados siguiendo parámetros especiales tanto a nivel del espacio físico como a nivel de la infraestructura tecnológica; así mismo se requiere especial cuidado a la hora de escoger el personal que va a atender y a brindar sus servicios en el Data Center.

A continuación se presentarán los requerimientos, características y clasificación de los Data Centers modernos y las principales aplicaciones que se pueden alojar en los mismos.

3.1 CLASIFICACIÓN Y CARACTERÍSTICAS DE LOS DATA CENTERS

The Uptime Institute (TUI), estableció una clasificación de los Data Centers en lo que ellos llamaron Tiers (niveles) que recopiló la evolución de los mismos desde su aparición hasta nuestros días, teniendo como criterio principal de clasificación la disponibilidad ofrecida en cada Tier. TUI determinó 4 Tiers para enmarcar en ellos los distintos tipos de Data Centers existentes y cada Tier posee unas características y requerimientos (criterios de diseño) propios que dan a los Data Centers diseñados bajo dichos criterios el nivel de disponibilidad que lo hace pertenecer a cada nivel.

3.1.1 DATA CENTER TIER I: BÁSICO

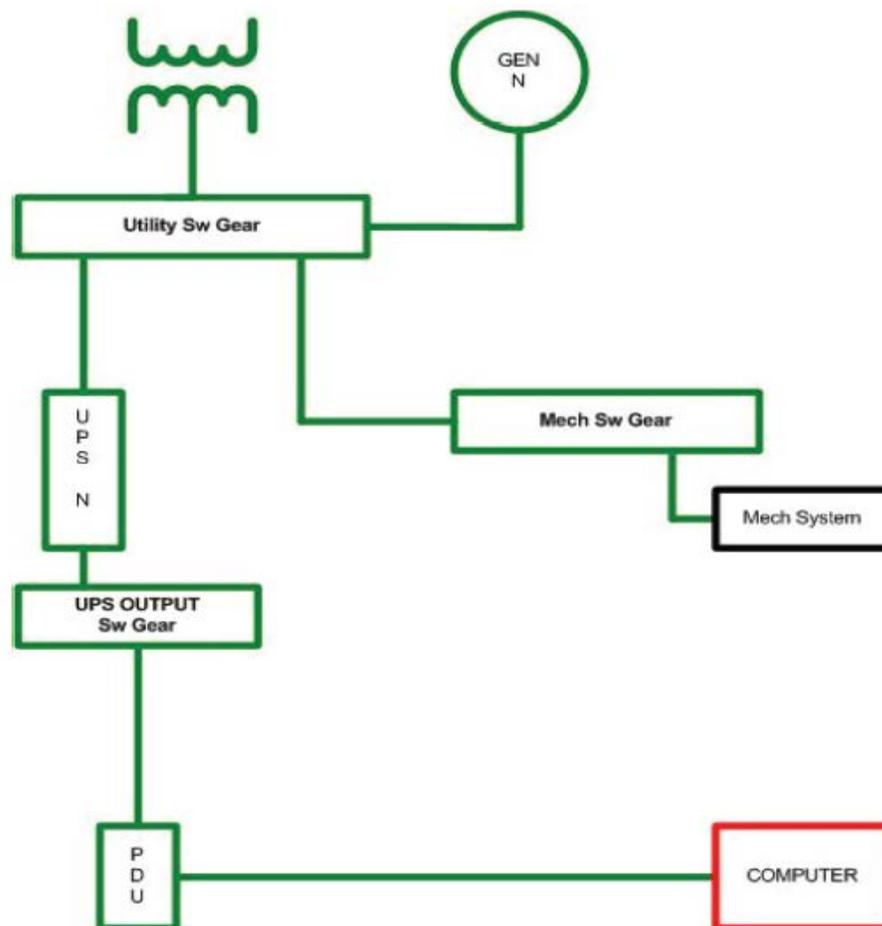
Un Data Center Tier I es susceptible de sufrir interrupciones, tanto planeadas como inesperadas. Posee sistemas de distribución eléctrica y de refrigeración pero puede que no tengan pisos falsos, UPS o generadores alternos de energía. En caso de tener generadores de energía o UPS, estos sistemas son simples y poseen muchos puntos de falla.

Este tipo de Data Centers debe ser apagado por completo para hacer cualquier mantenimiento mayor o una reparación de cualquier sistema que lo componga. Cuando se presentan situaciones críticas con frecuencia, son muy susceptibles de tener que apagarse en dichos casos. Cualquier error o daño en alguno de los sistemas causará una interrupción en el servicio.

El nivel de disponibilidad que un Data Center Tier I puede garantizar es de solo 99.671% [7]. La carga crítica de un Tier I, puede llegar al 100% de N (capacidad total).

Uno de los aspectos que más incide en la disponibilidad de un Data Center o de cualquier infraestructura tecnológica es el sistema eléctrico del cual se alimentan los equipos. Si este sistema es débil, todos los demás elementos que componen la infraestructura fallarán inexorablemente. A continuación se muestra un diagrama esquemático de un sistema eléctrico típico de un Data Center Tier I.

Figura 5. Esquema típico de un sistema eléctrico Tier I



De este esquema se puede destacar que una falla que ocurra en cualquiera de los elementos del mismo provocará una caída de toda la infraestructura del Data Center porque no existe ningún tipo de redundancia que garantice la continuidad de la operación.

A nivel de la planta física, arquitectónicamente hablando, los Data Center Tier I no cuentan con protecciones contra eventos físicos, naturales o intencionales; por ejemplo, normalmente están ubicados en edificaciones sin sismo resistencia, con sistemas de control de incendios deficientes. Esto se debe a que la aplicación de este tipo de Data Centers es para empresas pequeñas que trabajan con infraestructuras tecnológicas que no requieren de redundancia ni de calidad de servicio.

En el campo de las redes y telecomunicaciones, Tier I solo posee un proveedor de servicio y una sola ruta de cableado interno. Los equipos activos de red no cuentan con elementos redundantes. Cualquier falla en alguno de estos componentes de red afectará la operación del Data Center de manera negativa.

Por último, Con respecto a los componentes mecánicos de la infraestructura de Data Centers Tier I, estos cuentan con sistemas de aire acondicionado, en ocasiones múltiples, pero sin ninguna redundancia. Las tuberías solo siguen una única ruta de distribución y no cuentan con protecciones ante eventos naturales y/o intencionales.

Por sus características, los data center Tier I son aplicables para negocios pequeños en que su infraestructura de TI esté orientada a satisfacer solo necesidades internas, no brindar servicios al público en general. Son compañías que basan su negocio en Internet, pero que no requieren calidad de servicio [15].

3.1.2 DATA CENTER TIER II: REDUNDANTE

Los Data Center Tier II son sistemas con componentes redundantes, lo que los hace menos susceptibles a interrupciones tanto planeadas como no planeadas. Poseen pisos falsos, UPS y generadores de energía alterna pero su capacidad, por diseño, es N+1 lo que significa que poseen una sola ruta de distribución de cableado. El mantenimiento de cualquier elemento crítico del sistema requerirá una caída en la operación.

El nivel de disponibilidad que un Data Center Tier II puede garantizar es de solo 99.741% [7]. La carga crítica de un Tier II, puede llegar al 100% de N (capacidad total).

Tal como se explicó para los Data Center Tier I, se establecerán las características de Tier II en los aspectos eléctrico, arquitectónico, mecánico y de telecomunicaciones.

A nivel arquitectónico, un data center tier II tiene en cuenta algunos aspectos de seguridad física concernientes a protección contra eventos naturales o intencionales, pero a nivel mínimo. Un aspecto en que mejoran con respecto a Tier I es que poseen sistemas de control de acceso al área física del data center mediante puertas de seguridad.

Con respecto a los sistemas de telecomunicaciones, un data center tier II cuenta con equipos críticos redundantes, con fuentes de poder duales, multiprocesadores y también cuentan con más de un proveedor de servicio de Internet, lo que le brinda un nivel básico de redundancia, pero no automatizada.

A nivel eléctrico, Tier II cuenta con UPS redundante de diseño N+1, generadores redundantes, PDUs redundantes y preferiblemente alimentados de UPS separadas. Los gabinetes deben contar con dos circuitos eléctricos dedicados de 20A/120V. Por último, deben contar con un sistema EPO (Emergency Power-Off System). A continuación se presenta un diagrama esquemático de un sistema eléctrico típico de un data center Tier II.

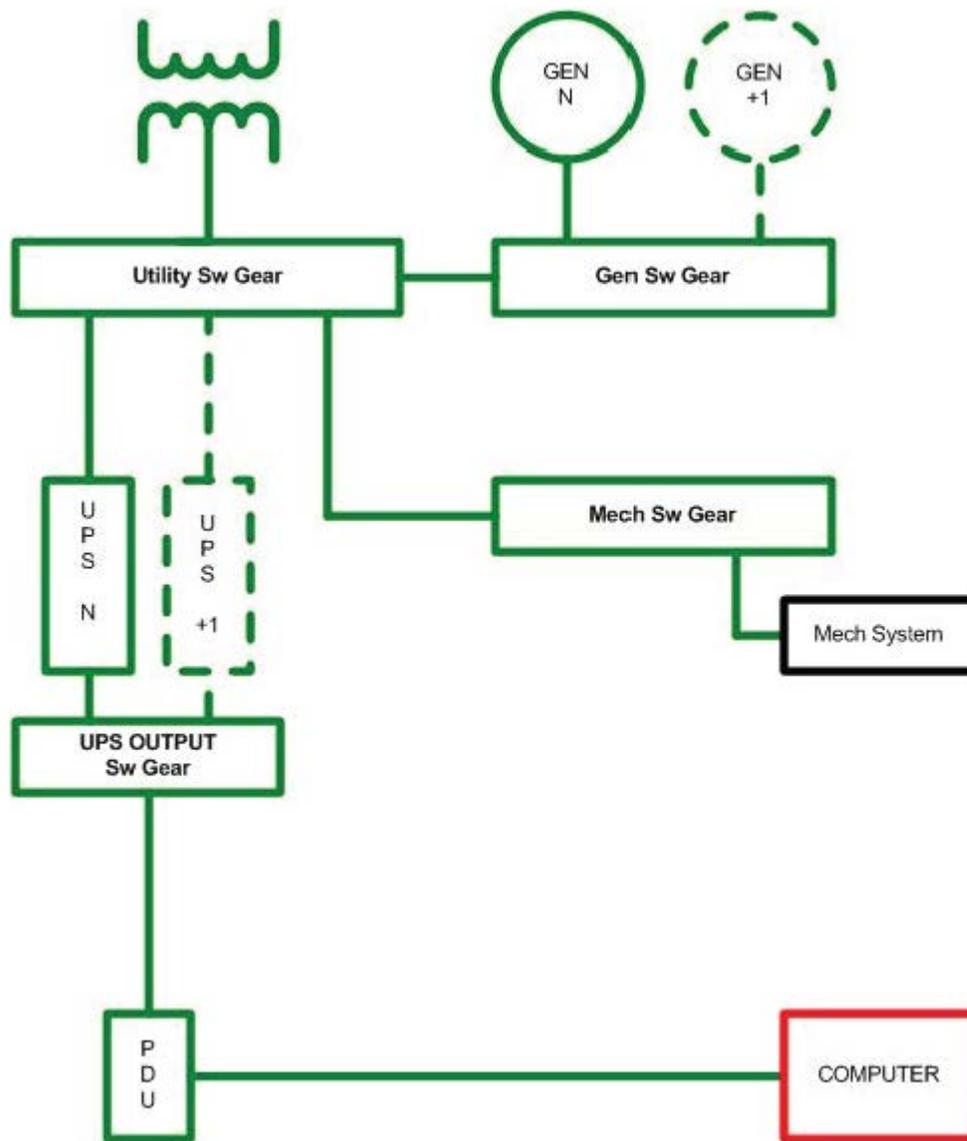
Con respecto a las condiciones mecánicas de un data center Tier II, estos normalmente cuentan con sistemas de enfriamiento que combinan control de temperatura y humedad y que operan 24x7x365.

Los Data Center Tier II cumplen con todas las especificaciones de los Tier I y las superan brindando mejores niveles de desempeño y seguridad que garantizar lograr los niveles de disponibilidad característicos de este tipo de espacios

Los Data Center Tier II siguen siendo aceptables para empresas de tamaño pequeño, con operación limitada a horarios normales de trabajo; compañías de software que no brindan servicios on-line o real-time

A continuación se presenta un esquema de un diseño de sistema eléctrico típico de un Data Center Tier II, donde se evidencian los sistemas redundantes característicos.

Figura 6. Esquema típico de un sistema eléctrico Tier II



3.1.3 DATA CENTER TIER III: CONCURRENTEMENTE MANTENIBLE

Los Data Centers clasificados en este nivel permiten la operación continua de toda la infraestructura cuando se hacen mantenimientos planificados. Dentro de las actividades programadas que soporta este nivel de Data Centers sin interrumpir su operación se cuentan mantenimientos preventivos y correctivos programados,

reparación y reemplazo de componentes de hardware, adición o eliminación de componentes para ampliar o reducir la capacidad de operación, pruebas de componentes y sistemas, entre otros.

Para poder lograr el mantenimiento concurrente de los sistemas que lo conforman, se debe asegurar suficiente capacidad en cada uno de los sistemas redundantes para operar sin alterar el rendimiento mientras el sistema alterno se encuentra fuera de línea.

Estos data centers, a pesar de contar con un alto nivel de disponibilidad, aún es susceptible a interrupciones por eventos no planeados (Ej. daños, caídas de los proveedores). La disponibilidad esperada de este tipo de Data Centers es de 99.982%. La carga crítica de un Tier I, puede llegar al 90% de N (capacidad total) [7].

En el aspecto arquitectónico, una mejora de los data centers Tier III con respecto de los niveles inferiores es la implementación de sistemas de seguridad perimetral, control de acceso electrónico, muros exteriores sin ventanas para evitar el acceso no autorizado a las instalaciones.

A nivel de telecomunicaciones, los data center Tier III cuentan con dos o más proveedores de servicio, dos cuartos de entrada de servicio, rutas y áreas redundantes, con elementos activos configurados para tolerar fallos en enlaces y fallas en equipos.

En cuanto al aspecto mecánico de las instalaciones, los data centers tier III cuentan con múltiples unidades de aire acondicionado con sistemas de control de temperatura y humedad de alta precisión; poseen tuberías y bombas duales y cuentan con sistemas de detección de derrames.

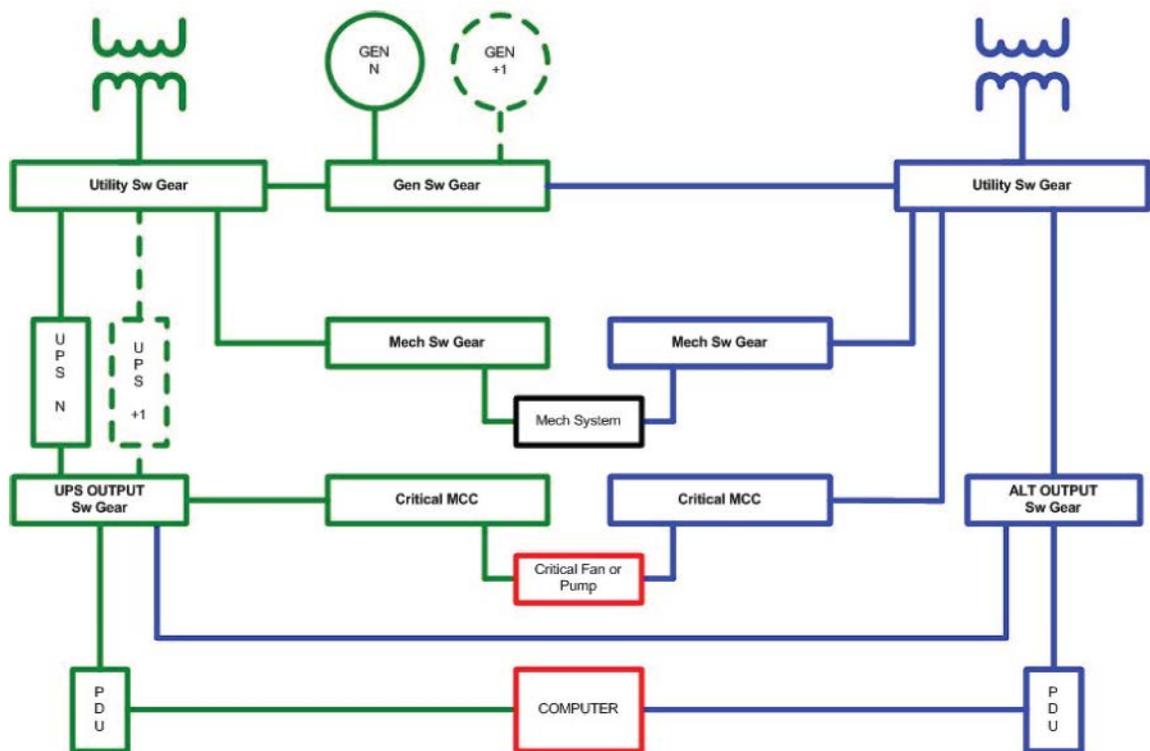
Por último, en cuanto a la infraestructura eléctrica, poseen como mínimo redundancia N+1 en generadores eléctricos, UPS y PDUs; cuentan con dos vías de distribución, una activa y una alterna; sistemas de aterrizaje y de protección para alumbrado; sistemas de control y monitoreo para gestionar la mayoría de los equipos eléctricos; servidor de monitoreo redundante para asegurar control y gestión continuos de la infraestructura eléctrica.

Los data center tier III están orientados para compañías que brindan servicios 24x7 como centros de servicio e información; también para compañías donde los recursos de TI deben brindar soporte a procesos automatizados y sin interrupciones; otro tipo de compañías son aquellas que cuentan con sucursales

en distintas partes del mundo, por lo que su personal accederá a los recursos de TI en distintos horarios.

A continuación se presenta un diagrama de la infraestructura eléctrica típica de un data center Tier III

Figura 7. Esquema típico de un sistema eléctrico Tier III



3.1.4 DATA CENTER TIER IV: TOLERANTE A FALLAS

Un data center Tier IV ofrece capacidad en su infraestructura para permitir cualquier actividad planeada sobre la misma sin interrupción de la carga crítica. Las funcionalidades de tolerancia a fallas permiten que, además, la infraestructura pueda sostenerse en, al menos, un caso de fuerza mayor, evento no planeado o en cualquier evento que no afecte la carga crítica.

Esta característica de tolerancia a fallas requiere de sistemas totalmente redundantes y activos de forma simultánea; esto significa que cada subsistema contará con un diseño N+1, lo cual generará un diseño global tipo 2(N+1).

La disponibilidad que un data center Tier IV puede ofrecer es de 99.995% y su carga crítica combinada global no superará el 90% de N [7].

Debido a la reglamentación sobre seguridad eléctrica, estos data centers deberán tener interrupciones en casos críticos como la activación de la alarma de incendios o cuando se active una EPO (Emergency Power-Off).

A nivel arquitectónico, un data center Tier IV está ubicado en un edificio con protecciones contra accidentes naturales como terremotos, inundaciones, huracanes, entre otros. Siempre se encuentra en un edificio independiente dedicado solo a la operación del data center y normalmente se encuentran alejados de lugares públicos susceptibles de ataques terroristas o accidentes graves (aeropuertos, líneas férreas).

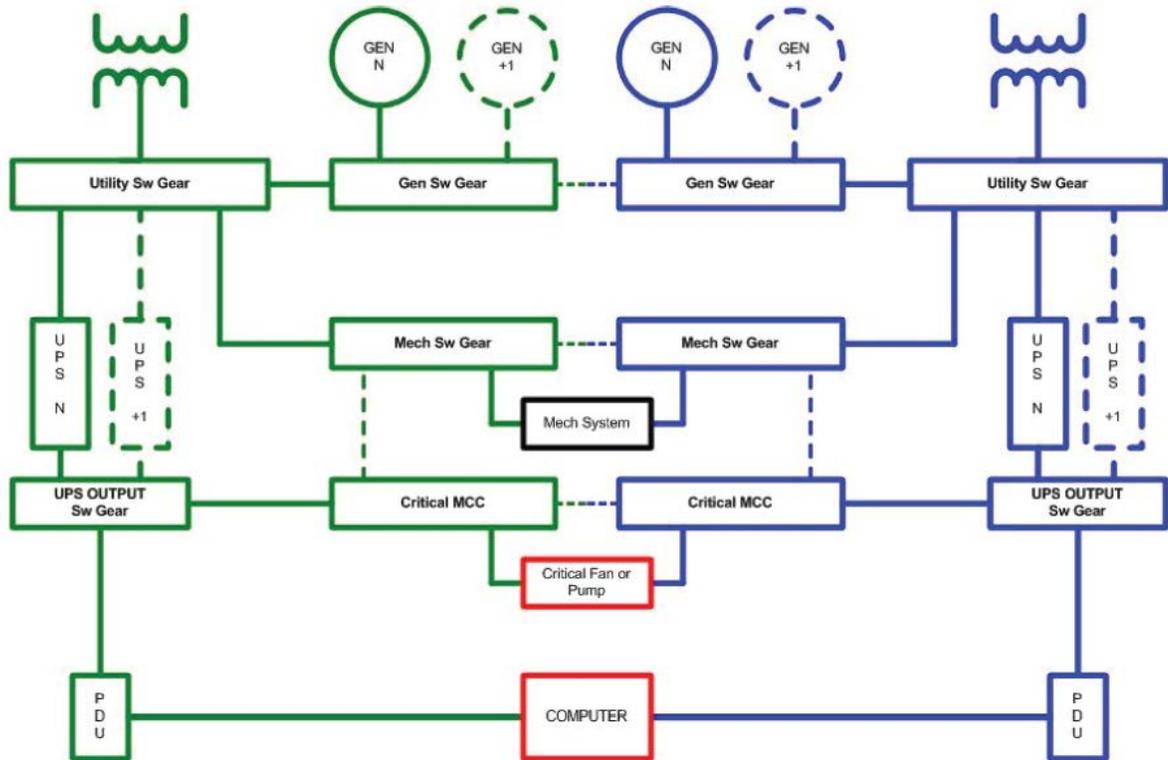
Con respecto a la conectividad y comunicaciones, se cuenta con múltiples proveedores activos simultáneamente, áreas de recepción de servicio, de distribución de los mismos, áreas de servidores y áreas de acceso y de gestión totalmente separadas entre sí.

En todos sus sistemas, un data center opera con total redundancia en caliente. A nivel eléctrico, cuenta con un diseño 2(N+1). Cada UPS debe contar con sistemas de bypass manual para mantenimientos o fallas. Se debe tener un sistema de monitoreo del estado de las baterías. Se debe contar también con energía provista por mínimo 2 subestaciones eléctricas independientes y que sean alimentadas por circuitos distintos y activos de manera simultánea. Así mismo la detección transferencia eléctrica debe ser totalmente automatizada entre los distintos subsistemas que lo conforman.

Este nivel de data centers es el más avanzado hasta el momento y ofrece los más altos niveles de disponibilidad para sus usuarios. Es por esto que su aplicación es principalmente para compañías con presencia en el mercado internacional, con operaciones 24x7x365 en mercados altamente competitivos; compañías basadas en el comercio electrónico, entidades financieras, bolsas de valores, proveedores de servicio de Internet de primer nivel, entre otras [15].

A continuación se presenta un diagrama de la infraestructura eléctrica típica de un data center Tier IV.

Figura 8. Esquema típico de un sistema eléctrico Tier IV



3.2 SOLUCIONES A LAS EXPECTATIVAS NO CUMPLIBLES ACERCA DE LA DISPONIBILIDAD DE UN DATA CENTER

Los data centers Tier IV son los que más altos niveles de disponibilidad pueden ofrecer a sus usuarios; sin embargo, no pueden garantizar la disponibilidad exigida por los usuarios de 99.999% (cinco nueves). Como ya fue enunciado, la disponibilidad máxima que un data center tier IV puede ofrecer es de 99.995%. Para lograrlo, se asume que solo ocurrirá una interrupción como resultado de una alarma de incendios o de la activación de EPO, una vez cada 5 años. Según el TUI, solo el 10% de los data centers tier IV en el mundo puede ofrecer este nivel de disponibilidad. A no ser que se pueda controlar y evitar las fallas en todas las actividades realizadas por el personal técnico del data center, se presume que al menos un incidente adicional puede ocurrir dentro de ese lapso de 5 años.

Aunque se asume que las interrupciones se solucionarán de manera casi instantánea, lo que requiere mantener personal operativo 24x7x365 en el data

center, la realidad es que se puede requerir al menos 4 horas para el restablecimiento total de la operación y del acceso a los datos y/o servicios afectados. La disponibilidad de 99.995% de un data center tier IV es una medida promedio calculada en un período de 5 años.

Una manera de lograr altas disponibilidades, incluso cercanas a los “cinco nueves”, es logrando evitar las activaciones accidentales de las alarmas de incendio o de los sistemas EPO (que son los únicos eventos que se permite puedan generar una interrupción de la operación del data center). Dentro de las medidas preventivas que se pueden implementar se cuentan las siguientes:

- Detección de humo de alta sensibilidad.
- Infraestructura física retardante a las llamas.
- Señalización de los elementos activadores de las alarmas o EPOs
- Entrenamiento constante del personal
- Limitar el acceso del personal no autorizado (personal no técnico) a las áreas críticas del data center
- Generar bienestar laboral en todo el personal a cargo del data center, que permita que el mismo adquiera un gran sentido de pertenencia con el mismo.

Todas estas medidas, si se implementan, minimizan los riesgos de interrupciones en la operación de un data center.

Otras soluciones que se pueden implementar son:

- Ubicación de los elementos redundantes de la infraestructura de IT en un lugar diferente y separado de la principal. Así, cualquier evento que pueda generar una interrupción en el servicio en uno de los sitios, no afectará a la parte redundante.
- Especializar a un grupo del personal de TI en el manejo eficiente y en la recuperación ante incidencias de las aplicaciones y/o sistemas de misión crítica; de esta manera, se reduce el tiempo de 4 horas promedio que se puede tardar la recuperación específicamente para esas aplicaciones y/o sistemas.

Todos estos aspectos operativos aplicados correctamente pueden mejorar la disponibilidad ofrecida por cualquier data center (incluso de niveles menores que tier IV), pero son particularmente importantes en los data center con disponibilidades tipo “cuatro nueves” que requieren disponibilidad de “cinco nueves” [15].

4. MODELAMIENTO PROPUESTO DE UN DATA CENTER ORIENTADO A BRINDAR SERVICIOS DE CLOUD COMPUTING

Se propone diseñar un Data Center que provea servicios de Cloud Computing involucra el desarrollo de una serie de conocimientos que va desde la capa física hasta la capa lógica.

Teniendo en cuenta los niveles de disponibilidad ofrecidos por los Data Centers, así como sus características de redundancia y tolerancia a fallos, se recomienda que el tier mínimo con que se debe diseñar un data center para prestar servicios de cloud computing es el Tier III. Esto se deduce de la falta de características en los Tier inferiores que permitan que los servicios de cloud computing sean ofertados con la confiabilidad y disponibilidad requeridos por este modelo.

Un Data Center Tier I no es viable para este tipo de modelos de negocio debido a que no ofrece ningún mecanismo de redundancia en sus subsistemas y cualquier falla que se presente en el mismo dejaría por fuera de operación todos los servicios prestados.

Un Data Center Tier II, a pesar de contar con algunos elementos redundantes dentro de su infraestructura, no cuenta con la posibilidad de la realización de mantenimiento en caliente sobre los sistemas críticos ni tampoco tolera una falla general de dichos sistemas. Es por esto que con un data center Tier II no se podría garantizar la operación ininterrumpida de los servicios.

Por el contrario, un Data Center Tier III cumple con la mayor parte de requerimientos del modelo Cloud Computing, como son, redundancia en todos los sistemas que minimiza la posibilidad de fallas que afecten toda la operación y además habilitan la posibilidad de hacer mantenimientos programados en caliente sin necesidad de tener que apagar los componentes críticos del Data Center.

Teniendo en cuenta el Tier III como la base para el diseño de data centers para brindar servicios de cloud computing, a continuación se presentarán las principales consideraciones que se deben tener en cuenta para lograr el objetivo de brindar servicios de calidad mediante este modelo con un diseño de data centers eficiente. Las consideraciones principalmente se tomarán de los estándares de diseño definidos por la norma TIA-942 [16], que describe los parámetros de diseño de los data centers según el nivel de disponibilidad que ofrecen (Tiers).

4.1 ARQUITECTURA DEL DATA CENTER

4.1.1 CONSIDERACIONES GENERALES

La estructura general de un data center debe estar construida en acero o concreto y debe cumplir con las normas de sismo resistencia aplicables en el lugar donde haya sido construido.

Las losas en pendiente deben tener un espesor mínimo de 127 mm y deben tener una capacidad de carga de 12 kPa. Las losas elevadas deben ser hechas de concreto de alta resistencia y deben tener un mínimo de 100 mm de cubierta sobre las vigas de metal en zonas sísmicas tipo 3 o 4, para permitir la aplicación adecuada de los anclajes epóxicos. El piso del área de UPS deben ser diseñados para soportar mínimo cargas de 12 a 24 kPa en las cubiertas y cerchas, 19.2 kPa para vigas principales, columnas y bases.

El techo debe diseñarse para soportar el peso real de los equipos más un remanente de 1.2 kPa para las cargas suspendidas. El techo del área donde se encuentran las UPS deben diseñarse para soportar una carga colgante de 1.4 kPa.

Todos los equipos deben anclarse adecuadamente a los elementos que los soportan. Con frecuencia, los equipos son sensibles a las vibraciones por lo cual debe garantizarse el control adecuado de las fuentes de vibración para no afectar la operación de los equipos. Los equipos que provocan vibraciones (aires acondicionados, generadores, entre otros) deben ubicarse sobre aisladores de vibración.

Todas las particiones internas de los cuartos deben ir desde el piso hasta el techo y deben tener una clasificación que permita el retardo del fuego por mínimo 1 hora.

Deben existir muelles de carga para el acceso de camiones que llevan o traen equipos y demás elementos hacia o desde el data center. Estos muelles deberán cumplir con todas las normas de seguridad para el acceso al igual que las demás entradas al complejo.

Deben existir áreas para almacenamiento con suficiente capacidad para allí alojar los inventarios de elementos como cables, papelería, tintas, cintas, entre otros, los cuales deben estar separados del área de almacenamiento de equipos críticos.

Se debe diseñar el techo de las áreas de equipos de manera tal que permita mantener dichas áreas totalmente limpias, especialmente cuando hay riesgo de contaminación por polvo u hojuelas provenientes de sistemas contra incendios.

En los data centers grandes, normalmente se recomienda la existencia de un área separada del resto del complejo para ubicar allí un centro de operaciones (NOC). Es recomendable que esta área tenga un acceso directo a los cuartos de equipos del data center.

4.1.2 UBICACIÓN FÍSICA

La ubicación de un data center debe realizarse en un terreno alto para así minimizar el riesgo de inundaciones pero, debido a que la dificultad para refrigerar los equipos aumenta con la altura, un Data Center no debe ubicarse por encima de los 3050 metros sobre el nivel del mar.

Debe ser de fácil acceso para el personal y debe estar libre de interferencia electromagnética. Se recomienda, según el estándar, que se tenga en cuenta mínimo acceso por dos vías diferentes para brindar redundancia en el acceso del personal a las instalaciones del data center.

Se recomienda que el edificio donde se ubique el data center sea especialmente dedicado para este propósito, ojalá construido para el mismo. Sin embargo, en caso que esto no sea posible y se deba compartir el área con otras organizaciones, estas no deben efectuar operaciones industriales. Se debe evitar principalmente la ubicación de un data center en el mismo edificio donde operen restaurantes o cafeterías para evitar el riesgo de incendios.

4.1.3 CONSIDERACIONES ARQUITECTÓNICAS

Además de las consideraciones recomendadas por la norma, se debe tener en cuenta unas características arquitectónicas adicionales que le brindan el soporte

para ofrecer la disponibilidad característica de este tipo de data centers. Dentro de estas características se puede mencionar:

- Debe poseer varios puntos de ingreso al edificio y debe tener puntos de control de seguridad para evitar el acceso a personal no autorizado.
- Debe tener varias vías de acceso para proveer redundancia y/o separar el acceso de los empleados y los proveedores
- El Data Center no debe tener ventanas hacia el exterior.
- La construcción del edificio debe proveer protección contra radiación electromagnética. Por ejemplo, las estructuras construidas con acero pueden proveer este blindaje. Como alternativa, se puede insertar en las paredes un blindaje para armar una Jaula de Faraday; este blindaje podría armarse con hojas de aluminio.
- Los sistemas de acceso al área de equipos del data center deben ser lo suficientemente seguros para prevenir que dos personas puedan ingresar haciendo uso de una sola credencial de acceso.
- Todo el perímetro del edificio debe protegerse con vallas de seguridad y deben monitorearse los alrededores mediante circuito cerrado de televisión con capacidad de visión nocturna así como con sistemas de detección de intrusos basados en microondas.
- El acceso al sitio debe asegurarse mediante mecanismos de identificación y autenticación de usuarios. Para las áreas críticas del data center se debe proveer sistemas de seguridad adicionales.
- El Data Center debe contar con un área dedicada al monitoreo y control de todos los sistemas de seguridad.
- La carga mínima que debe soportar el piso del área de equipos del data center es de 12 kPa (250 lb / ft²) para la carga principal y 2.4 kPa (50 lb / ft²) para los elementos colgantes debajo del piso falso.

4.2 REQUERIMIENTOS DE TELECOMUNICACIONES E INFRAESTRUCTURA ELÉCTRICA

Al igual que con los requerimientos arquitectónicos, un data center debe cumplir con todos los requerimientos de la norma y adicionalmente contará con requerimientos adicionales que se especifican a continuación:

El Data Center, en aras de poder mantener los servicios de cloud computing activos en todo momento, deberá contar con mínimo 2 proveedores de servicio de Internet independientes y que el servicio provenga de puntos de presencia distintos.

Los cables de acceso de los servicios entregados por cada proveedor deberán separarse mínimo 20 metros entre sí para poder ser considerados como rutas independientes de acceso.

Cada proveedor de acceso deberá ingresar al data center en un cuarto de entrada separado, preferiblemente ubicados en lugares opuestos del data center. En ningún caso se deberán compartir los sistemas eléctricos, de aire acondicionado o de cableado entre los distintos cuartos de entrada así como tampoco se deben compartir los equipos activos de recepción de los servicios entre los diferentes proveedores de acceso.

Se deberá contar con rutas de backbone redundantes entre los cuartos de entrada y los demás cuartos principales del data center (MDA, HDA).

El cableado horizontal entre el HDA y el MDA debe ser redundante tanto para las LAN como para las SAN. Las conexiones redundantes deben guiarse mediante enrutamiento independiente.

Se deberá contar con sistemas de respaldo pasivo en caliente para todos los sistemas críticos de telecomunicaciones (equipos de los proveedores, enrutadores y switches de core).

Todo el cableado deberá estar documentado de manera física y sistematizada, mediante bases de datos o con software especializado en administración de cableado estructurado. Esto es una condición obligatoria para que un data center pueda ser considerado tier III.

Algunos de los puntos de falla potenciales que existen en estos data center son:

- Cualquier evento catastrófico dentro del MDA puede interrumpir todas las operaciones del data center.
- Cualquier evento catastrófico dentro del HDA puede interrumpir todas las operaciones hacia el área de servidores.

Los equipos a instalar en el cuarto de comunicaciones y salas de equipos de un data center deben contar con los elementos de protección eléctrica tales que puedan ser conectados y aterrizados adecuadamente en el rack donde se van a instalar. Es por esto que los requerimientos de telecomunicaciones y eléctricos van de la mano dada la interacción que existe entre los dos. A continuación se presentan los requerimientos eléctricos de los data centers para permitir que nuestros equipos queden conectados adecuadamente y de manera segura.

4.2.1 SISTEMAS ELÉCTRICOS EN LOS DATA CENTER

4.2.1.1 Acometidas Principales: Se debe preferir la alimentación eléctrica proveniente de circuitos que brinden servicio también a instituciones como hospitales, los cuales tienen alta prioridad cuando existen fallas en el fluido. No son convenientes las acometidas provenientes de circuitos que también proveen servicio a empresas del sector industrial dado que estas poseen máquinas que pueden generar transientes y armónicos en las acometidas vecinas.

Se debe preferir las acometidas que llegan vía subterránea en lugar de las aéreas debido a que estas últimas son más propensas a accidentes por roce de vehículos, rayos, árboles y/o vandalismo.

El mecanismo de conexión primario debe diseñarse para permitir el crecimiento, mantenimiento y redundancia del sistema. Se recomienda un sistema de doble fin (principal – unión – principal) o una configuración aislada redundante. La conexión primaria debe ser sobredimensionada desde el principio ya que este sistema es el menos expandible una vez inicia la operación del data center.

Los interruptores deben ser intercambiables y el sistema debe diseñarse para permitir el mantenimiento de cada uno de los elementos que lo conforman. Se debe instalar sistemas de supresión de transientes de voltaje en cada nivel del sistema de distribución eléctrica.

4.2.1.2 Sistemas de UPS: En cuanto a las UPS, el estándar no define si se debe utilizar UPS online o de línea interactiva pero si deben tener el tiempo de respaldo suficiente para permitir que los generadores de energía alterna entren en funcionamiento y se pueda garantizar la operación continua de todos los equipos. Por esto se define que la función primaria del sistema de UPS es garantizar energía suficiente a los equipos durante el tiempo en que los generadores alternativos arrancan o cuando regresa el fluido eléctrico principal del edificio.

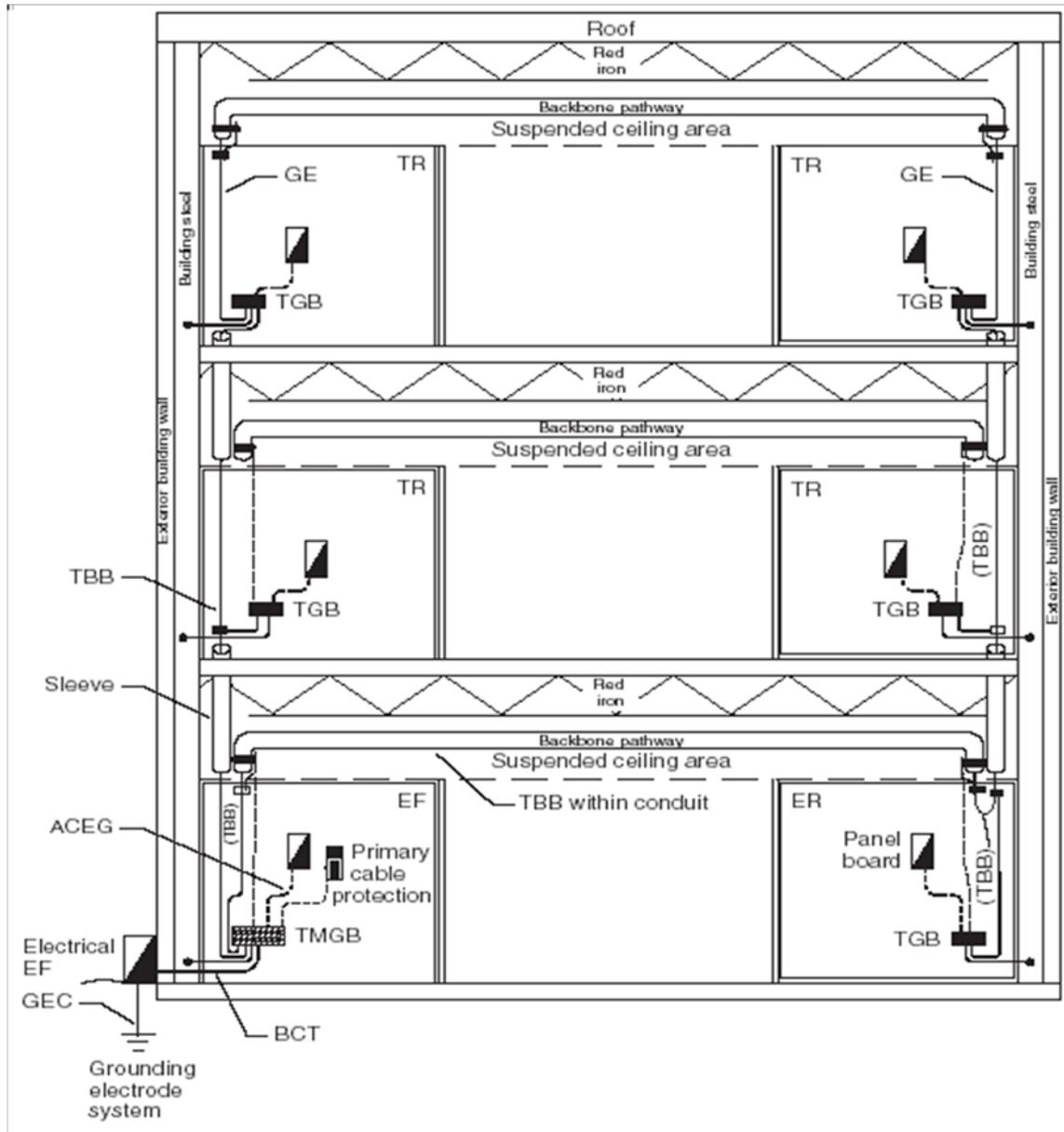
Se deben preferir sistemas de UPS que permitan monitorear el estado de operación y vida útil de las baterías de manera constante para así poder planificar cualquier mantenimiento y/o reemplazo de dichos elementos en caso de ser necesario.

Se sugiere que las UPS y baterías se encuentren en un cuarto independiente de los equipos activos a los cuales protegen y dicho cuarto debe contar con sistemas de enfriamiento de precisión que permitan mantener la temperatura adecuada de operación de las baterías, las cuales sufren una degradación de su vida útil cuando se someten a altas temperaturas.

4.2.1.3 Sistemas de Tierras: Se debe instalar un perímetro de aterrizaje eléctrico alrededor de todo el edificio, ubicando varillas de cobre de mínimo 3 metros de longitud por $\frac{3}{4}$ " de grosor, separándolas entre si cada 6 a 12 metros. Para unir las varillas entre si se debe utilizar cable #4/0 AWG como mínimo, el cual deberá enterrarse a una profundidad mínima de 1 metro y el sistema debe separarse de las paredes exteriores del edificio una distancia mínima de 1 metro. A este sistemas se deben conectar todos los elementos mayores del sistema eléctrico del edificio como los generadores, transformadores, UPS, sistemas de telecomunicaciones y pararrayos.

Para los data centers, existe una norma específica para este propósito que es la TIA-J-STD-607-A la cual recomienda el estándar de aterrizaje adecuado para este tipo de infraestructura. A continuación se presenta un diagrama con los elementos principales de los sistemas de tierras en un Data Center, el cual se analizará con más detalle y se determinarán los elementos necesarios para construirlo.

Figura 9. Especificación del aterrizaje de un Data Center



<p>ACEG: Alternating current equipment ground BC: Bonding Conductor BCT: Bonding Conductor for Telecommunications EF: Entrance Facility ER: Equipment Room GE: Grounding Equalizer</p>	<p>GEC: Grounding Electrode Conductor TBB: Telecommunications Bonding Backbone TGB: Telecommunications Grounding Busbar TMGB: Telecommunications Main Grounding Busbar TR: Telecommunications Room</p>
---	--

A continuación se detallan las consideraciones de diseño exigidas por la norma TIA-J-STD-607-A para los elementos principales del sistema de aterrizaje

El TBB o Cableado de backbone principal del sistema de aterrizaje debe dimensionarse teniendo en cuenta las distancias a recorrer, según la siguiente tabla:

Tabla 3. Dimensionamiento del conductor de unión vertical (TBB)

Longitud en metros del TBB	Calibre adecuado del TBB (AWG)
Menos de 4	6
4-6	4
6-8	3
8-10	2
10-13	1
13-16	1/0
16-20	2/0
Más de 20	3/0

Para las barras de los TGB y TMGB, estas deben cumplir con los siguientes requerimientos mínimos:

- Deben estar hechas de cobre. Se recomiendan platinadas para evitar corrosión.
- Perforaciones de 5/16" con centros a 5/8" y perforaciones de 7/16" con centros a 1"
- Espacio mínimo de 2" entre soporte y barra
- La barra debe quedar aislada de su soporte

Cada gabinete o rack de telecomunicaciones se debe conectar al TGB del TR respectivo mediante patch cords con terminales de ponchado irreversible, con 2 perforaciones. En el rack se debe contar con un elemento de conexión a tierra y si no se cuenta con él, se debe pelar la pintura para poder realizar el aterrizaje del mismo al TGB. Para esto se deben utilizar tornillos especiales para remover la pintura al apretar y el área de conexión se debe pulir y preparar con antioxidante para evitar la corrosión.

El aterrizaje de los elementos ubicados en el rack (Ej. servidores, elementos de red) al gabinete se debe realizar a través del chasis de los mismos, según las instrucciones de cada fabricante.

En cada gabinete se deben instalar manillas de descarga antiestática para que los técnicos las utilicen al manipular los equipos y elementos electrónicos ubicados en el. Estas manillas deben estar conectadas al chasis del gabinete para asegurar la descarga adecuada de la energía hacia el sistema de tierras.

4.2.2 CONSIDERACIONES DE DISEÑO A NIVEL ELÉCTRICO

- Deben contar con redundancia de nivel N+1 en cada subsistema eléctrico (Ej. enrutamientos, generadores, UPS, PDUs).
- Al menos debe haber 2 alimentaciones eléctricas principales para el data center, que puedan proveer voltaje medio o alto (mayor a 600 V). Estas alimentaciones deben poderse escoger de manera automática.
- Debe contar con sistemas de intercambio automático que permitan detectar una falla en el fluido eléctrico principal y de inmediato inicie el proceso de arranque de los generadores alternos y transfiera la carga a estos.
- Se debe contar con un centro de monitoreo de todo el sistema eléctrico. Además se debe instalar sistemas tipo PLC (Programmable Logic Control) para administrar los sistemas mecánicos y optimizar su eficiencia e indicar las alarmas necesarias en caso de fallas.

4.3 VIRTUALIZACIÓN

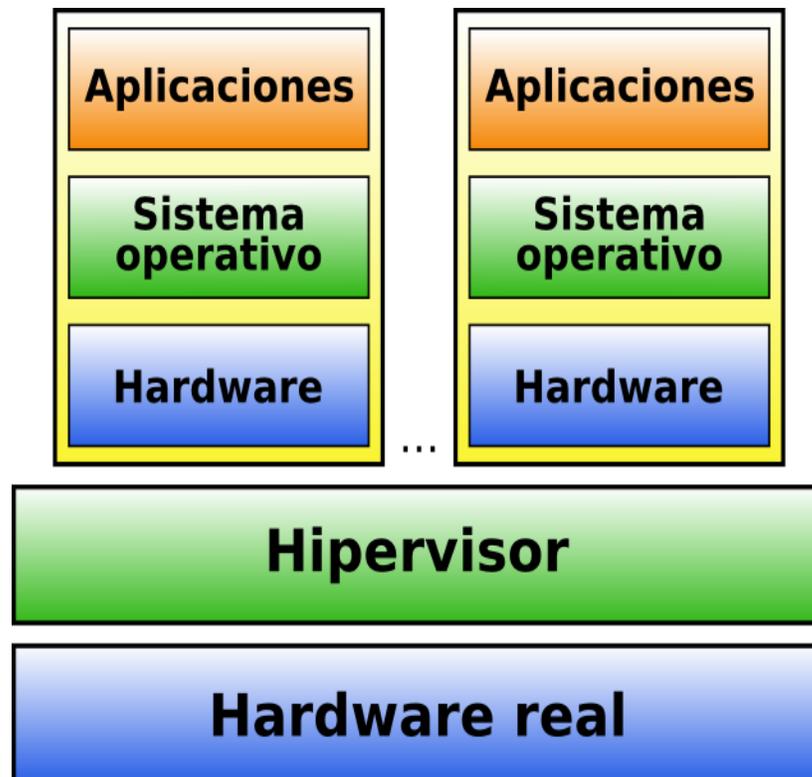
Dentro de los aspectos de diseño adicionales que se debe tener en cuenta para llevar un Data Center al nivel de poder brindar servicios de Cloud Computing de manera óptima, está el requerimiento de la virtualización, uno de los pilares en que se desarrollaron los modelos de servicio de Cloud Computing.

Un Data Center Tier III que esté diseñado pensando en virtualización es suficiente para brindar servicios de Cloud Computing.

Una plataforma de virtualización es un conjunto de software y hardware que simula la ejecución de equipos o sistemas operativos distintos a los reales. Esto se consigue ocultando las características de la plataforma real y proporcionando otra plataforma abstracta y simulada.

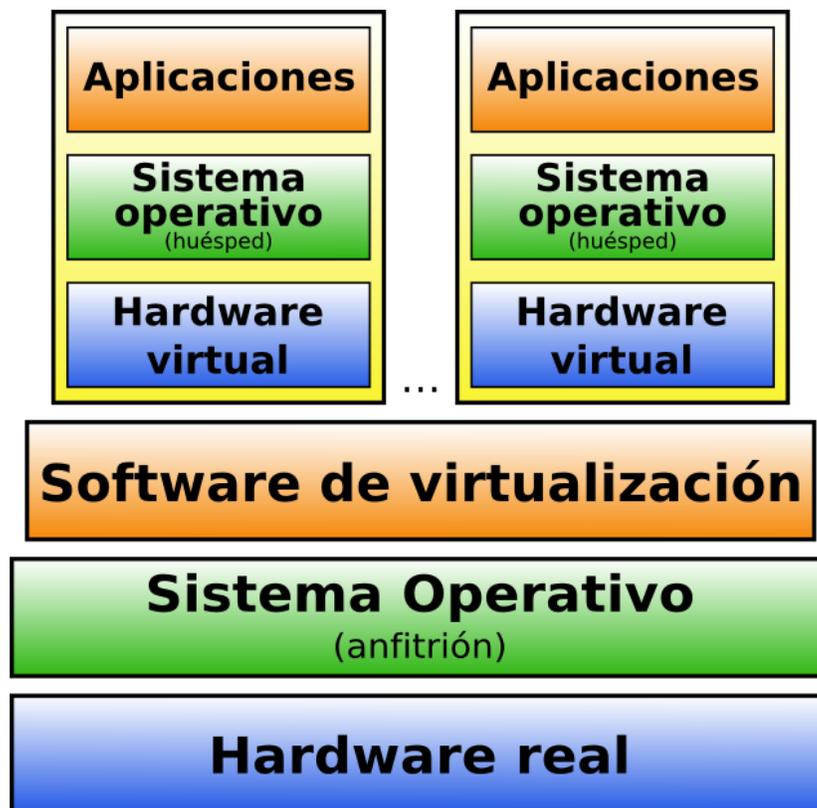
Existen dos tipos de virtualización. La Tipo 1, en la cual se instala un hypervisor (monitor de máquinas virtuales) directamente sobre la máquina física y administra directamente los sistemas operativos de cada máquina virtual que se cree sobre el mismo. Entre el hypervisor y la máquina real no existe ninguna capa de software adicional.

Figura 10. Virtualización Tipo 1



La virtualización Tipo 2, se basa en software de virtualización que se instala sobre un sistema operativo anfitrión que se encuentra instalado en la máquina física real y éste se encarga de gestionar las máquinas virtuales y sus sistemas operativos. Este tipo de virtualización agrega una capa de software adicional (Sistema Operativo anfitrión) entre el hypervisor (software de virtualización) y la máquina real.

Figura 11. Virtualización Tipo 2



En un Data Center preparado para aplicaciones de cloud computing, se debe minimizar las capas existentes entre la máquina física y las máquinas virtuales para así evitar errores o fallos adicionales que se puedan presentar en dichas capas. Es por esto que se recomienda la virtualización tipo 1 en la cual no hay necesidad de instalar un sistema operativo anfitrión en la máquina real para poder crear las máquinas virtuales requeridas.

4.4 COMO PUEDE UNA EMPRESA COLOMBIANA MIGRAR SU DATA CENTER PARA BRINDAR SERVICIOS EN LA NUBE A NIVEL PRIVADO

Para una empresa colombiana de gran tamaño, la cual puede contar con centros de datos para soportar la operación de su negocio, puede resultar fructífera una inversión para migrar dicho centro de datos a un nivel tal que le permita migrar sus sistemas a la nube y establecer nuevas maneras de comunicación y gestión de la información a nivel interno, con sus proveedores y con sus clientes. Para ello es

necesario que el centro de datos cumpla con la mayoría de requerimientos estudiados anteriormente y a continuación se estudiará como llevar ese data center al cumplimiento de dichos requerimientos, especialmente a nivel de conectividad e infraestructura de servidores.

Primero partamos de algunos supuestos que son comunes en la mayoría de compañías grandes en Colombia y que servirán de base para el ejercicio de diseño que se va a realizar:

- La compañía cuenta con varias sucursales a nivel nacional.
- La compañía tiene una planta de personal bastante grande, entre empleados de planta, contratistas y personal que brinda servicios externalizados (outsourcing).
- Cuentan con centros de datos donde tienen centralizada la información crítica de la empresa, con cierta cantidad de servidores especializados para los sistemas de información requeridos para la gestión de dicha información.
- Tienen una gran cantidad de clientes y proveedores para sus insumos y servicios.
- Cuentan con estabilidad financiera y una buena planificación del uso de los recursos económicos.
- Cuentan con un departamento de TI especializado para la administración de los centros de datos que poseen y su personal conoce de las tecnologías básicas de redes e infraestructura para su gestión.

4.4.1 ELECCIÓN DE LOS PROVEEDORES DE ACCESO

El primer paso a realizar es la escogencia adecuada de los proveedores de acceso a Internet, los cuales deben brindar altos niveles de calidad en el servicio y una alta disponibilidad. Entre los operadores que pueden garantizar un compromiso de calidad y disponibilidad suficientes para permitir la operación continua del centro de datos, podemos mencionar:

- **ISA Internexa S.A. E.S.P.:** Ofrece una disponibilidad de 99.6% en el servicio de Internet y cuenta con una red de fibra óptica a nivel nacional e internacional con topología anillada que garantiza una alta disponibilidad en caso de fallas. Cuenta con salida internacional a través de los principales cables submarinos que llegan a Colombia, principalmente en el cable ARCOS 1. Su centro principal de operaciones está en Medellín.
- **UNE EPM Telecomunicaciones:** Ofrece una disponibilidad en el servicio de 99.6% y cuenta con el apoyo para su transporte nacional a través de la infraestructura de ISA Internexa. Tiene su centro de operaciones en Medellín y también posee salida internacional a través de los principales cables submarinos que llegan al país.
- **Telefónica Telecom:** Ofrece una disponibilidad de 99.6% para los canales dedicados de Internet y cuenta con salida internacional propia a través del cable submarino SAM-1 y también tiene acuerdos de capacidad en otros cables que llegan al país. Su centro de operaciones principal se encuentra en la ciudad de Bogotá.
- **Global Crossing:** Es un operador de Internet Level 1 y cuenta con su propia infraestructura a nivel mundial con punto de presencia en Colombia. Su oficina principal se encuentra en la ciudad de Bogotá.
- **Columbus Networks:** Esta compañía es la propietaria y operadora del cable submarino ARCOS-1, principal cable submarino que llega a Colombia con configuración anillada, y del cable CFX que es una línea directa entre la Florida y Cartagena. Su cabecera de llegada de cable se encuentra en Cartagena.

Para el cumplimiento de los requerimientos exigidos para la computación en la nube, se debe escoger mínimo dos operadores de acceso para así poder brindar la redundancia requerida.

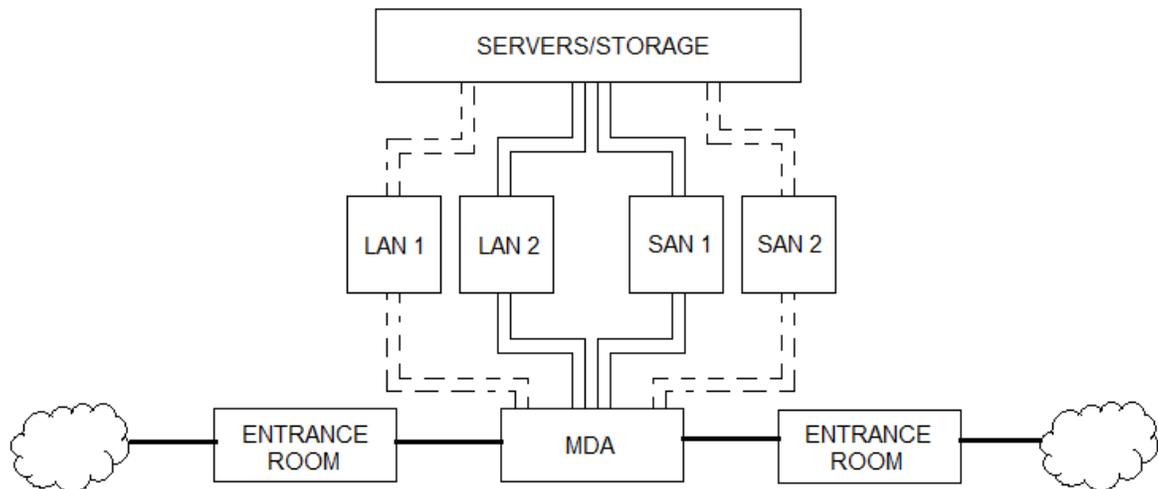
4.4.2 INFRAESTRUCTURA DE RED

Un Data Center que vaya a prestar servicios en la nube, debe diseñarse tipo Tier III, para poder garantizar la continuidad de la operación de los servicios. Es por

esto que en todos sus subsistemas, incluyendo el de comunicaciones, debe proveer redundancia de tipo N+1.

A continuación se presenta un esquema de diseño de la arquitectura de networking de un data center Tier III:

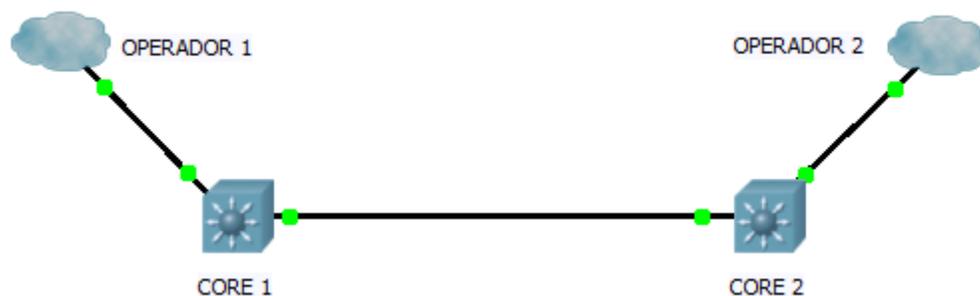
Figura 12. Arquitectura de networking de un data center Tier III.



Como primera medida, para cumplir con este requisito, se debe contar con mínimo 2 sistemas de borde (core) en el MDA del data center, uno activo y uno de respaldo. Además, todas las acometidas de cableado deben ser redundantes, como se evidencia en la figura 12, estando siempre una componente activa y una pasiva con arranque en caliente.

A continuación se presenta una alternativa de diseño para el core del data center que cumple con estos requerimientos:

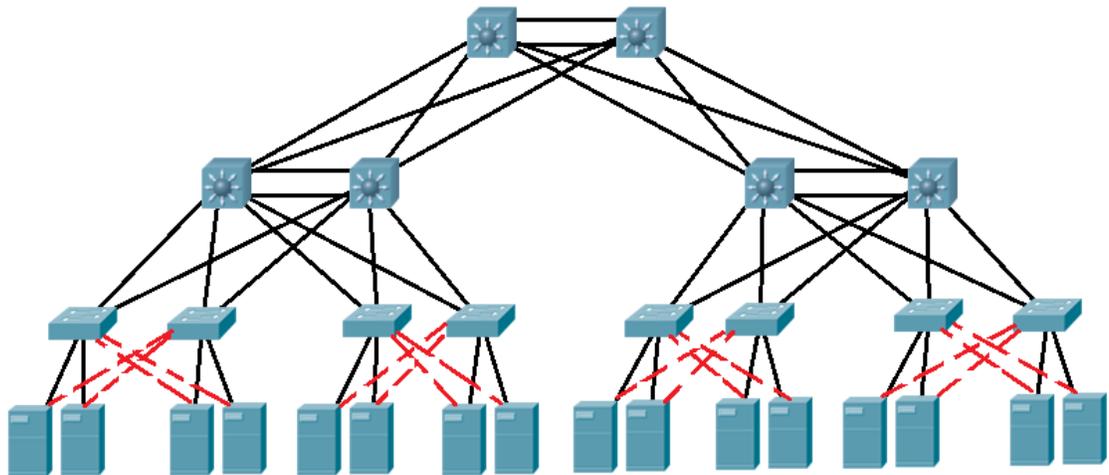
Figura 13. Estructura de la capa Core de un Data Center



La estructura de red lógica de un data center tier III se podría diseñar de acuerdo con el siguiente diagrama, con el objetivo de lograr la redundancia requerida y

tener suficientes equipos para salvaguardar la información. El diseño consiste en la capa de core, vista anteriormente, la capa de agregación/distribución y la capa de acceso:

Figura 14. Diagrama lógico de red



En la figura 14 vemos que los servidores de mantienen conexiones redundantes a la red, pero una de ellas es la que se encuentra activa en un momento de terminado. Con este requerimiento mínimo se puede lograr una alta disponibilidad y se puede brindar servicios de cloud computing con seguridad y eficiencia.

No se debe olvidar la virtualización como base de la infraestructura de servicios de la nube, con el objetivo de maximizar la eficiencia en el uso de los recursos y poder mantener servidores paralelos para la redundancia requerida y facilitar la recuperación ante desastres movilizand las máquinas virtuales de un servidor a otro de manera sencilla, utilizando las herramientas de virtualización que se ofrecen en el mercado.

Existen varias alternativas para la virtualización tipo 1 que se recomienda para este tipo de data centers. El líder mundial en estas tecnologías es VMware, el cual cuenta con herramientas gratuitas y muy eficientes para el manejo de la virtualización lo cual permitirá a las empresas ahorrar costos en licenciamiento. Esta alternativa es el producto VMware ESXi, que ofrece las características básicas para virtualización (hypervisor) con una interfaz de administración fácil y rápida. Para el uso de esta herramienta se debe considerar la escogencia de servidores directamente compatibles con la misma; se puede verificar en el sitio

web de vmware la lista de marcas y modelos de servidores compatibles con esta plataforma de virtualización.

Pero, existen otros proveedores de herramientas de virtualización como las versiones Data Center de Microsoft Windows Server, las cuales incluyen sistemas de virtualización licenciados que pueden dar muy buenos resultados para los que requieren montar su infraestructura basada en sistema operativo Windows.

A nivel de los servicios que se pueden brindar a través de esta infraestructura para el uso de la empresa se pueden mencionar, entre otros, los siguientes:

- Sistemas de CRM
- Correo electrónico corporativo
- Software de ERP
- Calendarios y Agendas
- Software de gestión documental

CONCLUSIONES Y RECOMENDACIONES

Luego de culminada la presente investigación se puede concluir que el Cloud Computing es aún un concepto muy reciente que está en plena expansión a nivel mundial, siendo las grandes empresas de Internet como Google o Amazon, por mencionar algunas, las que están impulsando su utilización y aprovechamiento generando servicios basados en esta tecnología y masificándolos a nivel mundial.

Implementar servicios de Cloud Computing requiere la implementación de Data Centers con altos niveles de disponibilidad y seguridad que garanticen una operación continua de los servicios en ellos implementados. Esto requiere la inversión de recursos económicos importantes pues este tipo de tecnología no es barata y se debe tener en cuenta que se debe implementar redundancia en todos los sistemas para poder garantizar la disponibilidad exigida por Cloud Computing.

Al estudiar el panorama del cloud computing en Colombia, el resultado es poco alentador, ya que en el país no existen muchas compañías grandes que tengan la capacidad de financiar la implementación de infraestructuras de data centers capaces de proveer servicios en la nube de manera globalizada. Hasta la fecha, solo la multinacional Telefónica Telecom ha promocionado oficialmente servicios bajo el modelo de Cloud Computing y esto lo ha logrado debido a que es una compañía que tiene el respaldo internacional de Telefónica en España, su casa matriz. El resto de compañías de Internet locales, poseen data centers cuyo propósito es la prestación de Internet y sus servicios asociados, pero no fueron diseñados pensando en Cloud Computing.

Uno de los mecanismos que puede reducir costos de implementación de data centers para cloud computing y que permite que los mismos brinden estos servicios adecuadamente es la virtualización, tecnología que permite un mayor aprovechamiento de los recursos tecnológicos instalados en los Data Centers y asegura la posibilidad de migrar esos servicios a otras infraestructuras con un mínimo esfuerzo, siendo esto fundamental para lograr altos niveles de disponibilidad de los servicios ofrecidos.

La revisión de la documentación y literatura obtenida durante el proceso de investigación permite generar conclusiones relacionadas con la adopción de la tecnología de Cloud Computing en Colombia, sus ventajas y obstáculos para su implementación o adopción.

La mayoría de empresas en el país caen en la categoría de pymes, las cuales se pueden beneficiar de los servicios en la nube pero sin implementar ellos mismos esta tecnología sino contratando los servicios con los proveedores mundiales que ya los tienen implementados.

En Colombia, las empresas pueden migrar sus recursos de TI a la nube si logran visualizar las oportunidades de crecimiento y ahorro de recursos que esta tecnología les puede brindar. Pero las oportunidades y amenazas existentes en el proceso son distintas para una compañía grande que para una pyme. Una empresa grande puede divisar las siguientes oportunidades al migrar a cloud computing:

- Posibilidad de reducir costos operativos.
- El Cloud Computing le permite a las empresas centrar sus recursos en su negocio principal.
- Le brinda la oportunidad de ser de los primeros en adoptar las nuevas tecnologías.

Pero también existen obstáculos al proceso de migración a la nube para este tipo de empresas, entre las que se pueden mencionar:

- Los sistemas actuales que poseen son bastante costosos como para deshacerse de ellos del todo después de tan grande inversión.
- Temor del mal manejo de la información de la empresa por parte de terceros.
- Percepción de la pérdida del control sobre sus datos y sistemas.

Para una pequeña o mediana empresa (pymes) las oportunidades y obstáculos son similares, pero existen diferencias. Dentro de las oportunidades para las pymes podemos encontrar:

- Menor costo de inversión inicial en TI
- Posibilidad de reducir costos operativos

- Ganar acceso a las últimas tecnologías

Se puede observar que las oportunidades son casi las mismas, pero a nivel de obstáculos si se presentan diferencias, como son:

- Percepción de que es algo innecesario y que se puede resolver montando infraestructuras pequeñas privadas.
- Desconocimiento del concepto cloud.
- Desconocimiento de nuevas tecnologías y sus beneficios.
- Desconocimiento de proyectos de financiación a nivel gubernamental para la adopción de nuevas tecnologías para las PyMES.

Después de visualizar estas oportunidades y obstáculos para la adopción del cloud computing en Colombia se puede concluir que las grandes empresas, que ya cuentan con centros de datos de alto nivel para el manejo de sus operaciones, pueden invertir en esas infraestructuras y llevarlas a un nivel tal que les permitan implementar servicios tipo cloud computing a nivel privado para su operación, lo cual les daría la oportunidad de establecer mejores relaciones con sus clientes, proveedores y su propio personal mediante la centralización de la información y el acceso ubicuo a la misma mediante la plataforma de Internet.

Para las pymes, por otra parte, la mejor opción para ahorrar costos puede ser la contratación de servicios prestados por los grandes operadores, los cuales brindan altos niveles de confiabilidad, seguridad y disponibilidad. De esta manera sus recursos, que son limitados, pueden ser invertidos en la operación del negocio y no en la infraestructura de TI que se requeriría para apoyar dicha operación. En la nube se pueden encontrar todas las herramientas necesarias para manejar la información de una empresa pequeña o mediana a un costo bastante económico.

Sin embargo, independientemente de que las empresas sean grandes o pequeñas y de los recursos económicos con los cuales cuenten, se recomienda que cada una analice desde su perspectiva las posibilidades de implementación de servicios en la nube teniendo en cuenta las ventajas que esto puede traerles a futuro para permitirles un crecimiento sin traumatismos y con poca inversión. Se recomienda la implementación de soluciones híbridas en las cuales se adopten algunos

servicios en la nube y otros locales pero tratando de integrarlos adecuadamente para que el crecimiento de la empresa pueda realizarse de manera organizada.

Por último, se recomienda continuar con el proceso de investigación acerca de estos temas, enfocando los objetivos hacia la determinación de los requerimientos a nivel de interconexión, relacionado con las distintas tecnologías de red que conforman o confluyen en los data centers así como los equipos activos adecuados para brindar servicios con alta disponibilidad. Se puede complementar la investigación con un análisis financiero de la implementación de estas infraestructuras teniendo en cuenta principalmente los aspectos relacionados con las redes y equipos de interconexión, pero sin dejar de lado los demás aspectos relacionados con el diseño de data centers (Ej. eléctricos, arquitectónicos, mecánicos, seguridad).

REFERENCIAS BIBLIOGRÁFICAS

- [1] Cloud Computing – Una Perspectiva para Colombia, Abril de 2010. Mesa Sectorial Cloud Computing (CINTEL, NEC) p.8.
http://www.interactic.com.co/dmdocuments/clud_computing.pdf
- [2] Deloitte, Un Paseo por las e-Nubes; 2010, p.2 -4.
<http://www.deloitte.com/assets/Dcom-Ecuador/Local%20Assets/Documents/ERS/Cloud%20computing%20-%20Un%20paseo%20por%20las%20e-nubes.pdf>
- [3] Wikipedia, Centros de Procesamiento de Datos. [Artículo de Internet] http://es.wikipedia.org/wiki/Centro_de_procesamiento_de_datos [Consulta: 14 de Mayo del 2011].
- [4] Arregocés, Mauricio. Cisco Press, Data Center Fundamentals p.5-26 <http://www.ciscopress.com> [Consulta: 18 de Mayo 2011].
- [5] Arregoces, Mauricio. Cisco Press, Data Center Fundamentals <http://www.ciscopress.com> [Consulta: 20 de Mayo 2011].
- [6] Wikipedia, Data Centers. [Artículo de Internet] http://en.wikipedia.org/wiki/Data_center [Consulta 1 Junio del 2011].
- [7] The Uptime Institute, Inc. White Paper; 2010. Data Center Tier Classification p.2 – 4.
- [8] Furht, Borko. Escalante, Armando. Handbook of Cloud Computing: Springer; 2010, 655 p.
- [9] T.Velte, Antony. Velte, Toby J. Eisenpeter, Robert. Cloud Computing A Practical Approach: McGraw Hill: 2009, 353 p (SaaS). P. 11 – 12.
- [10] T.Velte, Antony. Velte, Toby J. Eisenpeter, Robert. Cloud Computing A Practical Approach: McGraw Hill: 2009, 353 p (PaaS). P. 12-13.
- [11] T.Velte, Antony. Velte, Toby J. Eisenpeter, Robert. Cloud Computing A Practical Approach: McGraw Hill: 2009, 353 p (IaaS). P. 13-15.
- [12] T.Velte, Antony. Velte, Toby J. Eisenpeter, Robert. Cloud Computing A Practical Approach: McGraw Hill: 2009, 353 p (dSaaS). P.136.

[13] T.Velte, Antony. Velte, Toby J. Eisenpeter, Robert. Cloud Computing A Practical Approach: McGraw Hill: 2009, 353 p (Tipos de Cloud). P. 7.

[14] The Uptime Institute, Inc. White Paper; 2010. Tire Classification Define Site Infrastructure Performance. P.14 - 17.

[15] Grupo Electrotecnia, Monge, Jose Miguel. Estándares sobre Diseño y Funcionamiento de Data Center. P 5 - 47.

[16] Telecommunications Industry Association, ANSI-TIA-942-2005, Standard Definition.